

一种基于大语言模型的数据处理方法、装置及设备

技术领域

[01] 本申请涉及人工智能技术领域，尤其是涉及一种基于大语言模型（Large Language Model，简称 LLM）的数据处理方法、装置及设备。

5 背景技术

[02] 随着网络技术的快速发展，网络数据越来越多，为了从大量网络数据中检索到用户所需数据，需要由用户提供关键词，服务器基于关键词从大量网络数据中检索到用户所需数据。

10 [03] 然而，由于网络数据的数据量很大，服务器需要消耗大量计算资源才能够检索到用户所需数据，且服务器需要花费大量时间，从而导致服务器的大量资源消耗，且用户很长时间才能够获知所需数据。比如说，实际场景会部署大量摄像机，这些摄像机会采集大量图片，服务器为了从这些图片中分析用户所需数据，需要消耗大量计算资源，花费大量时间。

发明内容

15 [04] 有鉴于此，本申请提供一种基于大语言模型的数据处理方法、装置及设备，能够节省服务器的计算资源，并节省时间开销，能够快速确定出目标答案。

[05] 本申请提供一种基于大语言模型的数据处理方法，所述方法包括：

[06] 获取原始数据格式的第一输入问题；若所述原始数据格式不是文本数据格式，则将所述第一输入问题转换为文本数据格式的第二输入问题；

20 [07] 将所述第二输入问题输入给已训练的目标大语言模型，由所述目标大语言模型输出所述第二输入问题对应的文本数据格式的第一目标答案；

[08] 将所述第一目标答案转换为所述原始数据格式的第二目标答案。

[09] 本申请提供一种基于大语言模型的数据处理装置，所述装置包括：

25 [10] 获取模块，用于获取原始数据格式的第一输入问题；若所述原始数据格式不是文本数据格式，将所述第一输入问题转换为文本数据格式的第二输入问题；

[11] 处理模块，用于将所述第二输入问题输入给已训练的目标大语言模型，由目标大语言模型输出所述第二输入问题对应的文本数据格式的第一目标答案；

[12] 转换模块，用于将所述第一目标答案转换为原始数据格式的第二目标答案。

30 [13] 本申请提供一种电子设备，包括：处理器和机器可读存储介质，所述机器可读存储介质存储有能够被所述处理器执行的机器可执行指令；所述处理器用于执行机器可执行指令，以实现上述的基于大语言模型的数据处理方法。

[14] 由以上技术方案可见，本申请实施例中，可以将第一输入问题转换为文本数据格式

- 的第二输入问题，将第二输入问题输入给目标大语言模型，由目标大语言模型输出第二输入问题对应的文本数据格式的第一目标答案，并将第一目标答案转换为原始数据格式的第二目标答案。从而借助目标大语言模型得到目标答案，且通过输入文本数据格式的输入问题得到准确可靠的目标答案。只需要消耗少量计算资源就能够从大量数据中查询到输入问题对应的目标答案。获取目标答案的时间更短，从而节省服务器的计算资源，并节省时间开销。能够快速确定出输入问题对应的目标答案，提供逻辑分析推理能力，利用目标大语言模型的分析能力给出准确可靠的目标答案。而且，通过将多模态数据（如图片、视频、语音等数据）转换为文本数据输入给目标大语言模型，使得目标大语言模型支持多模态数据。目标大语言模型只需要支持文本数据的输入和输出即可，从而满足多模态数据的输入输出需求，支持多模态大语言模型的功能。

附图说明

- [15]为了更加清楚地说明本申请实施例或者现有技术中的技术方案，下面将对本申请实施例或者现有技术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申请中记载的一些实施例，对于本领域普通技术人员来讲，还可以根据本申请实施例的这些附图获得其他的附图。
- [16]图 1 是一种实施方式中的基于大语言模型的数据处理方法的流程示意图；
- [17]图 2 是一种实施方式中的目标大语言模型的结构示意图；
- [18]图 3 是一种实施方式中的基于大语言模型的数据处理方法的流程示意图；
- [19]图 4A 是一种实施方式中的 BLIP2 算法的示意图；
- [20]图 4B 是一种实施方式中的 whisper 算法的示意图；
- [21]图 4C 是一种实施方式中的 SD 算法的示意图；
- [22]图 5 是一种实施方式中的基于大语言模型的数据处理装置的结构示意图；
- [23]图 6 是一种实施方式中的电子设备的硬件结构图。

具体实施方式

- [24]在本申请实施例使用的术语仅仅是出于描述特定实施例的目的，而非限制本申请。本申请和权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其它含义。还应当理解，本文中使用的术语“和/或”是指包含一个或多个相关联的列出项目的任何或所有可能组合。
- [25]应当理解，尽管在本申请实施例可能采用术语第一、第二、第三等来描述各种信息，但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本申请范围的情况下，第一信息也可以被称为第二信息，类似地，第二信息也可以被称为第一信息。取决于语境，此外，所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。
- [26]本申请实施例中提出一种基于大语言模型的数据处理方法，该方法可以应用于任意

电子设备，如服务器等，参见图 1 所示，该方法可以包括：

[27] 步骤 101、获取原始数据格式的第一输入问题；若该原始数据格式不是文本数据格式，则将该第一输入问题转换为文本数据格式的第二输入问题。

5 [28] 步骤 102、将该第二输入问题输入给已训练的目标大语言模型，由目标大语言模型输出该第二输入问题对应的文本数据格式的第一目标答案。

[29] 步骤 103、将第一目标答案转换为原始数据格式的第二目标答案。

[30] 一个例子中，将该第一输入问题转换为文本数据格式的第二输入问题，可以包括但不限于：若原始数据格式是图片数据格式，则采用图片与文本的转换算法，将图片数据格式的第一输入问题转换为文本数据格式的第二输入问题。

10 [31] 或者，若原始数据格式是语音数据格式，则采用语音与文本的转换算法，将语音数据格式的第一输入问题转换为文本数据格式的第二输入问题。

[32] 或者，若原始数据格式是视频数据格式，则采用视频与文本的转换算法，将视频数据格式的第一输入问题转换为文本数据格式的第二输入问题。

15 [33] 一个例子中，将第一目标答案转换为原始数据格式的第二目标答案，可以包括但不限于：若原始数据格式是图片数据格式，则采用文本与图片的转换算法，将文本数据格式的第一目标答案转换为图片数据格式的第二目标答案。

[34] 或者，若原始数据格式是语音数据格式，则采用文本与语音的转换算法，将文本数据格式的第一目标答案转换为语音数据格式的第二目标答案。

20 [35] 或者，若原始数据格式是视频数据格式，则采用文本与视频的转换算法，将文本数据格式的第一目标答案转换为视频数据格式的第二目标答案。

[36] 一个例子中，将该第二输入问题输入给已训练的目标大语言模型，由目标大语言模型输出该第二输入问题对应的文本数据格式的第一目标答案，可以包括但不限于：获取与该原始数据格式对应的第一 token 标识和第二 token 标识。

25 [37] 在该第二输入问题的前面添加该第一 token 标识，在该第二输入问题的后面添加该第二 token 标识；其中，该第一 token 标识用于表示该第二输入问题的开始位置，该第二 token 标识用于表示该第二输入问题的结束位置。

[38] 将修改后的第二输入问题输入给目标大语言模型，由目标大语言模型输出该第二输入问题对应的文本数据格式的第一目标答案。

30 [39] 其中，该第一目标答案的前面具有第一 token 标识，该第一目标答案的后面具有第二 token 标识，且该第一 token 标识用于表示该第一目标答案的开始位置，该第二 token 标识用于表示该第一目标答案的结束位置。

[40] 一个例子中，将第一目标答案转换为原始数据格式的第二目标答案包括：基于目标大语言模型输出的数据内容，将位于第一 token 标识与第二 token 标识之间的数据内容确定为第一目标答案。确定与第一 token 标识或第二 token 标识对应的原始数据格式。

35 将第一目标答案转换为原始数据格式的第二目标答案。

[41]一个例子中，在得到原始数据格式的第一输入问题之后，若该原始数据格式是文本数据格式，则直接将该第一输入问题输入给已训练的目标大语言模型，由目标大语言模型输出该第一输入问题对应的文本数据格式的第一目标答案。

5 [42]一个例子中，针对目标大语言模型的训练过程，可以包括但不限于：获取样本数据集，该样本数据集可以包括多个样本问题和每个样本问题对应的样本标签，该样本问题可以是文本数据格式，该样本标签可以是文本数据格式。

[43]基于该样本数据集中的样本问题（如多个样本问题）和该样本问题对应的样本标签对待训练的初始大语言模型进行训练，得到已训练的候选大语言模型。

[44]将候选大语言模型作为目标大语言模型。或者，

10 [45]从样本数据集中选取部分样本问题作为参考样本问题，在该参考样本问题的前面添加任一数据格式的第一 token 标识，在该参考样本问题的后面添加该数据格式的第二 token 标识；基于修改后的参考样本问题和该参考样本问题对应的样本标签对候选大语言模型进行训练，得到目标大语言模型。

15 [46]一个例子中，上述执行顺序只是为了方便描述给出的示例，在实际应用中，还可以改变步骤之间的执行顺序，对此执行顺序不做限制。而且，在其它实施例中，并不一定按照本说明书示出和描述的顺序来执行相应方法的步骤，其方法所包括的步骤可以比本说明书所描述的更多或更少。此外，本说明书中所描述的单个步骤，在其它实施例中可能被分解为多个步骤进行描述。本说明书中所描述的多个步骤，在其它实施例也可能被合并为单个步骤进行描述。

20 [47]由以上技术方案可见，本申请实施例中，可以将第一输入问题转换为文本数据格式的第二输入问题，将第二输入问题输入给目标大语言模型，由目标大语言模型输出第二输入问题对应的文本数据格式的第一目标答案，并将第一目标答案转换为原始数据格式的第二目标答案。从而借助目标大语言模型得到目标答案，且通过输入文本数据格式的输入问题得到准确可靠的目标答案。只需要消耗少量计算资源就能够从大量数据中查询到输入问题对应的目标答案。获取目标答案的时间更短，从而节省服务器的计算资源，并节省时间开销。能够快速确定出输入问题对应的目标答案，提供逻辑分析推理能力，利用目标大语言模型的分析能力给出准确可靠的目标答案。而且，通过将多模态数据（如图片、视频、语音等数据）转换为文本数据输入给目标大语言模型，使得目标大语言模型支持多模态数据。目标大语言模型只需要支持文本数据的输入和输出即可，从而满足多模态数据的输入输出需求，支持多模态大语言模型的功能。

25

30

[48]以下结合具体应用场景，对本申请实施例的上述技术方案进行说明。

[49]由于网络数据（如文本数据、视频数据、语音数据、图片数据等）的数据量很大，服务器需要消耗大量计算资源才能够检索到用户所需数据，且需要花费大量时间，从而导致服务器的大量资源消耗，且用户很长时间才能够获知所需数据。比如说，实际场景会部署大量摄像机，这些摄像机采集大量图片，为了从这些图片中分析用户所需数据，需要消耗大量计算资源，花费大量时间。

35

[50]针对上述发现，可以借助大语言模型（Large Language Model, LLM）对数据进行分

析, 比如说, 将问题数据输入给大语言模型, 由大语言模型输出目标答案。这样, 只需要消耗少量计算资源就能够从大量数据中查询到问题数据对应的目标答案, 获取目标答案的时间更短, 从而节省服务器的计算资源, 节省时间开销, 利用大语言模型的分析能力给出准确可靠的目标答案。

- 5 [51] 比如说, ChatGPT 是一种基于自然语言处理和机器学习的大语言模型, 可以模拟人类的语言交流, 实现与用户的智能对话, 已经广泛应用于智能客服、聊天机器人、语音助手等领域, 为人们提供了更方便、快捷和智能化的服务。基于此, 可以借助 ChatGPT 对数据进行分析, 得到问题数据对应的目标答案。

- 10 [52] 随着人工智能的快速发展, 多模态大语言模型被提出并使用, 多模态大语言模型是指: 多模态大语言模型支持输入数据是视频数据、语音数据、图片数据和文本数据。即, 在将视频数据、语音数据、图片数据和文本数据的问题数据输入给多模态大语言模型时, 多模态大语言模型均能够输出问题数据对应的目标答案, 即多模态大语言模型能够对各种格式的问题数据进行处理。

- 15 [53] 为了实现上述功能, 需要采用多模态的数据对多模态大语言模型进行训练, 得到已训练的多模态大语言模型。即, 需要采用视频数据、语音数据、图片数据和文本数据对多模态大语言模型进行训练。为了对多模态大语言模型进行训练, 需要准备大量高质量数据, 如视频数据、语音数据、图片数据和文本数据等, 训练代价比较高。需要服务器消耗大量计算资源才能够完成多模态大语言模型的训练, 且需要花费大量时间才能够完成多模态大语言模型的训练。

- 20 [54] 在采用多模态大语言模型时, 多模态大语言模型具有编码(Encode)、解码(Decode)和适配(Linear Adapt)等过程, 输入给多模态大语言模型的是经过适配的 Embedding(嵌入)数据, 这种适配在输入侧和输出侧均需要执行。基于此, 需要对多模态大语言模型和 Linear Projector(或称为 Adaptor, 如全连接层等)做对齐训练, 这个对齐训练工作需要准备大量高质量数据, 训练代价高。

- 25 [55] 针对上述发现, 本申请提供一种基于大语言模型的数据处理方法, 通过将多模态数据(如图片、视频、语音等数据)转换为文本数据输入给目标大语言模型, 使得目标大语言模型支持多模态数据。目标大语言模型只需要支持文本数据的输入和输出即可, 从而满足多模态数据的输入输出需求, 支持多模态功能。比如说, 设计统一的多模态大模型架构下的目标大语言模型, 目标大语言模型的输入和输出都是文本形式, 从而避免输入侧及输入侧的对齐训练。
- 30

[56] 参见图 2 所示, 为目标大语言模型的结构示意图, 目标大语言模型的输入数据是文本数据格式的输入数据, 目标大语言模型的输出数据是文本数据格式的输入数据。比如说, 将文本数据格式的输入数据提供给目标大语言模型, 由目标大语言模型对输入数据进行处理后, 输出文本数据格式的输入数据。

- 35 [57] 参见图 2 所示, 若获取到图片数据格式的输入数据, 则将图片数据格式的输入数据转换为文本数据格式的输入数据。将文本数据格式的输入数据提供给目标大语言模型, 由目标大语言模型输出文本数据格式的输入数据。然后, 将文本数据格式的输入数据转换为图片数据格式的输入数据。

[58] 参见图 2 所示, 若获取到语音数据格式的输入数据, 则将语音数据格式的输入数据转换为文本数据格式的输入数据。将文本数据格式的输入数据提供给目标大语言模型, 由目标大语言模型输出文本数据格式的输入数据。然后, 将文本数据格式的输入数据转换为语音数据格式的输入数据。

- 5 [59] 参见图 2 所示, 若获取到视频数据格式的输入数据, 则将视频数据格式的输入数据转换为文本数据格式的输入数据。将文本数据格式的输入数据提供给目标大语言模型, 由目标大语言模型输出文本数据格式的输入数据。然后, 将文本数据格式的输入数据转换为视频数据格式的输入数据。

- 10 [60] 参见图 2 所示, 若获取到文本数据格式的输入数据, 则直接将文本数据格式的输入数据提供给目标大语言模型, 由目标大语言模型输出文本数据格式的输入数据, 这样, 可以直接得到文本数据格式的输入数据。

[61] 基于已训练的目标大语言模型, 本申请实施例中提出一种基于大语言模型的数据处理方法, 参见图 3 所示, 为该方法的流程图, 该方法可以包括:

[62] 步骤 301、获取原始数据格式的第一输入问题。

- 15 [63] 比如说, 可以接收用户输入的原始数据格式的问题, 将该问题记为第一输入问题, 可以采用其它方式获取原始数据格式的第一输入问题, 对此不做限制。

[64] 步骤 302、确定第一输入问题的原始数据格式。若原始数据格式是图片数据格式, 则执行步骤 303。若原始数据格式是语音数据格式, 则执行步骤 304。若原始数据格式是视频数据格式, 则执行步骤 305。若原始数据格式是文本数据格式, 则执行步骤 306。

- 20 当然, 上述只是原始数据格式的几个示例, 对此不做限制。

- [65] 一个例子中, 若第一输入问题是图片数据格式的输入问题, 则确定原始数据格式是图片数据格式。或者, 若第一输入问题是语音数据格式的输入问题, 则确定原始数据格式是语音数据格式。或者, 若第一输入问题是视频数据格式的输入问题, 则确定原始数据格式是视频数据格式。或者, 若第一输入问题是文本数据格式的输入问题, 则确定原始数据格式是文本数据格式。
- 25

[66] 步骤 303、若原始数据格式是图片数据格式 (即 image 数据格式), 则采用图片与文本的转换算法, 将图片数据格式的第一输入问题转换为文本数据格式的第二输入问题。其中, 在步骤 303 之后, 还可以执行步骤 306。

- [67] 一个例子中, 图片与文本的转换算法可以包括但不限于 BLIP (Bootstrapping Language-Image Pre-training, 引导语言图像预训练) 算法、BLIP2 算法, 可以通过 BLIP 算法或 BLIP2 算法将图片数据格式的第一输入问题转换为文本数据格式的第二输入问题。当然, BLIP 算法或 BLIP2 算法只是示例, 对此不做限制, 能够将图片数据格式的第一输入问题转换为文本数据格式的第二输入问题即可。
- 30

- [68] 比如说, BLIP2 算法的一个示例可以参见图 4A 所示, BLIP2 算法可以包括 Image Encoder (图像编码器) 和 Q-Former (Q 模型), Q-Former 可以包括第一子网络 (多个) 和第二子网络 (多个)。第一子网络可以包括 Self Attention (自我注意力) 网络层、Cross Attention (交叉注意力) 网络层和 Feed Forward (前馈) 网络层。第二子网络可以包括
- 35

Self Attention 网络层和 Feed Forward 网络层。

[69]在训练过程中，可以将样本图像提供给 Image Encoder 得到多个图像特征，将多个图像特征和 Learned Queries（查询向量）作为第一子网络的输入特征，由第一子网络输出图像文本匹配特征。可以将样本图像对应的文本描述特征作为第二子网络的输入特征，
5 由第二子网络输出基于图像的文本生成特征。然后，基于图像文本匹配特征和基于图像的文本生成特征对 Q-Former 进行训练。

[70]在使用过程中，可以将待检测图像（即图片数据格式的第一输入问题，即第一输入问题是图像）提供给 Image Encoder 得到多个图像特征，将多个图像特征和 Learned Queries 作为第一子网络的输入特征。由第一子网络输出文本数据格式的第二输入问题，
10 即第二输入问题是采用文本对第一输入问题进行描述。

[71]至此，可以采用 BLIP2 算法将图片数据格式的第一输入问题转换为文本数据格式的第二输入问题，第二输入问题是采用文本对第一输入问题进行描述。

[72]一个例子中，BLIP2 算法的 Q-Former 可以支持多种模式，如 LM 模式和 ITM 模式等。若选择 Q-Former 的 LM 模式，则 Q-Former 可以输出文本数据。若选择 Q-Former
15 的 ITM 模式，则 Q-Former 可以输出 Embedding 数据。

[73]基于此，为了输出文本数据，则可以选择 Q-Former 的 LM 模式，在 LM 模式下，可以将图片数据格式的第一输入问题提供给 Image Encoder 得到多个图像特征，将多个图像特征和 Learned Queries 作为第一子网络的输入特征，由第一子网络输出文本数据格式的第二输入问题，对此过程不再重复赘述。

20 [74]步骤 304、若原始数据格式是语音数据格式（即 speech 数据格式），则采用语音与文本的转换算法，将语音数据格式的第一输入问题转换为文本数据格式的第二输入问题。其中，在步骤 304 之后，还可以执行步骤 306。

[75]一个例子中，语音与文本的转换算法可以包括但不限于 whisper 算法，可以通过 whisper 算法将语音数据格式的第一输入问题转换为文本数据格式的第二输入问题。当然，whisper 算法只是一个示例，对此不做限制，只要能够将语音数据格式的第一输入
25 问题转换为文本数据格式的第二输入问题即可。

[76]比如说，whisper 算法的一个示例可以参见图 4B 所示，whisper 算法可以包括第一子网络（如第一子网络包括 2 个卷积层和 1 个 GELU 层）、第二子网络（如第二子网络包括多个 Encoder Block）、第三子网络（如第三子网络包括多个 Decoder Block），且第二子网络与第三子网络之间为 Cross Attention 网络层。
30

[77]可以将待检测语音（即语音数据格式的第一输入问题，第一输入问题是语音）提供给 whisper 算法，由 whisper 算法对待检测语音进行处理，输出文本数据格式的第二输入问题，即第二输入问题是采用文本对第一输入问题进行描述。

[78]至此，可以采用 whisper 算法将语音数据格式的第一输入问题转换为文本数据格式的第二输入问题，第二输入问题是采用文本对第一输入问题进行描述。
35

[79]步骤 305、若原始数据格式是视频数据格式（即 video 数据格式），则采用视频与文

本的转换算法，将视频数据格式的第一输入问题转换为文本数据格式的第二输入问题。其中，在步骤 305 之后，还可以执行步骤 306。

5 [80] 一个例子中，可以采用视频与文本的转换算法将视频数据格式的第一输入问题（即第一输入问题是视频）转换为文本数据格式的第二输入问题，即第二输入问题是采用文本对第一输入问题进行描述。对此转换算法不做限制，能够将视频数据格式的第一输入问题转换为文本数据格式的第二输入问题即可。

10 [81] 一个例子中，可以将视频数据格式的第一输入问题（即视频）分解成多个图像，对此分解过程不做限制。针对每个图像，可以采用图片与文本的转换算法，将该图像转换为文本数据格式的文本，如 BLIP 算法、BLIP2 算法等。然后，将多个图像对应的多个文本组合起来，得到文本数据格式的第二输入问题。

[82] 至此，可以将视频数据格式的第一输入问题转换为文本数据格式的第二输入问题，且第二输入问题是采用文本对第一输入问题进行描述。

[83] 步骤 306、将文本数据格式的输入问题输入给目标大语言模型。

15 [84] 一个例子中，针对图片数据格式的第一输入问题，在将图片数据格式的第一输入问题转换为文本数据格式的第二输入问题之后，可以获取与图片数据格式对应的第一 token 标识和第二 token 标识。第一 token 标识用于表示第二输入问题的开始位置，第二 token 标识用于表示第二输入问题的结束位置。

20 [85] 在第二输入问题的前面添加第一 token 标识，在第二输入问题的后面添加第二 token 标识，得到修改后的文本数据格式的第二输入问题。在得到修改后的第二输入问题之后，将修改后的第二输入问题输入给目标大语言模型。

[86] 比如说，图片数据格式对应的第一 token 标识可以为，图片数据格式对应的第二 token 标识可以为，则修改后的第二输入问题可以是：AAA，表示第二输入问题的开始位置，AAA 表示文本数据格式的第二输入问题，表示第二输入问题的结束位置。

25 [87] 一个例子中，针对语音数据格式的第一输入问题，在将语音数据格式的第一输入问题转换为文本数据格式的第二输入问题之后，可以获取与语音数据格式对应的第一 token 标识和第二 token 标识。第一 token 标识用于表示第二输入问题的开始位置，第二 token 标识用于表示第二输入问题的结束位置。

30 [88] 在第二输入问题的前面添加第一 token 标识，在第二输入问题的后面添加第二 token 标识，得到修改后的文本数据格式的第二输入问题。在得到修改后的第二输入问题之后，将修改后的第二输入问题输入给目标大语言模型。

35 [89] 比如说，语音数据格式对应的第一 token 标识可以为<SPEECH>，语音数据格式对应的第二 token 标识可以为</SPEECH>，则修改后的第二输入问题可以是：<SPEECH>AAA</SPEECH>，<SPEECH>表示第二输入问题的开始位置，AAA 表示文本数据格式的第二输入问题，</SPEECH>表示第二输入问题的结束位置。

[90] 一个例子中，针对视频数据格式的第一输入问题，在将视频数据格式的第一输入问

题转换为文本数据格式的第二输入问题之后，可以获取与视频数据格式对应的第一 token 标识和第二 token 标识。第一 token 标识用于表示第二输入问题的开始位置，第二 token 标识用于表示第二输入问题的结束位置。

5 [91]在第二输入问题的前面添加第一 token 标识，在第二输入问题的后面添加第二 token 标识，得到修改后的文本数据格式的第二输入问题。在得到修改后的第二输入问题之后，将修改后的第二输入问题输入给目标大语言模型。

10 [92]比如说，视频数据格式对应的第一 token 标识可以为<VD>，视频数据格式对应的第二 token 标识可以为<\VD>，则修改后的第二输入问题可以是：<VD>AAA<\VD>，<VD>表示第二输入问题的开始位置，AAA 表示文本数据格式的第二输入问题，<\VD>表示第二输入问题的结束位置。

[93]一个例子中，针对文本数据格式的第一输入问题，可以直接将文本数据格式的第一输入问题输入给目标大语言模型，对此过程不做限制。

[94]步骤 307、通过目标大语言模型输出文本数据格式的第一目标答案。

15 [95]一个例子中，若输入给目标大语言模型的是AAA<\IMG>，则目标大语言模型可以将与<\IMG>之间的数据内容作为输入问题（即第二输入问题）。然后，目标大语言模型经过分析推理得到该输入问题对应的文本数据格式的第一目标答案，并输出文本数据格式的第一目标答案。

20 [96]在目标大语言模型输出第一目标答案时，第一目标答案的前面具有第一 token 标识，第一目标答案的后面具有第二 token 标识，且第一 token 标识表示第一目标答案的开始位置，第二 token 标识表示第一目标答案的结束位置。比如说，目标大语言模型输出BBB<\IMG>，表示第一目标答案的开始位置，<\IMG>表示第一目标答案的结束位置，BBB 表示文本数据格式的第一目标答案。

25 [97]一个例子中，若输入给目标大语言模型的是<SPEECH>AAA<\SPEECH>，则目标大语言模型可以将<SPEECH>与<\SPEECH>之间的数据内容作为输入问题（即第二输入问题）。然后，目标大语言模型经过分析推理得到该输入问题对应的文本数据格式的第一目标答案，并输出文本数据格式的第一目标答案。

30 [98]在目标大语言模型输出第一目标答案时，第一目标答案的前面具有第一 token 标识，第一目标答案的后面具有第二 token 标识，第一 token 标识表示第一目标答案的开始位置，第二 token 标识表示第一目标答案的结束位置。如目标大语言模型输出<SPEECH>BBB<\SPEECH>，<SPEECH>表示第一目标答案的开始位置，<\SPEECH>表示第一目标答案的结束位置，BBB 表示第一目标答案。

35 [99]一个例子中，若输入给目标大语言模型的是<VD>AAA<\VD>，则目标大语言模型可以将<VD>与<\VD>之间的数据内容作为输入问题（即第二输入问题）。然后，目标大语言模型经过分析推理得到该输入问题对应的文本数据格式的第一目标答案，并输出文本数据格式的第一目标答案。

[100] 在目标大语言模型输出第一目标答案时，第一目标答案的前面具有第一 token 标识，第一目标答案的后面具有第二 token 标识，且第一 token 标识表示第一目标答案

的开始位置，第二 token 标识表示第一目标答案的结束位置。比如说，目标大语言模型输出<VD>BBB<\VD>，<VD>表示第一目标答案的开始位置，<\VD>表示第一目标答案的结束位置，BBB 表示文本数据格式的第一目标答案。

5 [101] 一个例子中，若输入给目标大语言模型的是 AAA，则目标大语言模型可以将 AAA 作为输入问题（即第一输入问题）。目标大语言模型经过分析推理得到该输入问题对应的文本数据格式的第一目标答案，并输出第一目标答案。

10 [102] 一个例子中，目标大语言模型在得到文本数据格式的输入问题（如第一输入问题或第二输入问题）时，可以对输入问题进行 Tokenizer（分词器）转换得到 Embedding 数据，基于 Embedding 数据进行分析推理得到推理结果，对推理结果进行 DeTokenizer（反分词器）转换得到文本数据格式的第一目标答案。

[103] 步骤 308、将第一目标答案转换为原始数据格式的第二目标答案。

15 [104] 一个例子中，若目标大语言模型输出的是BBB，则基于目标大语言模型输出的数据内容，可以将位于第一 token 标识与第二 token 标识之间的数据内容（BBB）确定为第一目标答案。然后，确定与第一 token 标识或者第二 token 标识对应的原始数据格式，即或者对应图片数据格式。然后，可以将第一目标答案转换为图片数据格式的第二目标答案。

[105] 比如说，由于原始数据格式是图片数据格式，因此，采用文本与图片的转换算法，将文本数据格式的第一目标答案转换为图片数据格式的第二目标答案。

20 [106] 比如说，文本与图片的转换算法可以包括但不限于 SD（Stable Diffusion，稳定扩散）算法，可以通过 SD 算法将文本数据格式的第一目标答案转换为图片数据格式的第二目标答案。当然，SD 算法只是一个示例，对此不做限制，能够将文本数据格式的第一目标答案转换为图片数据格式的第二目标答案即可。

[107] 比如说，SD 算法的一个示例可以参见图 4C 所示，SD 算法可以包括 Text Model（文本模型）、UNet 网络、autoencoder（自动编码器）等。

25 [108] 在使用过程中，可以将待检测文本（即文本数据格式的第一目标答案，即第一目标答案是文本）提供给 Text Model 得到 text embeddings（文本嵌入特征）。将 text embeddings 输入给 UNet 网络和 autoencoder，可以得到图片数据格式的第二目标答案，即第二目标答案是采用图片对第一目标答案进行描述。

30 [109] 一个例子中，若目标大语言模型输出的是<SPEECH>BBB<\SPEECH>，则基于目标大语言模型输出的数据内容，可以将位于第一 token 标识与第二 token 标识之间的数据内容（BBB）确定为第一目标答案。然后，确定与第一 token 标识或第二 token 标识对应的原始数据格式，即<SPEECH>或<\SPEECH>对应语音数据格式。然后，可以将第一目标答案转换为语音数据格式的第二目标答案。

35 [110] 比如说，由于原始数据格式是语音数据格式，因此，采用文本与语音的转换算法，将文本数据格式的第一目标答案转换为语音数据格式的第二目标答案。

[111] 比如说，文本与语音的转换算法（也可以称为语音合成算法）可以包括但不限

于 TTS (Text-To-Speech, 文本转语音) 算法、Azure TTS 算法、Amazon Polly 算法等, 可以通过 TTS 算法、Azure TTS 算法、Amazon Polly 算法将文本数据格式的第一目标答案转换为语音数据格式的第二目标答案。当然, TTS 算法、Azure TTS 算法、Amazon Polly 算法只是示例, 对此不做限制, 能够将文本数据格式的第一目标答案转换为语音数据格式的第二目标答案即可。

[112] 一个例子中, 若目标大语言模型输出的是<VD>BBB<\VD>, 则基于目标大语言模型输出的数据内容, 可以将位于第一 token 标识与第二 token 标识之间的数据内容 (BBB) 确定为第一目标答案。然后, 可以确定与第一 token 标识或者第二 token 标识对应的原始数据格式, 即<VD>或<\VD>对应视频数据格式。然后, 可以将第一目标答案转换为视频数据格式的第二目标答案。

[113] 比如说, 由于原始数据格式是视频数据格式, 因此, 采用文本与视频的转换算法, 将文本数据格式的第一目标答案转换为视频数据格式的第二目标答案。

[114] 比如说, 可以采用文本与视频的转换算法将文本数据格式的第一目标答案转换为视频数据格式的第二目标答案, 对此转换过程不做限制。或者, 可以采用文本与图片的转换算法将文本数据格式的第一目标答案转换为多帧图像, 然后, 可以将多帧图像组合成视频数据格式的第二目标答案。

[115] 至此, 可以得到原始数据格式的第二目标答案, 完成分析推理过程。

[116] 一个例子中, 针对目标大语言模型的训练过程, 可以包括但不限于:

[117] 获取样本数据集, 该样本数据集可以包括多个样本问题和每个样本问题对应的样本标签 (即样本问题对应的样本答案), 该样本问题可以是文本数据格式, 该样本标签可以是文本数据格式。基于该样本数据集中的样本问题 (如多个样本问题) 和该样本问题对应的样本标签对待训练的初始大语言模型进行训练, 得到已训练的候选大语言模型, 本实施例对此训练过程不做限制。

[118] 在得到候选大语言模型后, 可以将候选大语言模型作为目标大语言模型。

[119] 或者, 在得到候选大语言模型后, 还可以对候选大语言模型进行微调, 以使候选大语言模型支持 token 标识, 得到微调后的目标大语言模型。

[120] 针对候选大语言模型的微调过程, 可以从样本数据集中选取部分样本问题作为参考样本问题。针对每个参考样本问题, 在该参考样本问题的前面添加任一数据格式 (如图片数据格式、语音数据格式、视频数据格式) 的第一 token 标识, 在该参考样本问题的后面添加该数据格式的第二 token 标识。

[121] 比如说, 针对部分参考样本问题, 在这些参考样本问题的前面添加, 在这些参考样本问题的后面添加<\IMG>。针对部分参考样本问题, 在这些参考样本问题的前面添加<SPEECH>, 在这些参考样本问题的后面添加<\SPEECH>。针对部分参考样本问题, 在这些参考样本问题的前面添加<VD>, 在这些参考样本问题的后面添加<\VD>。基于上述处理, 可以得到修改后的参考样本问题。

[122] 基于修改后的参考样本问题和该参考样本问题对应的样本标签对候选大语言模

型进行训练，得到目标大语言模型，本实施例对此训练过程不做限制。

[123] 一个例子中，在对初始大语言模型进行训练得到候选大语言模型时，初始大语言模型的输入和输出均为文本数据，省去输入侧和输出侧的 Adaptor 模块（用于对齐的线性映射层），避免繁琐的输入侧对齐和输出侧对齐训练。

5 [124] 在训练得到候选大语言模型之后，可以对候选大语言模型进行微调训练即可得到目标大语言模型，目标大语言模型支持 token 标识即可。

10 [125] 由以上技术方案可见，本申请实施例中，借助目标大语言模型得到目标答案，且通过输入文本数据格式的输入问题得到准确可靠的目标答案。只需要消耗少量计算资源就能够从大量数据中查询到输入问题对应的目标答案。获取目标答案的时间更短，从而节省服务器的计算资源，并节省时间开销。能够快速确定出输入问题对应的目标答案，提供逻辑分析推理能力，利用目标大语言模型的分析能力给出准确可靠的目标答案。而且，通过将多模态数据（如图片、视频、语音等数据）转换为文本数据输入给目标大语言模型，使得目标大语言模型支持多模态数据。目标大语言模型只需要支持文本数据的输入和输出即可，从而满足多模态数据的输入输出需求，支持多模态大语言模型的功能。

15 [126] 目标大语言模型直接处理文本数据，取消输入与输出侧的线性映射层，不需要输入侧与目标大语言模型、输出侧与目标大语言模型的对齐操作，无需繁琐的对齐训练，可以大大简化目标大语言模型的训练过程，满足更多业务需求。

20 [127] 基于同一发明构思，还提供与上述基于大语言模型的数据处理方法对应的基于大语言模型的数据处理装置及电子设备，由于数据处理装置及电子设备解决问题的原理与基于大语言模型的数据处理方法相似，因此，数据处理装置及电子设备的实施可以参见基于大语言模型的数据处理方法，重复之处不再赘述。

[128] 基于与上述方法同样的申请构思，本申请实施例中提出一种基于大语言模型的数据处理装置，参见图 5 所示，为该装置的结构示意图，该装置可以包括：

25 [129] 获取模块 51，用于获取原始数据格式的第一输入问题；若原始数据格式不是文本数据格式，将所述第一输入问题转换为文本数据格式的第二输入问题；

[130] 处理模块 52，用于将所述第二输入问题输入给已训练的目标大语言模型，由目标大语言模型输出所述第二输入问题对应的文本数据格式的第一目标答案；

[131] 转换模块 53，用于将第一目标答案转换为原始数据格式的第二目标答案。

30 [132] 一个例子中，所述获取模块 51 将所述第一输入问题转换为文本数据格式的第二输入问题时具体用于：若所述原始数据格式是图片数据格式，则采用图片与文本的转换算法，将图片数据格式的第一输入问题转换为文本数据格式的第二输入问题。或者，若所述原始数据格式是语音数据格式，则采用语音与文本的转换算法，将语音数据格式的第一输入问题转换为文本数据格式的第二输入问题。或者，若所述原始数据格式是视频数据格式，则采用视频与文本的转换算法，将视频数据格式的第一输入问题转换为文本数据格式的第二输入问题。

35 [133] 一个例子中，所述转换模块 53 将所述第一目标答案转换为原始数据格式的第二

目标答案时具体用于：若所述原始数据格式是图片数据格式，则采用文本与图片的转换算法，将文本数据格式的第一目标答案转换为图片数据格式的第二目标答案。或者，若所述原始数据格式是语音数据格式，则采用文本与语音的转换算法，将文本数据格式的第一目标答案转换为语音数据格式的第二目标答案。或者，若所述原始数据格式是视频数据格式，则采用文本与视频的转换算法，将文本数据格式的第一目标答案转换为视频数据格式的第二目标答案。

[134] 一个例子中，所述处理模块 52 将所述第二输入问题输入给已训练的目标大语言模型，由目标大语言模型输出所述第二输入问题对应的文本数据格式的第一目标答案时具体用于：获取与所述原始数据格式对应的第一 token 标识和第二 token 标识。在所述第二输入问题的前面添加所述第一 token 标识，在所述第二输入问题的后面添加所述第二 token 标识。其中，所述第一 token 标识表示所述第二输入问题的开始位置，所述第二 token 标识表示所述第二输入问题的结束位置。将修改后的第二输入问题输入给所述目标大语言模型，由所述目标大语言模型输出第二输入问题对应的文本数据格式的第一目标答案。

[135] 其中，所述第一目标答案的前面具有所述第一 token 标识，所述第一目标答案的后面具有所述第二 token 标识，所述第一 token 标识表示所述第一目标答案的开始位置，所述第二 token 标识表示所述第一目标答案的结束位置。

[136] 一个例子中，所述转换模块 53 将所述第一目标答案转换为原始数据格式的第二目标答案时具体用于：基于所述目标大语言模型输出的数据内容，将位于所述第一 token 标识与所述第二 token 标识之间的数据内容确定为所述第一目标答案。确定与所述第一 token 标识或所述第二 token 标识对应的所述原始数据格式。将所述第一目标答案转换为所述原始数据格式的第二目标答案。

[137] 一个例子中，所述装置还包括：训练模块，用于训练得到所述目标大语言模型。其中，所述训练模块训练得到所述目标大语言模型时具体用于：

[138] 获取样本数据集，所述样本数据集包括多个样本问题和每个样本问题对应的样本标签，所述样本问题是文本数据格式，所述样本标签是文本数据格式；

[139] 基于所述样本数据集中的样本问题和该样本问题对应的样本标签对待训练的目标大语言模型进行训练，得到已训练的候选大语言模型；

[140] 将所述候选大语言模型作为所述目标大语言模型；或，

[141] 从所述样本数据集中选取部分样本问题作为参考样本问题，在参考样本问题的前面添加任一数据格式的第一 token 标识，在参考样本问题的后面添加该数据格式的第二 token 标识；基于修改后的参考样本问题和该参考样本问题对应的样本标签对所述候选大语言模型进行训练，得到所述目标大语言模型。

[142] 基于与上述方法同样的构思，本申请一个例子中提出一种电子设备，参见图 6 所示，电子设备包括处理器 611 和机器可读存储介质 612，机器可读存储介质 612 存储有能够被处理器 611 执行的机器可执行指令；处理器 611 用于执行机器可执行指令，以实现上述示例公开的基于大语言模型的数据处理方法。

[143] 一个例子中, 处理器 611 可以包括一个或多个处理核心, 如 4 核心处理器、8 核心处理器等。处理器 611 可以采用 DSP (Digital Signal Processing, 数字信号处理)、FPGA (Field - Programmable Gate Array, 现场可编程门阵列)、PLA (Programmable Logic Array, 可编程逻辑阵列) 中的至少一种硬件形式来实现。

- 5 [144] 处理器 611 也可以包括主处理器和协处理器, 主处理器是用于对在唤醒状态下的数据进行处理的处理单元, 也称 CPU (Central Processing Unit, 中央处理器)。协处理器是用于对在待机状态下的数据进行处理的低功耗处理器。

[145] 在一些实施例中, 处理器 611 可以集成有 GPU (Graphics Processing Unit, 图像处理单元), GPU 用于负责显示屏所需要显示的内容的渲染和绘制。

- 10 [146] 一个例子中, 电子设备还可选包括有: 外围设备接口 613 和至少一个外围设备。处理器 611 和外围设备接口 613 之间可以通过总线或信号线相连。各个外围设备可以通过总线、信号线或电路板与外围设备接口 613 相连。外围设备可以包括: 射频电路 614 和电源 615 中的至少一种。

- [147] 射频电路 614 用于接收和发射 RF (Radio Frequency, 射频) 信号, 也称电磁信号。射频电路 614 通过电磁信号与通信网络以及其它通信设备进行通信。射频电路 614 将电信号转换为电磁信号进行发送, 或者, 将接收到的电磁信号转换为电信号。可选地, 射频电路 614 包括: 天线系统、RF 收发器、一个或多个放大器、调谐器、振荡器、数字信号处理器、用户身份模块卡等等。
- 15

- [148] 射频电路 614 可以通过至少一种无线通信协议来与用户设备进行通信。该无线通信协议包括但不限于: 万维网、城域网、内联网、各代移动通信网络 (2G、3G、4G 及 5G)、无线局域网和/或 WiFi (Wireless Fidelity, 无线保真) 网络。
- 20

[149] 电源 615 用于为电子设备中的各个组件进行供电, 电源 615 可以是交流电、直流电、一次性电池或可充电电池。

- [150] 基于与上述方法同样的申请构思, 本申请实施例还提供一种机器可读存储介质, 所述机器可读存储介质上存储有若干计算机指令, 所述计算机指令被处理器执行时, 能够实现本申请上述示例公开的基于大语言模型的数据处理方法。
- 25

- [151] 其中, 上述机器可读存储介质可以是任何电子、磁性、光学或其它物理存储装置, 可以包含或存储信息, 如可执行指令、数据, 等等。例如, 机器可读存储介质可以是: RAM (Random Access Memory, 随机存取存储器)、易失存储器、非易失性存储器、闪存、存储驱动器 (如硬盘驱动器)、固态硬盘、任何类型的存储盘 (如光盘、dvd 等), 或者类似的存储介质, 或者它们的组合。
- 30

- [152] 上述实施例阐明的系统、装置、模块或单元, 具体可以由计算机实体实现, 或者由具有某种功能的产品来实现。一种典型的实现设备为计算机, 计算机的具体形式可以是个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件收发设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任意几种设备的组合。
- 35

[153] 为了描述的方便, 描述以上装置时以功能分为各种单元分别描述。当然, 在实

施本申请时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[154] 本领域内的技术人员应明白，本申请的实施例可提供为方法、系统、或计算机程序产品。因此，本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[155] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可以由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其它可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其它可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[156] 而且，这些计算机程序指令也可以存储在能引导计算机或其它可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或者多个流程和/或方框图一个方框或者多个方框中指定的功能。

[157] 这些计算机程序指令也可装载到计算机或其它可编程数据处理设备上，使得在计算机或者其它可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其它可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[158] 以上所述仅为本申请的实施例而已，并不用于限制本申请。对于本领域技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等，均应包含在本申请的权利要求范围之内。

权利要求书

- 1、一种基于大语言模型的数据处理方法，其特征在于，所述方法包括：
获取原始数据格式的第一输入问题；若所述原始数据格式不是文本数据格式，则将所述第一输入问题转换为文本数据格式的第二输入问题；
- 5 将所述第二输入问题输入给已训练的目标大语言模型，由所述目标大语言模型输出所述第二输入问题对应的文本数据格式的第一目标答案；
将所述第一目标答案转换为所述原始数据格式的第二目标答案。
- 2、根据权利要求1所述的方法，其特征在于，
所述将所述第一输入问题转换为文本数据格式的第二输入问题，包括：
- 10 若所述原始数据格式是图片数据格式，则采用图片与文本的转换算法，将图片数据格式的第一输入问题转换为文本数据格式的第二输入问题；或者，
若所述原始数据格式是语音数据格式，则采用语音与文本的转换算法，将语音数据格式的第一输入问题转换为文本数据格式的第二输入问题；或者，
若所述原始数据格式是视频数据格式，则采用视频与文本的转换算法，将视频数据
- 15 格式的第一输入问题转换为文本数据格式的第二输入问题。
- 3、根据权利要求1或2所述的方法，其特征在于，
所述将所述第一目标答案转换为所述原始数据格式的第二目标答案，包括：
若所述原始数据格式是图片数据格式，则采用文本与图片的转换算法，将文本数据格式的第一目标答案转换为图片数据格式的第二目标答案；或者，
- 20 若所述原始数据格式是语音数据格式，则采用文本与语音的转换算法，将文本数据格式的第一目标答案转换为语音数据格式的第二目标答案；或者，
若所述原始数据格式是视频数据格式，则采用文本与视频的转换算法，将文本数据格式的第一目标答案转换为视频数据格式的第二目标答案。
- 4、根据权利要求1所述的方法，其特征在于，
- 25 所述将所述第二输入问题输入给已训练的目标大语言模型，由所述目标大语言模型输出所述第二输入问题对应的文本数据格式的第一目标答案，包括：
获取与所述原始数据格式对应的第一 token 标识和第二 token 标识；
在所述第二输入问题的前面添加所述第一 token 标识，在所述第二输入问题的后面添加所述第二 token 标识；其中，所述第一 token 标识表示所述第二输入问题的开始位置，
- 30 所述第二 token 标识表示所述第二输入问题的结束位置；
将修改后的第二输入问题输入给所述目标大语言模型，由所述目标大语言模型输出第二输入问题对应的文本数据格式的第一目标答案；
其中，所述第一目标答案的前面具有所述第一 token 标识，所述第一目标答案的后面具有所述第二 token 标识，所述第一 token 标识表示所述第一目标答案的开始位置，
- 35 所述第二 token 标识表示所述第一目标答案的结束位置。
- 5、根据权利要求4所述的方法，其特征在于，
所述将所述第一目标答案转换为所述原始数据格式的第二目标答案，包括：
基于所述目标大语言模型输出的数据内容，将位于所述第一 token 标识与所述第二 token 标识之间的数据内容确定为所述第一目标答案；
- 40 确定与所述第一 token 标识或所述第二 token 标识对应的所述原始数据格式；

将所述第一目标答案转换为所述原始数据格式的第二目标答案。

6、根据权利要求1所述的方法，其特征在于，

针对所述目标大语言模型的训练过程，包括：

5 获取样本数据集，所述样本数据集包括多个样本问题和每个样本问题对应的样本标签，所述样本问题是文本数据格式，所述样本标签是文本数据格式；

基于所述样本数据集中的样本问题和该样本问题对应的样本标签对待训练的初始大语言模型进行训练，得到已训练的候选大语言模型；

将所述候选大语言模型作为所述目标大语言模型；或者，

10 从所述样本数据集中选取部分样本问题作为参考样本问题，在参考样本问题的前面添加任一数据格式的第一 token 标识，在参考样本问题的后面添加该数据格式的第二 token 标识；基于修改后的参考样本问题和该参考样本问题对应的样本标签对所述候选大语言模型进行训练，得到所述目标大语言模型。

7、一种基于大语言模型的数据处理装置，其特征在于，所述装置包括：

15 获取模块，用于获取原始数据格式的第一输入问题；若所述原始数据格式不是文本数据格式，将所述第一输入问题转换为文本数据格式的第二输入问题；

处理模块，用于将所述第二输入问题输入给已训练的目标大语言模型，由目标大语言模型输出所述第二输入问题对应的文本数据格式的第一目标答案；

转换模块，用于将所述第一目标答案转换为原始数据格式的第二目标答案。

20 8、根据权利要求7所述的装置，其特征在于，所述获取模块将所述第一输入问题转换为文本数据格式的第二输入问题时具体用于：

若所述原始数据格式是图片数据格式，则采用图片与文本的转换算法，将图片数据格式的第一输入问题转换为文本数据格式的第二输入问题；或者，

若所述原始数据格式是语音数据格式，则采用语音与文本的转换算法，将语音数据格式的第一输入问题转换为文本数据格式的第二输入问题；或者，

25 若所述原始数据格式是视频数据格式，则采用视频与文本的转换算法，将视频数据格式的第一输入问题转换为文本数据格式的第二输入问题。

9、根据权利要求7或8所述的装置，其特征在于，所述转换模块将所述第一目标答案转换为原始数据格式的第二目标答案时具体用于：

30 若所述原始数据格式是图片数据格式，则采用文本与图片的转换算法，将文本数据格式的第一目标答案转换为图片数据格式的第二目标答案；或者，

若所述原始数据格式是语音数据格式，则采用文本与语音的转换算法，将文本数据格式的第一目标答案转换为语音数据格式的第二目标答案；或者，

若所述原始数据格式是视频数据格式，则采用文本与视频的转换算法，将文本数据格式的第一目标答案转换为视频数据格式的第二目标答案。

35 10、根据权利要求7所述的装置，其特征在于，所述处理模块将所述第二输入问题输入给已训练的目标大语言模型，由目标大语言模型输出所述第二输入问题对应的文本数据格式的第一目标答案时具体用于：

获取与所述原始数据格式对应的第一 token 标识和第二 token 标识；

40 在所述第二输入问题的前面添加所述第一 token 标识，在所述第二输入问题的后面添加所述第二 token 标识；其中，所述第一 token 标识表示所述第二输入问题的开始位

置, 所述第二 token 标识表示所述第二输入问题的结束位置;

将修改后的第二输入问题输入给所述目标大语言模型, 由所述目标大语言模型输出第二输入问题对应的文本数据格式的第一目标答案;

- 5 其中, 所述第一目标答案的前面具有所述第一 token 标识, 所述第一目标答案的后面具有所述第二 token 标识, 所述第一 token 标识表示所述第一目标答案的开始位置, 所述第二 token 标识表示所述第一目标答案的结束位置。

11、根据权利要求 10 所述的装置, 其特征在于, 所述转换模块将所述第一目标答案转换为原始数据格式的第二目标答案时具体用于:

- 10 基于所述目标大语言模型输出的数据内容, 将位于所述第一 token 标识与所述第二 token 标识之间的数据内容确定为所述第一目标答案;

确定与所述第一 token 标识或所述第二 token 标识对应的所述原始数据格式;

将所述第一目标答案转换为所述原始数据格式的第二目标答案。

12、根据权利要求 7 所述的装置, 其特征在于, 还包括:

训练模块, 用于训练得到所述目标大语言模型;

- 15 其中, 所述训练模块训练得到所述目标大语言模型时具体用于:

获取样本数据集, 所述样本数据集包括多个样本问题和每个样本问题对应的样本标签, 所述样本问题是文本数据格式, 所述样本标签是文本数据格式;

基于所述样本数据集中的样本问题和该样本问题对应的样本标签对待训练的初始大语言模型进行训练, 得到已训练的候选大语言模型;

- 20 将所述候选大语言模型作为所述目标大语言模型; 或,

从所述样本数据集中选取部分样本问题作为参考样本问题, 在参考样本问题的前面添加任一数据格式的第一 token 标识, 在参考样本问题的后面添加该数据格式的第二 token 标识; 基于修改后的参考样本问题和该参考样本问题对应的样本标签对所述候选大语言模型进行训练, 得到所述目标大语言模型。

- 25 13、一种电子设备, 其特征在于, 包括: 处理器和机器可读存储介质, 所述机器可读存储介质存储有能够被所述处理器执行的机器可执行指令; 所述处理器用于执行机器可执行指令, 以实现权利要求 1-7 任一所述的方法。