Claims

1. A human-artificial intelligence (AI) interaction method based on a large language model (LLM), comprising:

(S1) performing pre-management on domain-specific information, wherein the domain-specific information is not covered by training data of the LLM, and the pre-management is performed by constructing a dataset using a non-vector database and comprises:

subjecting a content of the domain-specific information to tokenization and parsing;

taking a tokenization label as a metadata of the content of the domain-specific information; and

constructing the dataset in a format of relational data or non-relational data for management;

(S2) when receiving a user problem or a user request, preprocessing the user problem or the user request by performing tokenization and parsing on the user problem or the user request, wherein the user problem or the user request refers to a user problem or a user request without context-before information;

(S3) adding the domain-specific information to a retrieval scope, and performing indexing through tokenization and keyword weight to find an information fragment that is most similar to the user problem or the user request, wherein after the domain-specific information is added, the retrieval scope comprises the domain-specific information without excluding original public data and semi-public data;

(S4) processing the information fragment as required to obtain a processed information fragment as context information of the user problem or the user request; and

(S5) transmitting the context information together with the user problem or the user request to the LLM for processing according to interface usage method or requirements of the LLM, for the LLM to feed back related information;

wherein a sequential order of steps (S1) and (S2) is not limited.

2. The human-AI interaction method of claim 1, wherein in step (S4), the information fragment is processed through a direct processing approach and/or a reprocessing approach, wherein the reprocessing approach is selected from the group consisting of copying, editing, summarization, sorting, screening, translation, compression, filter, re-encoding, re-retrieving, multiple re-retrieving, code execution based on a retrieval result, and a combination thereof.

3. The human-AI interaction method of claim 1 or 2, wherein in step (S4), the processed information fragment is further used as data required by a future LLM interface.

4. The human-AI interaction method of any one of claims 1 to 3, wherein in step (S5), the transmitting comprises: encrypting the context information and the user problem or the user request followed by transmission to the LLM; or respectively transmitting the context information and the user problem or the user request to the LLM through independent interfaces.

5. A human-artificial intelligence (AI) interaction system based on a large language model (LLM), comprising:

a pre-managing module;

a problem preprocessing module;

a retrieval module;

an information fragment processing module; and

a transmitting module;

wherein the pre-managing module is configured to perform pre-management on domain-specific information, wherein the domain-specific information is not covered by training data of the LLM, and the pre-management is performed by constructing a

dataset using a non-vector database and comprises: subjecting a content of the domain-specific information to tokenization and parsing; taking a tokenization label as a metadata of the content of the domain-specific information; and constructing the dataset in a format of relational data or non-relational data for management;

the problem preprocessing module is configured to, when receiving a user problem or a user request, preprocess the user problem or the user request by performing tokenization and parsing on the user problem or the user request refers to a user problem or a user request without context-before information;

the retrieval module is configured to add the domain-specific information to a retrieval scope, and perform indexing through tokenization and keyword weight to find an information fragment that is most similar to the user problem or the user request, wherein after the domain-specific information is added, the retrieval scope includes the domain-specific information without excluding original public data and semi-public data;

the information fragment processing module is configured to process the information fragment as required to obtain a processed information fragment as context information of the user problem or the user request; and

the transmitting module is configured to transmit the context information and the user problem or the user request to the LLM for processing according to interface usage method or requirements of the LLM, for the LLM to feed back related information.

6. The human-AI interaction system of claim 5, wherein the information fragment processing module is configured to process the information fragment through a direct processing approach and/or a reprocessing approach, wherein the reprocessing approach is selected from the group consisting of copying, editing, summarization, sorting, screening, translation, compression, filtration, re-encoding, re-retrieving, multiple re-retrieving, code execution based on a retrieval result, and a combination thereof.

7. The human-AI interaction system of claim 5 or 6, wherein the information fragment processing module is further configured to take the processed information fragment as data required by a future LLM interface.

8. The human-AI interaction system of any one of claims 5 to 7, wherein the transmitting module is configured to encrypt the context information and the user problem or the user request followed by transmission to the LLM, or to respectively transmit the context information and the user problem or the user request to the LLM through independent interfaces.