|  |  |  |  |
| --- | --- | --- | --- |
| **Inventory of supplemental materials** | | | |
| **File name** | **Specific content** | | **Corresponding parts in the manuscript** |
| Table S1 | The data layer of multimodal CPM-KG  (More details are available in the three versions of multimodal CPM-KG) | Nodes of 1 primary knowledge field | Section 3.1 (Fig. 2) |
| Nodes of 7 secondary knowledge fields | Section 3.1 (Fig. 2 and Table 3) |
| Nodes of 36 tertiary knowledge fields | Section 3.1 (Fig. 2 and Table 3) |
| Nodes of 195 documents | Section 3.1 (Figs. 2 and 3) |
| Nodes of 1,375 document contents | Section 3.1 (Figs. 2 and 3) |
| 43 triples of [knowledge field, has subfield of, knowledge field] | Section 3.1 (Fig. 2) |
| 242 triples of [tertiary knowledge field, involves, document] | Section 3.1 (Fig. 2) |
| The 18 example triples of [document, contains, document content] | Section 3.1 (Figs. 2 and 3) |
| Pre-assessment of CPM-KG by five experts before its application | Section 3.1 |
| Table S2 | The CPM-QA test dataset containing 2,435 questions | | Section 3.3 (Fig. 5) |
| Table S3 | The 38,960 answers to 2,435 questions from 8 LLMs before and after the integration of multimodal CPM-KG | | Sections 4.1, 4.2, and 4.3 |
| Performance comparison of 8 LLMs on CRCEEs | | Section 4.1 (Fig. 7) |
| Performance comparison of 8 LLMs across 36 knowledge subfields | | Section 4.2 (Fig. 8 and Table 5) |
| Performance comparison of 8 LLMs on different question types | | Section 4.3 (Figs. 9-10 and Table 5) |
| Fig. S1 | 43 triples of [knowledge field, has subfield of, knowledge field] | | Section 3.1 (Fig. 2) |
| Fig. S2 | 278 triples of [tertiary knowledge field, involves, document] | | Section 3.1 (Fig. 2) |
| Fig. S3 | 1,375 triples of [document, contains, document content] | | Section 3.1 (Fig. 2) |
| Video S1 | Video clips of running original LLMs | | Section 3.2 |
| Video S2 | Video clips of running LLMs with CPM-KG | | Section 3.2 |
| Multiple CPM-KG | Chroma-formatted vectorized CPM-KG | | Section 3.1 |
| Neo4j-readable CPM-KG dump | | Section 3.1 |
| JSON-formatted CPM-KG | | Section 3.1 |
| Codes | Codes for running original large language models | Codes for original ChatGLM3-6B | Section 3.2 |
| Codes for original GPT-3.5-turbo, GPT-4, and GPT-4-turbo | Section 3.2 |
| Codes for original ERNIE-Bot, ERNIE-Bot 4.0, Qianfan-Chinese-Llama-2-7B, and Llama-2-70B-Chat | Section 3.2 |
| Codes for running large language models integrated with CPM-KG | Codes for ChatGLM3-6B with CPM-KG | Section 3.2 |
| Codes for GPT-3.5-turbo with CPM-KG, GPT-4 with CPM-KG, and GPT-4-turbo with CPM-KG | Section 3.2 |
| Codes for ERNIE-Bot with CPM-KG, ERNIE-Bot 4.0 with CPM-KG, Qianfan-Chinese-Llama-2-7B with CPM-KG, and Llama-2-70B-Chat with CPM-KG | Section 3.2 |

Due to too many materials, all supplemental materials are provided in GitHub repository (https://github.com/0AnonymousSite0/QA\_for\_CPM). Besides the GitHub repository, the CPM-QA test dataset is also shared in Hugging Face repository (<https://huggingface.co/datasets/AnonymousSite/QA_dataset_for_CPM>).