|  |  |  |  |
| --- | --- | --- | --- |
| **Inventory of supplemental materials** | | | |
| **File name** | **Specific content** | | **Corresponding parts in the manuscript** |
| Table S1 | Textbook chapters and sections integration and related documents screening | Chapter merging for three textbooks | Section 3.1 (Figure 3 and Table 4) |
| Section merging for three textbooks | Section 3.1 (Figure 3 and Table 4) |
| Extracting the document name entities and removing the non-knowledge documents | Section 3.1 (Figure 3) |
| CPM-related knowledge documents of the same meaning merging | Section 3.1 (Figure 3) |
| Five experts participated in integrating textbooks and screening documents | Section 3.1 |
| Table S2 | The instance 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 (Figure 2) |
| Nodes of 7 secondary knowledge fields | Section 3.1 (Figures 2-3 and Table 4) |
| Nodes of 36 tertiary knowledge fields | Section 3.1 (Figures 2-3 and Table 4) |
| Nodes of 195 documents | Section 3.1 (Figures 2-4) |
| Nodes of 1,375 document contents | Section 3.1 (Figures 2-4) |
| 43 triples of [knowledge field, has subfield of, knowledge field] | Section 3.1 (Figure 2) |
| 242 triples of [tertiary knowledge field, involves, document] | Section 3.1 (Figure 2) |
| The 18 example triples of [document, contains, document content] | Section 3.1 (Figure 2) |
| Pre-assessment of CPM-KG by five experts before its application | Section 3.1 |
| Table S3 | The CPM-QA testing set containing 2,435 multiple-choice questions | | Section 3.3 (Figure 7) |
| 100 case-based questions from CRCEEs | | Section 4.5 |
| Table S4 | The 38,960 answers to 2,435 multiple-choice questions from 8 GLMs before and after the integration of multimodal CPM-KG | | Sections 4.1, 4.2, and 4.3 |
| Precision and recall of 2,435 multiple-choice questions from 8 GLMs before and after the integration of multimodal CPM-KG | | Sections 4.1, 4.2, and 4.3 |
| The 1,600 answers to 100 case-based questions from 8 GLMs before and after the integration of multimodal CPM-KG | | Section 4.4 |
| Review of each case-based question by three experts | | Section 4.4 |
| The 38,960 answers to 2,435 multiple-choice questions from 8 GLMs based on single- and multi-modal CPM-KG | | Section 4.5 |
| Performance comparison of 8 GLMs on CRCEEs | | Section 4.1 (Figure 9) |
| Performance comparison of 8 GLMs across 36 knowledge subfields | | Section 4.2 (Figure 10) |
| Performance comparison of 8 GLMs on different question types | | Section 4.3 (Figures 11 and 12) |
| Performance comparison between single- and multi-modal CPM-KGs | | Section 4.4 (Figures 13 and 14) |
| Performance comparison in case-based questions | | Section 4.5 (Figure 15) |
| Table S5 | Functional validation of the Smart QA of CPM | | Section 5.3 (Figures 16 and 17) |
| 10 Samples of using the self-explanation prompt | | Section 5.4 |
| Figure S1 | 43 triples of [knowledge field, has subfield of, knowledge field] | | Section 3.1 (Figure 2) |
| Figure S2 | 242 triples of [tertiary knowledge field, involves, document] | | Section 3.1 (Figure 2) |
| Figure S3 | 1,375 triples of [document, contains, document content] | | Section 3.1 (Figure 2) |
| Video S1 | Video clips of running original GLMs | | Section 3.2 |
| Video S2 | Video clips of running GLMs with CPM-KG | | Section 3.2 |
| **File name** | **Specific content** | | **Corresponding parts in the manuscript** |
| Video S3 | Video clips of Smart QA of CPM | | Section 5.3 (Figures 16 and 17) |
| Multimodal 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 |
| Single-modal CPM-KG | Chroma-formatted vectorized CPM-KG | | Section 4.4 |
| Neo4j-readable CPM-KG dump | | Section 4.4 |
| JSON-formatted CPM-KG | | Section 4.4 |
| Codes | Codes for running original GLMs | 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 GLMs integrated with CPM-KG | Codes for ChatGLM3-6B with CPM-KG | Section 3.2 |
| Codes for GPT-3.5-turbo, GPT-4, and GPT-4-turbo with CPM-KG | Section 3.2 |
| Codes for ERNIE-Bot, ERNIE-Bot 4.0, Qianfan-Chinese-Llama-2-7B, and Llama-2-70B-Chat with CPM-KG | Section 3.2 |
| Codes for deploying and running the CPM-QA prototype | | Section 5.3 |
| Dmeta-embedding-zh | More details on Dmeta-embedding-zh and its parameters | | Section 3.2 (Figure 6) |

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>).