|  |  |  |  |
| --- | --- | --- | --- |
| **Inventory of supplemental materials** | | | |
| **File name** | **Specific content** | | **Corresponding parts in the manuscript** |
| Table S1 | Experts for developing CCLR knowledge graph and QA dataset | Experts for helping confirm CCLR knowledge graph before its application | Section 3.1 |
| Experts for formulating the self-designed questions | Section 3.3 |
| Table S2 | The data layer of CCLR knowledge graph  (More details are available in the three versions of CCLR knowledge graph) | Nodes of 1 first-tier knowledge domain | Section 3.1 (Figs. 2 and 3) |
| Nodes of 8 second-tier knowledge domains | Section 3.1 (Figs. 2 and 3) |
| Nodes of 193 third-tier knowledge domains | Section 3.1 (Figs. 2 and 3) |
| Nodes of 276 laws and regulations | Section 3.1 (Figs. 2 and 3) |
| Nodes of 2,069 clauses | Section 3.1 (Figs. 2 and 3) |
| 201 triples of [knowledge domain, has subdomain of, knowledge domain] | Section 3.1 (Figs. 2 and 3) |
| 657 triples of [knowledge domain, is related to, law and regulation] | Section 3.1 (Figs. 2 and 3) |
| The 8 example triples of [law and regulation, contains, clause] | Section 3.1 (Figs. 2 and 3) |
| Table S3 | The QA dataset containing 6,359 questions | | Section 3.3 (Fig. 6) |
| Table S4 | The 89,026 answers to 6,359 questions from 7 LLMs before and after the integration of CCLR knowledge graph | | Sections 4.1, 4.2, and 4.3 |
| Effects analysis of LLMs under different settings | | Section 4.5 |
| Table S5 | The performance of LLMs based on different embedding models and chunking parameters configurations | The performance and the scoring rate of 5 LLMs in pilot dataset based on 7 embedding models | Section 3.2.1 |
| The performance and scoring rate of 7 LLMs in pilot dataset based on 12 [chunk size, overlap size] configurations | Section 3.2.1 |
| Table S6 | Experts cross-checking in 4,119 self-designed questions | 4,119 self-designed questions from 16 experts in Table S1 | Section 3.3 |
| Expert consistency on self-designed questions by Gwet’s AC1 coefficient | Section 3.3 |
| The specific scoring process of each expert for each question | Section 3.3 |
| Table S7 | The scoring rates of 4 legal LLMs in CCLR QA dataset | | Section 4.4 |
| The scoring rates of ChatLaw and knowledge-incorporated LLMs in 2023 first-level RCQE | | Section 4.4 |
| Table S8 | The three source documents of each question | | Section 4.5 (Fig. 11) |
| Table S9 | The data layer of HKCLR knowledge graph | Nodes of 1 first-tier knowledge domain | Section 5.3 (Fig. 14) |
| Nodes of 8 second-tier knowledge domains | Section 5.3 (Fig. 14) |
| Nodes of 403 laws and regulations | Section 5.3 (Fig. 14) |
| Nodes of 2,727 clauses | Section 5.3 (Fig. 14) |
| 8 triples of [knowledge domain, has subdomain of, knowledge domain] | Section 5.3 (Fig. 14) |
| 663 triples of [knowledge domain, is related to, law and regulation] | Section 5.3 (Fig. 14) |
| 2,727 triples of [law and regulation, contains, clause] | Section 5.3 (Fig. 14) |
| Table S10 | The HKCLR QA dataset containing 150 questions and 2,100 answers from 7 LLMs before and after the integration of HKCLR knowledge graph | | Section 5.3 (Fig. 14) |
| Fig. S1 | 201 triples of [knowledge domain, has subdomain of, knowledge domain] in CCLR knowledge graph | | Section 3.1 (Fig. 3a) |
| Fig. S2 | 657 triples of [knowledge domain, is related to, law and regulation] in CCLR knowledge graph | | Section 3.1 (Fig. 3b) |
| Fig. S3 | 2,213 triples of [law and regulation, contains, clause] in CCLR knowledge graph | | Section 3.1 (Fig. 3c) |
| Fig. S4 | 663 triples of [knowledge domain, is related to, law and regulation] in HKCLR knowledge graph | | Section 5.3 (Fig. 14b) |
| Fig. S5 | 2,727 triples of [law and regulation, contains, clause] in HKCLR knowledge graph | | Section 5.3 (Fig. 14c) |
| Video S1 | Video clips of running original LLMs | | Section 3.3 |
| Video S2 | Video clips of running LLMs with CCLR knowledge graph | | Section 3.3 |
| CCLR knowledge graph | Chroma-formatted vectorized CCLR knowledge graph | | Section 3.1 |
| Neo4j-readable CCLR knowledge graph dump | | Section 3.1 |
| JSON-formatted CCLR knowledge graph | | Section 3.1 |
| Codes | Codes for running original large language models | Codes for original text-davinci-003, GPT-3.5-turbo, and GPT-4 | Section 3.2 |
| Codes for original ERNIE-Bot 4.0, Qianfan-Chinese-Llama-2-7B, and Llama-2-70b | Section 3.2 |
| Codes for original ChatGLM2-6B and ChatGLM3-6B | Section 3.2 |
| Codes for running large language models integrated with knowledge graph | Codes for text-davinci-003 with knowledge graph, GPT-3.5-turbo with knowledge graph, and GPT-4 with knowledge graph | Section 3.2 |
| Codes for ERNIE-Bot 4.0 with knowledge graph, Qianfan-Chinese-Llama-2-7B with knowledge graph, and Llama-2-70b with knowledge graph | Section 3.2 |
| Codes for ChatGLM2-6B with knowledge graph and ChatGLM3-6B with knowledge graph | Section 3.2 |
| Codes for deploying and running the CEM-QA prototype | | Section 5.2 |

Due to too many materials, all supplemental materials are provided in GitHub repositor(https://huggingface.co/datasets/AnonymousSite/QA\_dataset\_for\_CCLR). Besides the Github repository, the CCLR QA dataset is also shared in Hugging Face repository (https://github.com/0AnonymousSite0/QA\_for\_CCLR).