

A.2.2 HUMAN EVALUATION

In this section, we demonstrate the cases of human analysis. In our self-constructed evaluation dataset, the instructions are either manually written data or manual revised from public datasets, such as CLiB⁶, C-Eval (Huang et al., 2023), FacTool (Chern et al., 2023b), LeetCode⁷, etc.

In terms of each case, we demonstrate the responses and Elo ratings⁸ of all models for comparison. Specifically, as the data in our human evaluation are in Chinese, we also provide their translations in English.

Knowledge		
Responses		
Model	Response	Elo
Qwen-7B-Chat (SFT)	<p>北京大学的前身是什么学校? What was the predecessor of Peking University?</p>	1037
Qwen-14B-Chat (SFT)	<p>北京大学的前身是京师大学堂，创建于1898年，是中国近代第一所国立大学。 The predecessor of Peking University was the Imperial University of Peking, established in 1898. It was the first national university in modern China and also the earliest higher education institution to be named a “university”.</p>	1011
Qwen-14B-Chat (RLHF)	<p>北京大学的前身是京师大学堂，创建于1898年，是中国近代第一所国立综合性大学。京师大学堂最初设立的学科有经学、文学、史学、哲学、法律、教育、农学、工学、商学、医学等，后来逐渐扩大到社会科学、自然科学、工程技术、艺术等多个领域。1912年，京师大学堂改名为北京大学，成为中国最高学府之一，也是中国现代高等教育的发源地之一。 Peking University, originally known as the Imperial University of Peking, was founded in 1898 and is the first comprehensive national university in modern China. The initial disciplines established at the Imperial University of Peking included classics, literature, history, philosophy, law, education, agriculture, engineering, business, medicine, etc., gradually expanding to encompass social sciences, natural sciences, engineering and technology, arts, and other fields. In 1912, the Imperial University of Peking changed its name to Peking University, becoming one of the top universities in China and also one of the birthplaces of modern higher education in China.</p>	1090
GPT-3.5	<p>北京大学的前身是国立北京大学，建于1898年，是中国近代第一所国立大学。其成立初衷是为了培养国家的政治、经济、军事和外交人才，为中国的现代化进程做出贡献。 The predecessor of Peking University was National Peking University, established in 1898. It is the first national university in modern China. Its original purpose was cultivating talents in politics, economics, military, and diplomacy and contributing to China's modernization process.</p>	910

⁶<https://github.com/jeinlee1991/chinese-llm-benchmark>

⁷<https://leetcode.cn/>

⁸https://en.wikipedia.org/wiki/Elo_rating_system