

Table 5. Several Examples of the Corpus

Personality Knowledge Statements	Practical Scenario Cases
EN: You waste your time. ZH: 你浪费自己的时间。	In everyday life: A. you always use your time productively. B. you always spend time on meaningless activities. 在日常生活中: A. 你总是有效地利用时间。 B. 你总是在无意义的活动上花费时间。
EN: You complete tasks successfully. ZH: 你能成功完成任务。	When assigned a challenging project with a tight deadline: A. you are overwhelmed and have difficulty moving the process forward effectively, often resulting in incomplete or unsatisfactory results. B. you organise your work and manage your resources properly, and the project is often completed successfully and on time. 当被指派一个期限紧迫的具有挑战性的项目时: A. 你不知所措, 难以有效地推进进程, 常导致结果不完整或不令人满意。 B. 你组织工作, 妥善管理资源, 项目往往按时顺利完成。
EN: You shirk your duties. ZH: 你推卸责任。	When someone points out a mistake in your work: A. you take responsibility. B. you shirk your responsibility. 当别人指出你的工作失误: A. 你勇于承担责任。 B. 你推卸责任。
EN: You tend to find fault with others. ZH: 你喜欢挑剔别人的毛病。	When dealing with people: A. you tend to focus on the person's good points and strengths. B. you often pick on other people's faults and weaknesses. 在与人相处时: A. 你往往关注他的优点与长处。 B. 你常挑剔别人的缺点与毛病。
EN: You usually postpone finalizing decisions for as long as possible. ZH: 你通常会尽可能推迟最终决定。	When making choices: A. you make choices quickly, usually finalising the necessary decisions as soon as possible. B. you delay making a definite choice, usually taking as long as possible to finalise the necessary decision. 在做选择时: A. 你会迅速做出选择, 通常会尽快敲定必要的决定。 B. 你会推迟做出明确的选择, 通常会尽可能长时间地敲定必要的决定。
EN: You struggle with deadlines. ZH: 你很难在最后期限前完成任务。	You have a week to complete a work project: A. you always make sure that it is completed ahead of or on the deadline. B. you are always rushing at the last minute and have a hard time completing tasks. 你有一周的时间来完成一个工作项目: A. 你往往确保提前或在截止日期完成。 B. 你总是在最后一刻还在赶工, 很难完成任务。
EN: You remain calm in tense situations. ZH: 你在紧张情境中仍保持冷静。	When dealing with a conflict or a high-pressure problem: A. you become visibly agitated, finding it challenging to maintain composure. B. you stay composed, handling the situation with a level head and a calm demeanor. 在处理冲突或高压问题时: A. 你会明显变得焦躁不安, 发现保持镇定很有挑战性。 B. 你保持镇定, 以平和的心态和冷静的举止处理情况。
EN: You are the life of the party. ZH: 聚会时你是活跃气氛的人。	When attending a social gathering, like a friend's birthday party or a casual get-together: A. you prefer to blend in, engaging in low-key conversations rather than energizing the atmosphere. B. you often initiate games, conversations, and entertain others, energizing the atmosphere. 参加社交聚会, 如朋友的生日派对或休闲聚会时: A. 你喜欢融入其中, 低调地交谈, 而不是主动活跃气氛。 B. 你经常会主动发起游戏、谈话, 活跃气氛。

Table 6. Various Prompts of Personality Knowledge Questionnaire

Source	Prompt
16 Personalities Test	<i>You can only reply to me with numbers from 1 to 7. Score each statement on a scale of 1 to 7 with 1 being disagree, 4 being not sure and 7 being agree.</i>
MBTI-M Test	<i>Read the following statements carefully and rate each one from 1 to 7, with 7 meaning that it applies to you completely, 1 meaning that it doesn’t apply to you at all, and 4 meaning that you are not sure whether it applies to you or not.</i>
TDA-100 Test	<i>Below are several descriptions that may or may not fit you. Please indicate how much you agree or disagree with that statement by giving a specific number from 1 to 7. 1 means you totally disagree with the statement, 4 means you are not sure, and 7 means you totally agree with the statement.</i>
BFI-44-children adapted version	<i>Here are several statements that may or may not describe what you are like. Write the number between 1 and 7 that shows how much you agree or disagree that it describes you. 1 means you disagree strongly that the statement applies to you, 4 means you are not sure, and 7 means you agree strongly with the statement.</i>
Dario Nardi’s Cognitive Test	<i>Please read carefully each of the phrases below. For each phrase: Rate how often you do skillfully what the phrase describes between 1 and 7. 1 means the phrase is not me, 4 means that you are not sure, and 7 means that the phrase is exactly me.</i>

F. Significance Tests

In the following, we will apply significance tests to further demonstrate significant differences between the performance of LLMs and humans. We incorporated significance testing for the responses of LLMs and humans in the same experiment. Specifically, we performed permutation tests to compare LLMs’ results and human respondents’ results, yielding p-values significantly below 0.05 in experiments in Section 3.2 (corresponding to results in Table 1). This confirms substantial disparities between LLMs and humans in performance for each metric across both experiments. The specific p-values are outlined in Table 7.

Table 7. Results of significance testing for the responses of LLMs and humans in the experiment in Section 3.2

P-value of COMPARISON between LLMs & Human	Cosine Similarity	Spearman Rank Correlation Coefficient	Value Mean Difference	Proportion of Consistent Pairs
Results in Table 1	4.91e-05	4.91e-05	1.47e-04	2.46e-05

G. Additional Notes On Human Reviewers and Respondents

G.1. Recruitment of Scenario Providers, Reviewers and Human Respondents

We recruited individuals from undergraduate, postgraduate and PhD students. Taking the International English Language Testing System(IELTS), CET 6 exam results, and their GPA in English courses into account, we recruited 16, 10 and 35 native Chinese speakers as reviewers and respondents.

G.2. Instructions Given to Scenario Providers

Before requiring the individual to complete the following tasks, we asked the respondents whether they agreed to the anonymisation of their reviews for scientific research and subsequent publication. Only if the respondents gave their consent were they given the corpus to review. And we promised not to publish each individual’s MBTI results and specific practical scenario cases. Then, we investigated each person’s MBTI type and ensured that we ultimately recruited 16 individuals with distinct MBTI types. After this, we required the reviewers to accomplish the following tasks:

Please provide a practical scenario case for every statement in the personality knowledge questionnaire. Each scenario case comprises situations drawn from your own lives, along with two completely contrasting actions: Action A and Action B. Action A fully aligns with the statement, while Action B completely contradicts it. The content of Action A and Action B need to be kept basically the same length.

G.3. Instructions Given to Reviewers

We require the reviewers to accomplish the following tasks:

- Please determine whether the practical scenario case is consistent with its corresponding personality knowledge statement. If yes, rate 1. If not, rate 0.
- If you rate 0 for all of the practical scenario cases of a personality knowledge statement, please offer suggestions to improve the practical scenario design. It would be better if an example could be provided.

G.4. Instructions Given to Respondents

Before answering the questionnaires, we did not tell the respondents what kind of questionnaires they would be answering or how the questions were related to each other. In addition to this, we asked the respondents whether they agreed to the anonymisation of their answers for scientific research and subsequent publication. Only if the respondents gave their consent were they given the questionnaires to answer.

In all experiments that appeared in our research, human respondents received the exact same prompts that LLM received. The difference is that in the case of experiments with multiple prompts with similar meanings, LLM responded multiple times by prompt type, while human subjects read all the prompts and responded only once.