

Article

Not peer-reviewed version

---

# ChatGPTTest: A Mixed-Methods Study on AI-Generated Questionnaires Based on Theoretical Models

---

[Menghan Cheng](#) and [Changsheng Wang](#) \*

Posted Date: 12 May 2025

doi: 10.20944/preprints202505.0823.v1

Keywords: Artificial Intelligence; Mixed-methods research; Questionnaires



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

## Article

# ChatGPTTest: A Mixed-Methods Study on AI-Generated Questionnaires Based on Theoretical Models

Menghan Cheng <sup>1,2</sup> and Changsheng Wang <sup>3,\*</sup>

<sup>1</sup> School of Design & Arts, Beijing Institute of Technology, Beijing, CHN

<sup>2</sup> Tangshan Research Institute, Beijing Institute of Technology, Tangshan, Hebei, CHN

<sup>3</sup> Sejong University, Seoul, KOR

\* Correspondence: Changsheng Wang (e-mail: zhangshawang@gmail.com).

**Abstract:** This study employed a hybrid research design that combines quantitative expert assessment and qualitative in-depth interviews to systematically examine the performance of AI-generated questionnaires based on single and integrated theoretical models, comparing them with manually created questionnaires. The research found that, under a single model, AI questionnaires showed significant differences from manual questionnaires in three dimensions: accuracy, comprehensiveness, and clarity. AI demonstrated limitations when handling complex and abstract variables, lacked the flexibility of human-designed questions, and tended to satisfy model requirements rather than research topic requirements. Under the integrated models, significant differences existed between the AI and manual questionnaires in four dimensions: accuracy, comprehensiveness, clarity, and redundancy. As the model complexity increased, the probability of problems in the AI questionnaires correspondingly increased. However, regardless of whether single or integrated models were used, AI questionnaires exhibited advantages in language fluency and generation efficiency and surpassed manual questionnaires in objectivity. Researchers may benefit from using AI to generate initial questionnaire drafts followed by manual corrections. This study not only reveals the advantages and limitations of AI-generated questionnaires but also provides theoretical and practical guidance for researchers. Future research should improve AI's contextual understanding capabilities and generation logic to further enhance the effectiveness of the AI-assisted questionnaire design.

**Keywords:** artificial intelligence; mixed-methods research; questionnaires

## 1. Introduction

Over the past decade, artificial intelligence (AI) has developed rapidly, achieving significant breakthroughs in multimodal domains such as speech processing, video analysis, 3D modeling, and image generation, and has been widely applied [1,2]. AI technology can be used for natural language processing tasks such as text classification, sentiment analysis, and machine translation [3]; generating high-quality images through prompt inputs [4]; simulating real human voices [5]; generating videos [6] and 3D models [7]; and implementing multimodal large models [8].

AI technology also shows enormous potential for experimental research [9]. Scholars have begun to use AI to enhance research efficiency by viewing it as a collaborative researcher [10]. Iskender *et al.* used AI as an interviewee to explore its perspectives in the hotel, tourism, and education industries [11]. Xiaotian *et al.* utilized AI for quantitative analysis and found that ChatGPT can improve sports researchers' literature analysis capabilities and research efficiency [12]. Wang Yibo used ChatGPT to write Chinese abstracts and discovered that it produced more logical but homogeneous content [13]. AI technology has also provided new approaches to questionnaire design. Benjamin Lebrun *et al.* had humans and AI separately identify questionnaires filled out by AI, finding that humans could judge the identity of questionnaire authors with above-random accuracy, whereas AI does not yet possess

this ability [14]. Olivos *et al.* used AI for questionnaire pre-testing, suggesting that AI feedback can serve as an auxiliary stage for human prediction, reducing iterations, although researcher judgment remains indispensable [15].

Traditional questionnaire design requires researchers to select appropriate theoretical models as a foundation based on research objectives. Theoretical models provide a systematic framework to clarify key variables and their relationships, ensuring that data effectively reflects the research topic, and improving data validity [16]. Researchers can use a single model to design questionnaires. When research purposes are complex, they may need to integrate multiple models or variables to form an integrated model [17]. However, this process is time consuming and susceptible to subjective bias. Although AI offers new approaches to questionnaire design, current research on AI-generated questionnaires based on theoretical models remains insufficient. This study aimed to explore the following aspects:

(1) The performance of AI-generated questionnaires based on single theoretical models and their differences from manually created questionnaires. (2) The performance of AI-generated questionnaires based on integrated theoretical models and their differences from manually created questionnaires.

To address these questions, this study conducted three substudies. Study One randomly selected a theoretical model and an article using this model from the literature, using its questionnaire as the manual questionnaire, while simultaneously generating an AI questionnaire based on the same research objectives, and invited experts to evaluate both questionnaires. Study Two identified studies that integrated the model from Study One with other models or variables, randomly selected an article, used its questionnaire as the manual questionnaire, generated an AI questionnaire based on the same research objectives, and invited experts to evaluate both questionnaires. Study Three conducted expert interviews to explore the performance of AI-generated questionnaires and their differences from manual questionnaires in depth. By synthesizing the results of these three studies, this study answers the research questions and provides new perspectives for AI-assisted experimental research.

## 2. Literature Review

### 2.1. AI-Assisted Academic Research

AI-related research can be traced back to the 1990s when it was still in the experimental stage. After entering the 21st century, tech giants, such as Microsoft, began training machines to generate content. In 2012, Microsoft released a fully automated simultaneous interpretation system, and in 2014, the Generative Adversarial Network (GAN) proposed by Goodfellow could be used for text generation [18]. In 2019, the DVD-GAN model was capable of generating continuous videos [19], and in 2022, David Holz's studio developed Midjourney, which could generate images based on natural language input in just one minute [20].

With the advent of ChatGPT, AI has demonstrated enormous potential in assisting academic research. Trained by massive text data, ChatGPT can engage in fluent conversations with users through text generation. On March 14, 2023, GPT-4.0 was released, which, compared with previous versions, added an application marketplace, making it a feature-rich ecosystem that allows researchers to select appropriate plugin applications from the market according to task requirements. This product can process multimodal information, including images and audio, and has demonstrated near-human capabilities in various professional and academic benchmark tests [21]. This progress marks the beginning of a new era in AI-assisted academic research.

In recent years, many scholars have begun to apply AI technologies such as ChatGPT to academic research. Xiaotian *et al.* [12] explored the specific applications and contributions of ChatGPT in sports science research from multiple dimensions, including literature learning, research methodology reform, and research paradigm transformation. They found that large language models can improve the efficiency and accuracy of sports science research, although they also face some application challenges. The study proposed innovative solutions, emphasizing that the future effective use and improvement of this tool could have

far-reaching impacts on sports science research. Wang Yibo *et al.* [13] compared Chinese abstracts written by ChatGPT with 100 highly cited paper abstracts and found that ChatGPT's writing is logically strong, but its generated abstracts show high homogeneity, whereas abstracts written by scholars demonstrate significant individual differences. Dowling *et al.* [22] generated a complete finance paper including a literature review, research design, data collection, and empirical analysis by inputting prompts to ChatGPT sequentially and submitted it for expert blind review. The results showed that papers generated using precollected and organized data were of higher quality than those generated by ChatGPT's autonomous data search, indicating that researchers' professional skills remain a key factor in ensuring paper quality.

In terms of automated experiments, the MolFormer model was pre-trained on 1.1 billion molecules and surpassed state-of-the-art graph neural network models in multiple benchmark tests. Through large-scale pre-training, this model learned rich molecular representations and achieved excellent performance in various property-prediction tasks. BioPlanner is dedicated to optimizing biological experimental workflows, capable of transforming detailed experimental steps into structured pseudocode representations, greatly improving experimental planning efficiency. In practical testing, BioPlanner successfully generated and executed complete workflows for *E. coli* cultivation and cryopreservation [23]. These examples demonstrate that AI has demonstrated powerful capabilities in assisting academic research and is becoming a valuable assistant to researchers.

## 2.2. AI and Questionnaires

Questionnaires are one of the most commonly used research tools in academic research. With the continuous development of AI technology, many researchers have attempted to apply AI to various aspects of questionnaires, including their detection and completion.

Lebrun *et al.* [14] found that recruiting participants to complete online questionnaires through crowdsourcing platforms has become increasingly prevalent, owing to its ease of use and low cost. Many unscrupulous vendors use AI to complete online questionnaires automatically, reducing the research quality of online surveys. Through experiments in which both humans and AI were tasked with detecting AI-completed questionnaires, Lebrun *et al.* concluded that humans can identify the authorship of such questionnaires with above-random accuracy, whereas AI cannot currently detect such questionnaires. Olivos *et al.* [15] explored the use of ChatGPT as a tool for pre-testing survey questionnaires, suggesting that GPT can serve as an auxiliary stage before human prediction, helping reduce the number of questionnaire iterations. However, they also emphasized that researchers should incorporate their own judgments when interpreting and implementing AI feedback. Bing *et al.* [24] used ChatGPT to complete questionnaires and found that it could be developed into a useful tool to assist researchers in conducting experiments and questionnaire surveys, greatly saving research time and improving efficiency. Compared to traditional small-sample surveys and experiments, the questionnaire results completed by ChatGPT based on ultra-large-scale training sets are more representative, independently distributed, and suitable for sentiment analysis. However, problems, such as overly concentrated samples, insufficient questionnaire randomness, and limited information, exist.

Currently, researchers primarily use AI as an evaluation or completion tool for questionnaires. However, research on whether AI can generate questionnaires based on theoretical models is scarce. This study aims to fill this gap by thoroughly examining the performance of AI-generated questionnaires and their differences from manually created questionnaires, thereby providing new insights for AI-assisted experimental research.

## 3. Research Methods

This study employed a mixed research design that combines quantitative and qualitative methods to comprehensively evaluate the quality of AI-generated questionnaires and their differences from manually created questionnaires. For the quantitative portion, we invited experts to evaluate and score questionnaires generated by AI based on single models and integrated models, as



well as corresponding manually created questionnaires, comparing the performance of different types of questionnaires through quantitative analysis. In the qualitative portion, we conducted in-depth interviews to understand experts' views and feedback on AI-generated questionnaires and cross-validated these with quantitative results to gain a more comprehensive and in-depth understanding.

3.1. Experimental Design

Currently, multiple theoretical models are widely applied in fields such as technology acceptance, user behavior, psychology, education, and marketing. This study places high importance on the selection of experimental theoretical models, as theoretical models are directly related to questionnaire quality, with high-quality models reflecting research topics more effectively. Therefore, we first compiled commonly used theoretical models in various research fields and then randomly selected one as our experimental model. Ultimately, we selected the UTAUT (Unified Theory of Acceptance and Use of Technology) model as the experimental model for this study. The UTAUT model is a theoretical framework for explaining and predicting user acceptance and the use of new technology, as proposed by Venkatesh and other scholars in 2003. Its core elements include Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), Facilitating Conditions (FC), Behavioral Intention (BI), and Use Behavior (UB). PE refers to the degree to which users believe that technology can improve their work efficiency, EE refers to the degree of perceived ease of use, SI refers to the influence of surrounding people on users' technology use, and FC refers to the degree of objective environmental support for technology use.

3.2. Single Model Questionnaire Generation

After determining the experimental model, the next step is to determine the experimental topic. Considering that subsequent experiments would compare AI-generated questionnaires with manually created questionnaires, we decided to select published articles to ensure the quality of the manual questionnaires. The questionnaires used in the published articles were carefully designed by the authors and provided data support for subsequent research, making them highly valuable references. First, we searched for papers with "UTAUT" in their titles from the China National Knowledge Infrastructure (CNKI) and selected the 50 most recently published articles. We then organized each article, screened out those that only used the UTAUT model, and eliminated articles that incorporated other models or variables, resulting in 18 articles. Next, we eliminated articles that did not display questionnaires or had fewer than 3 items per variable, resulting in 14 articles. Finally, we randomly selected one article from these 14 articles used the manual questionnaire for the first study.

The randomly selected article was "Research on Willingness of Medical Students to Use Virtual Simulation Experiment Platforms Based on the UTAUT Model[25]." This study explores medical students' willingness to use virtual simulation experimental platforms and influencing factors based on the UTAUT model, providing a scientific basis for improving and promoting virtual simulation experiments. Each dimension ultimately adopted three items, all using a 5-point Likert scale, with scores from 1 to 5 representing "strongly disagree," "disagree," "neutral," "agree," and "strongly agree." strongly agree. The questionnaire contents are presented in Table 1.

**Table 1.** Medical Students' Virtual Simulation Experiment Platform Usage Willingness Questionnaire (Manually Created).

Dimension	Items
Performance Expectancy	1. The application of virtual simulation can compensate for scenarios that cannot be realized in real teaching (such as face-to-face interviews with patients).
	2. Using virtual simulation helps me develop practical operation skills.
	3. Virtual simulation experiments can give me a sense of learning achievement.
Effort Expectancy	1. The system function options of the virtual simulation platform are simple and easy to use.

Social Influence	2. The page response speed is fast during the use of the virtual simulation platform.
	3. The course content organization structure of the virtual simulation platform is clear and has a strong sense of hierarchy.
	1. Surrounding classmates are all using the virtual simulation platform, and I cannot fall behind.
Facilitating Conditions	2. Feel proud of being able to complete course learning from the virtual simulation platform.
	3. Recognize that learning on the virtual simulation platform is one of the main learning modes for future careers.
	1. The learning time and location of the virtual simulation platform are somewhat limited.
Behavioral Intention	2. The virtual simulation platform is inconvenient to switch between other programs or applications.
	3. The virtual simulation platform requires network and computer, which is very inconvenient for you.
	1. I plan to continue trying to use the educational resources of the virtual simulation platform.
	2. I am willing to recommend the teaching resources of the virtual simulation platform to classmates, friends, etc.
	3. I actively use the virtual simulation platform for learning.

After determining the manual questionnaire, the next step was to generate a questionnaire based on the same theoretical model and research topics. First, we opened ChatGPT-4.0 and created a new conversation. Then, we input the prompt "Please introduce the UTAUT model," and ChatGPT provides relevant information about the UTAUT model. Next, we input the research topic and content needed: "I now want to explore medical students' willingness to use virtual simulation experiment platforms and influencing factors. Please create a research questionnaire based on the UTAUT model. Set three questions for each variable'." The ChatGPT subsequently generated a questionnaire. Although we did not explicitly specify the scale type, the ChatGPT chose a five-point Likert scale consistent with the manual questionnaire. The contents of the AI-generated questionnaire are shown in Table 2.

**Table 2.** Medical Students' Virtual Simulation Experiment Platform Usage Willingness Questionnaire (AI-generated).

Dimension	Items
Performance Expectancy	1. I believe the virtual simulation experiment platform can help me improve my academic performance and experimental skills.
	2. I believe using the virtual simulation experiment platform can enhance my experimental operation abilities.
	3. I believe the virtual simulation experiment platform provides practical help for my medical studies.
Effort Expectancy	1. I believe the virtual simulation experiment platform is easy to operate.
	2. I believe the interface design of the virtual simulation experiment platform is simple and intuitive.
	3. I believe the process of learning how to use the virtual simulation experiment platform is not too complicated.
Social Influence	1. My classmates and teachers recommend using the virtual simulation experiment platform.
	2. I feel that my peers' use of the virtual simulation experiment platform influences my willingness to use it.
	3. I believe that using the virtual simulation experiment platform in learning is widely accepted and promoted.

Facilitating Conditions	1. I believe the school provides sufficient resources (such as computer equipment, network support) to use the virtual simulation experiment platform.
	2. I believe technical support and help for the virtual simulation experiment platform is easily accessible.
	3. I believe I have sufficient knowledge and skills to use the virtual simulation experiment platform.
Behavioral Intention	1. I am willing to continue using the virtual simulation experiment platform for learning.
	2. I plan to frequently use the virtual simulation experiment platform in my future studies.
	3. Even without mandatory requirements, I would choose to use the virtual simulation experiment platform for learning.
Use Behavior	1. In the past three months, I have used the virtual simulation experiment platform for learning.
	2. In the past three months, I have conducted experimental operations through the virtual simulation experiment platform.
	3. I frequently use the virtual simulation experiment platform for medical experimental operations.

3.3. Integrated Model Questionnaire Generation

Integrated model questionnaires were selected using the same method as in the first study. First, we searched for papers with "UTAUT" in their titles from the China National Knowledge Infrastructure (CNKI) and selected the 50 most recently published papers. We then organized each article and screened out those that combined the UTAUT model with other models or variables, resulting in 32 papers. Next, we eliminated articles that did not display questionnaires and those with fewer than three items per variable, leaving a total of 25 papers. Finally, we randomly selected one article from these 25 papers and used the manual questionnaire for the second study. The finally selected article was "Study on the Willingness of Chinese University Students to Learn AI Painting Tools Based on an Extended UTAUT Model" [26]. This study aimed to explore the willingness of design major students in Chinese higher education to accept and use AI painting tools. This study adopted an extended UTAUT model to systematically study the influence of individual innovativeness, AI anxiety, performance expectancy, effort expectancy, social influence, and facilitating conditions on student acceptance and use of AI painting tools. The questionnaire used in this study is presented in Table 3.

Table 3. College Students' Willingness to Learn AI Painting Tools Questionnaire (Manually generated).

Dimension	Items
Performance Expectancy	1. I believe learning AI painting tools can enhance my overall abilities.
	2. I think using AI painting tools can improve my creative efficiency.
	3. I believe AI painting technology will be very useful in the future.
Effort Expectancy	1. I find it easy to learn how to use AI painting tools.
	2. I find using AI painting tools is not complicated/does not require much mental effort.
	3. I can easily master using AI painting tools.
Social Influence	1. Friends around me occasionally discuss and exchange information about AI painting tools.
	2. My friends/teachers recommend that I use AI painting tools.
	3. My company/industry is gradually beginning to use AI painting tools.
Facilitating Conditions	1. I have sufficient resources to use AI painting tools.
	2. I have enough time to learn and use AI painting tools.
	3. I can access the information needed to learn AI painting tools.
Individual Innovativeness	1. I like to try new information technologies.

AI Anxiety	2. Among my peers, I am usually the first to try new information technologies.
	3. If I hear about a new information technology, I will try to experiment with it.
	1. I worry that AI painting tools might generate content that violates privacy.
Behavioral Intention	2. I am concerned that as AI painting technology develops, the skills I have previously learned may become less important.
	3. I worry that the widespread application of AI painting tools may affect my career prospects.
	1. I plan to use AI painting tools in my future research.
	2. I will recommend others to use AI painting tools.
	3. I am willing to follow the development and updates of AI painting tools.

After determining the manual questionnaire, we repeated the steps from the first study to generate a questionnaire based on the integrated model and the research topic. First, we opened ChatGPT-4.0 and created a new conversation. Then, we sequentially input the prompts "Please introduce the UTAUT model," "Please introduce individual innovativeness," and "Please introduce AI anxiety," to which ChatGPT provided relevant introductions about the UTAUT model, individual innovativeness, and AI anxiety. Next, we input the research topic and the content needed: "I now want to explore the willingness of design major students in Chinese higher education to accept and use AI painting tools. Please incorporate individual innovativeness and AI anxiety from the information technology field into the UTAUT model as key variables and create a research questionnaire. Set three questions for each variable'." ChatGPT subsequently generated a questionnaire adopting a five-point Likert scale that was consistent with the manual questionnaire. The contents of the AI-generated questionnaire are shown in Table 4.

Table 4. College Students' Willingness to Learn AI Painting Tools Questionnaire (AI-generated).

Dimension	Items
Performance Expectancy	1. I believe using AI painting tools can improve my design efficiency.
	2. Using AI painting tools can help me create more creative design works.
	3. AI painting tools can bring substantial help to my design process.
Effort Expectancy	1. I believe that operating AI painting tools is simple and easy to understand.
	2. I believe learning to use AI painting tools won't take too much time and effort.
	3. I believe the functional interface design of AI painting tools is intuitive and easy to use.
Social Influence	1. My classmates encourage me to use AI painting tools.
	2. My teachers recommend using AI painting tools to improve design abilities.
	3. Many people in my social circle use AI painting tools, and I feel their user experience is positive.
Facilitating Conditions	1. I can access sufficient resources (such as computer equipment, software, etc.) to use AI painting tools.
	2. I have adequate technical support to use AI painting tools.
	3. My school provides the necessary facilities and environment for using AI painting tools.
Individual Innovativeness	1. I am willing to try new technologies and tools to enhance my design abilities.
	2. I like to try unfamiliar technologies, even if they bring some uncertainty.
	3. I usually become an early adopter, using newly launched technologies or tools.
AI Anxiety	1. I worry that AI painting tools might replace my creativity and design thinking.
	2. I feel uncertain and anxious about how AI painting tools process my creations.
	3. I am concerned that AI painting tools might infringe on my personal creative privacy or leak my design data.
Behavioral Intention	1. I plan to use AI painting tools in future design projects.
	2. I am willing to use AI painting tools in my future learning process.
	3. I believe AI painting tools will be an important part of my future design work.



3.4. Expert Evaluation Design

Expert evaluation is a commonly used method to support decision making, but evaluation schemes should be customized according to research questions and expert characteristics [27,28]. This study proposed the following selection criteria for experts: familiarity with the UTAUT model, experience in designing questionnaires, and conducting research based on theoretical models. Eventually, 12 experts were selected (two doctoral students, three lecturers, three associate professors, and four professors, including seven males and five females). An expert evaluation form (Table 5) was developed by integrating the existing questionnaire evaluation methods [29] and research purposes. This form used a 5-point Likert scale, with scores from 1 to 5 representing "strongly disagree," "disagree," "neutral," "agree," and "strongly agree." The evaluated questionnaires included four sets (single model manual questionnaire, AI questionnaire, integrated model manual questionnaire, and AI questionnaire). To ensure an objective and fair evaluation, the questionnaires were randomly numbered, with their sources concealed. The evaluation was conducted online using a Questionnaire Star platform.

Table 5. Expert Evaluation Form on Questionnaire Quality.

Dimension	Items
Accuracy	The questions accurately reflect the variables intended to be studied.
Comprehensiveness	The questionnaire questions comprehensively reflect each variable.
Clarity	The language of the questions is straightforward, concise, without ambiguity, and without complex expressions or difficult terminology.
Redundancy	Each question expresses a different meaning, with no repetition between different questions.
Order and Structure	The questions in the questionnaire are arranged in a logical order, guiding respondents to answer smoothly.
Objectivity	The questionnaire does not contain questions that create emotional or cognitive bias in respondents; all questions are neutral and fair.
Ethics and Privacy Protection	The questionnaire fully considers ethics and respondent privacy protection.

3.5. Deep Interviews

Six additional experts (two doctoral students, one lecturer, one associate professor, and two professors, including four males and two females) were selected for the in-depth interviews. With the experts' consent, the interviews were conducted face-to-face, lasting 30-50 minutes, with an average of approximately 40 minutes. Prior to the interviews, the experimental background and purpose were communicated to the experts, and agreement was reached on matters such as recording and confidentiality. The interview content covered aspects such as the advantages and limitations of the AI-generated questionnaires and their future application prospects. To ensure the completeness and accuracy of the materials, researchers promptly transcribed the recordings into text after each interview and carefully verified them to ensure consistency between the transcribed content and recordings. The transcribed texts were coded using a unified numbering system to facilitate subsequent references and analyses.

4. Results

4.1. Quantitative Research Results

To evaluate the differences between the AI-generated questionnaires and manual questionnaires under a single model, independent sample t-tests were conducted on the evaluation results of 12 experts. The results revealed significant differences between the AI and manual questionnaires in the following three dimensions: Accuracy ( $t=5.322$ ,  $p<0.001$ ), comprehensiveness ( $t=6.127$ ,  $p<0.001$ ), and clarity ( $t=5.409$ ,  $p<0.001$ ). However, there were no significant differences in the four dimensions: redundancy ( $t=1.787$ ,  $p=0.088$ ), order and structure ( $t=0.793$ ,  $p=0.436$ ), objectivity ( $t=-0.348$ ,  $p=0.731$ ), and ethics and privacy protection ( $t=0.842$ ,  $p=0.409$ ). As shown in Table 6.

Table 6. Analysis Results of Single Model Questionnaire Evaluation.

Dimension	Type	Mean±SD	t	p
Accuracy	Manual	4.00±0.85	5.322	0.000**
	AI	2.17±0.83		
Comprehensiveness	Manual	4.00±0.74	6.127	0.000**
	AI	2.08±0.79		
Clarity	Manual	4.08±0.67	5.409	0.000**
	AI	2.25±0.97		
Redundancy	Manual	4.17±0.58	1.787	0.088
	AI	3.67±0.78		
Order & Structure	Manual	4.58±0.51	0.793	0.436
	AI	4.42±0.51		
Objectivity	Manual	4.33±0.65	-0.348	0.731
	AI	4.42±0.51		
Ethics & Privacy Protection	Manual	4.75±0.45	0.842	0.409
	AI	4.58±0.51		

Similarly, independent samples t-tests were conducted on the evaluation results of the integrated model questionnaires. The results showed that AI and manual questionnaires had significant differences in the following four dimensions: accuracy (t=4.811, p<0.001), comprehensiveness (t=5.244, p<0.001), clarity (t=9.601, p<0.001), and redundancy (t=8.509, p<0.001). There were no significant differences in the three dimensions: order and structure (t=0.632, p=0.534), objectivity (t=-0.327, p=0.747), and ethics and privacy protection (t=0.793, p=0.436). As shown in Table 7.

Table 7. Analysis Results of Integrated Model Questionnaire Evaluation.

Dimension	Type	Mean±SD	t	p
Accuracy	Manual	4.25±0.87	4.811	0.000**
	AI	2.42±1.00		
Comprehensiveness	Manual	4.33±0.65	5.244	0.000**
	AI	2.67±0.89		
Clarity	Manual	4.50±0.52	9.601	0.000**
	AI	2.25±0.62		
Redundancy	Manual	4.08±0.51	8.509	0.000**
	AI	2.33±0.49		
Order & Structure	Manual	4.42±0.67	0.632	0.534
	AI	4.25±0.62		
Objectivity	Manual	4.08±0.79	-0.327	0.747
	AI	4.17±0.39		
Ethics & Privacy Protection	Manual	4.58±0.51	0.793	0.436
	AI	4.42±0.51		

4.2. Qualitative Research Results

4.2.1. Qualitative Research Results for Single Theoretical Model

Through analysis of the interview materials, the performance of AI questionnaires and their differences from manual questionnaires were categorized into the following types, with interviewees' narratives interspersed in a contextualized manner during writing to restore their authentic feelings and experiences.

First, the experts pointed out that the AI questionnaires showed inadequate understanding of the variables. They mentioned that "AI questionnaires did not clearly articulate newly introduced variables and failed to accurately grasp the connotations of variables," and "although the questions

contained variables, the core being examined was not the variables themselves." For example, in the effort expectancy dimension, the item "I believe the interface design of the virtual simulation experiment platform is simple and intuitive," while interface design is one aspect of the usage process and is less important than factors such as operational difficulty. Additionally, some experts noted that "the questions excessively applied the model framework without flexible adjustments based on actual research needs." This was mainly reflected in the AI questionnaire covering not only "behavioral intention" but also "use behavior," while the research purpose actually only needed to examine "behavioral intention."

Second, experts believed that the questions in the AI questionnaires were not sufficiently comprehensive. They suggested that "the questions provided by AI only addressed certain aspects of the variables, with many angles not yet considered," and "the questions only examined part of the content of the variables, with insufficient coverage." For example, in the social influence dimension, AI-generated questions mainly focused on teachers, classmates, and peers without covering other possible social influence factors.

Third, experts found that the expressions of questions in the AI questionnaires were not sufficiently clear. They pointed out that "the expression of AI questionnaire questions was too general" and "the expression of questions was not concise enough." For example, in the performance expectancy dimension, the item "I believe the virtual simulation experiment platform provides practical help for my medical studies," where the scope of "provides practical help" is too broad and should use a more specific expression. Additionally, some experts mentioned that "the expression of questions is not clear enough, which may lead to multiple interpretations by different respondents'."

Besides pointing out the shortcomings of AI questionnaires, experts have also affirmed the excellent performance of AI questionnaires in some aspects. They believed that the structural design of AI questionnaires was reasonable and met the requirements of the theoretical frameworks. At the same time, the expression of questions in the AI questionnaires was objective and standardized without emotional bias or emotional guidance. They did not involve ethical issues or infringe on the privacy of the subjects.

#### 4.2.2. Qualitative Research Results for Integrated Theoretical Model

Through an analysis of the interview results for the integrated theoretical model, we found that AI questionnaires mainly had the following problems: insufficient understanding of variables, non-comprehensive examination of questions, and unclear expression of questions. These findings are consistent with those found in a single theoretical model.

In addition, the experts pointed out that the questions in the AI questionnaires had problems with repetitive meanings. They mentioned that "the meanings expressed between questions were similar, lacking clear distinction," and "although the wording of questions was different, the content they actually wanted to express was the same." For example, in the individual innovativeness dimension, the three questions "I am willing to try new technologies and tools," "I like to try unfamiliar technologies," and "I usually become an early adopter" have overlapping meanings. Moreover, some experts pointed out that "there was an inclusive relationship between AI questionnaire questions, with one question encompassing the content of other questions." For example, in the performance expectancy dimension, the third question "AI painting tools can bring substantial help to my design process," where "substantial help" already included "improving design efficiency" and "creating more creative design works" mentioned in the first two questions.

## 5. Discussion

### 5.1. Performance of AI in Generating Questionnaires Based on Single Theoretical Model

This study found that the AI and manual questionnaires showed significant differences in accuracy. Quantitative results indicated that the score difference between the two in this dimension reached a statistically significant level ( $t=5.322$ ,  $p<0.001$ ), and in-depth interviews also revealed the problem of the AI questionnaires' insufficient understanding of variables. This finding is consistent with previous

research conclusions that AI performs well in handling basic questionnaire problems, but tends to show understanding bias when facing complex problems [15]. The reason might be that AI relies mainly on learning from large amounts of training data to generate questions. For common models or research topics, AI demonstrates high accuracy in understanding; however, when facing complex models or novel topics, AI may find it difficult to deeply understand the connotations of problems owing to a lack of relevant learning experience.

Second, there were significant differences between the AI questionnaires and manual questionnaires in comprehensiveness. The quantitative results showed that the score difference between the two in this dimension was statistically significant ( $t=6.127$ ,  $p<0.001$ ). Qualitative research results also pointed out that AI-generated questions often only address certain aspects of variables, without comprehensively covering their connotations. This echoes the findings of previous research [30], possibly stemming from AI's tendency of AI to simplify variable connotations when processing complex theoretical models to pursue higher matching with relevant samples, ultimately leading to a partial understanding of variables.

Third, in terms of clarity, there were significant differences between the AI and the manual questionnaires. The score difference between the two in the clarity dimension was statistically significant ( $t=5.409$ ,  $p<0.001$ ). In qualitative interviews, experts expressed that the questions in the AI questionnaires were not straightforward and easy to understand for the target audience and could easily lead to multiple interpretations. This phenomenon might stem from the following reasons: first, AI tends to follow standardized language rules and formal academic expressions, possibly neglecting the colloquial nature and ease of understanding in questionnaire item expressions; second, AI might lack sufficient consideration of the target group's background and comprehension abilities, failing to adjust question expression methods according to audience characteristics. This finding supplements previous research on AI filling out questionnaires [14].

We also made unexpected discoveries. The AI-generated questionnaire not only included "behavioral intention" but also covered "use behavior." Although the UTAUT model itself includes both variables, the research topic only needs to explore medical students' willingness to use virtual simulation experiment platforms, not their actual usage behavior. Therefore, the manual questionnaire removed "use behavior," but the AI questionnaire retained this variable. This indicates that when the theoretical model and research topic are not completely aligned, AI tends to prioritize meeting the requirements of the theoretical model rather than the research topic. Additionally, the three items in the "facilitating conditions" dimension of the manual questionnaire adopted a reverse questioning approach, whereas the AI questionnaire used positive questioning. In questionnaire design, reverse questioning is usually used to detect whether respondents are answering seriously. If respondents do not carefully read the questions, they might give answers completely opposite to their actual situation. Unless specifically requested by researchers, AI typically does not use reverse questioning proactively. This finding is consistent with previous research [24], which indicates that, compared to AI, manually designed questionnaires have greater flexibility, whereas AI usually relies on understanding data and template-based rules.

At the same time, quantitative research results showed that there were no significant differences between AI and manual questionnaires in three dimensions: order and structure ( $t=0.793$ ,  $p=0.436$ ), objectivity ( $t=-0.348$ ,  $p=0.731$ ), and ethics and privacy protection ( $t=0.842$ ,  $p=0.409$ ). Notably, in terms of objectivity, the average AI questionnaire score was even higher than that of manual questionnaires, making it the only dimension among the seven, where AI questionnaires outperformed manual questionnaires. The reason for this phenomenon might be that AI's generation logic is based on neutral data, effectively avoiding emotional bias and thus possessing higher objectivity. This finding suggests that AI might have advantages in questionnaire design tasks that require high standardization and neutrality.

## 5.2. Performance of AI in Generating Questionnaires Based on Integrated Theoretical Models

The quantitative results showed that under integrated models, AI questionnaires and manual questionnaires had significant differences in four aspects: accuracy, comprehensiveness, clarity, and redundancy. Unlike the single models, AI questionnaires and manual questionnaires also showed significant differences in redundancy under the integrated models ( $t=8.509$ ,  $p<0.001$ ). Qualitative analysis results also

indicated that AI questionnaires based on integrated models exhibited repetitive meanings in the questions. The reasons for the significant differences in the first three aspects have been discussed in the previous section and are not repeated here. Combined with previous research [31], the significant difference in redundancy in integrated models may be because to ensure comprehensive coverage, questions designed by AI typically include more information and background. As model complexity increases, AI's ability to understand decreases, making it difficult to break down variables from multiple angles, and it can only express the same meaning by changing wording.

Consistent with the single models, there were no significant differences between AI questionnaires and manual questionnaires in three dimensions: order and structure, objectivity, and ethics and privacy protection. Whether with single models or integrated models, AI questionnaires performed the same as manual questionnaires in these three aspects and remained relatively stable. Therefore, when researchers use AI to design questionnaires, they need not invest excessive effort or make large-scale modifications to these three areas.

## 6. Conclusion

This study employed a hybrid research design combining quantitative expert evaluation and qualitative in-depth interviews to comprehensively explore the performance of AI-generated questionnaires based on single and integrated theoretical models, comparing them with manually created questionnaires. The research found that, under a single model, AI questionnaires showed significant differences from manual questionnaires in three dimensions: accuracy, comprehensiveness, and clarity. AI was limited when handling complex and abstract variables, and manual questionnaires would use reverse questioning to check whether respondents were filling out seriously, while AI rarely proactively used this approach. Furthermore, when the theoretical model and research topic were not completely aligned, AI tended to satisfy the model rather than topic requirements. In the integrated models, there were significant differences between the two in four dimensions: Accuracy, comprehensiveness, clarity, and redundancy. As the model complexity increased, the probability of problems in the AI questionnaires correspondingly increased. At the same time, regardless of whether single or integrated models were used, there were no significant differences between the two in three dimensions: order and structure, objectivity, and ethics and privacy protection, with AI questionnaires even outperforming manual questionnaires in objectivity. This indicates that AI has advantages in questionnaire design, as it requires high standardization and neutrality.

These results suggest that although AI-generated questionnaires have advantages, such as fluent language and efficiency, they still have limitations when processing complex and abstract research variables. Researchers can use AI for initial questionnaire generation and then manually modify the questionnaire according to research purposes, with human judgment being indispensable in this process. To further improve the quality of AI-generated questionnaires, future research could combine expert feedback and theoretical support to improve AI models' contextual understanding capabilities and generation logic, achieving a more precise, comprehensive, and easily understood questionnaire design. With continuous technological advancements, AI is expected to play an increasingly important role in future academic research.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data is unavailable due to privacy or ethical restrictions.

**Acknowledgments:** During the preparation of this manuscript/study, the author(s) used ChatGPT for the purposes of AI-assisted questionnaire generation. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Wang C. AI-driven digital image art creation: Methods and case analysis[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(3): 406-414.



2. Zhang C, Lu Y. Study on artificial intelligence: The state of the art and future prospects[J]. *Journal of Industrial Information Integration*, 2021, 23: 100224.
3. Min B, Ross H, Sulem E, et al. Recent advances in natural language processing via large pre-trained language models: A survey[J]. *ACM Computing Surveys*, 2023, 56(2): 1-40.
4. Xiangdong L I, Hanfei X I A, Yifei S, et al. Opportunities and challenges: automatic generation technologies for graphical user interfaces[J]. *Journal of Graphics*, 2024, 45(3): 409.
5. Genelza G G. A systematic literature review on AI voice cloning generator: A game-changer or a threat?[J]. *Journal of Emerging Technologies*, 2024, 4(2): 54-61.
6. Ayyagari K C. Rapid Video Prototyping and Content Creation Using Generative Artificial Intelligence[J]. 2024.
7. Liang J, Shan X, Chung J. A Study on Process of Creating 3D Models Using the Application of Artificial Intelligence Technology[J]. *International Journal of Advanced Culture Technology*, 2023, 11(4): 346-351.
8. Zhang D, Yu Y, Dong J, et al. Mm-llms: Recent advances in multimodal large language models[J]. *arXiv preprint arXiv:2401.13601*, 2024.
9. AlZaabi A, ALamri A, Albalushi H, et al. ChatGPT applications in academic research: a review of benefits, concerns, and recommendations[J]. *Biorxiv*, 2023: 2023.08.17.553688.
10. CURTIS N, CHATGPT, 2023. To ChatGPT or not to ChatGPT? the impact of artificial intelligence on academic publishing [J]. *The Pediatric Infectious Disease Journal*, 42(4), 275.
11. ISKENDER A, 2023. Holy or unholy? interview with open AI's ChatGPT [J]. *European Journal of Tourism Research*, 34, 3414-3414.
12. LI X, HU L, SHANG X, et al. Reconstructing the Paradigm of Sports Science Research through Large Language Models: Applications and Insights—A Case Study of ChatGPT[J]. *Journal of Xi'an Physical Education University*, 2023, 40(04): 405-415. DOI:10.16063/j.cnki.issn1001-747x.2023.04.003.
13. Wang Y, Guo X, Liu Z. Detection and comparative study of differences between AI-generated and scholar-written Chinese abstracts[J]. *J Intell*, 2023, 42(9): 127-134.
14. Lebrun B, Temtsin S, Vonasch A, et al. Detecting the corruption of online questionnaires by artificial intelligence[J]. *Frontiers in Robotics and AI*, 2024, 10: 1277635.
15. Olivos F, Liu M. ChatGPTtest: opportunities and cautionary tales of utilizing AI for questionnaire pretesting[J]. *Field Methods*, 2024: 1525822X241280574.
16. Abu-Dalbouh H M. A questionnaire approach based on the technology acceptance model for mobile tracking on patient progress applications[J]. *J. Comput. Sci.*, 2013, 9(6): 763-770.
17. Putz D, Schilling J, Kluge A, et al. Measuring organizational learning from errors: Development and validation of an integrated model and questionnaire[J]. *Management learning*, 2013, 44(5): 511-536.
18. Goodfellow I. Pouget - Abadie J, Mirza M, et al. Generative adversarial nets[J]. *Advances in neural information processing systems*, 2014, 27: 2672-2680.
19. Clark A, Donahue J, Simonyan K. Adversarial video generation on complex datasets[J]. *arXiv preprint arXiv:1907.06571*, 2019.
20. Tsidylo I M, Sena C E. Artificial intelligence as a methodological innovation in the training of future designers: Midjourney tools[J]. *Information Technologies and Learning Tools*, 2023, 97(5): 203.
21. OpenAI. GPT - 4 [ EB/ OL]. [2023 - 04 - 27] <https://openai.com/research/gpt-4>.
22. Dowling M, Lucey B. ChatGPT for (finance) research: The Bananarama conjecture[J]. *Finance Research Letters*, 2023, 53: 103662.
23. Ramos M C, Collison C J, White A D. A review of large language models and autonomous agents in chemistry[J]. *Chemical Science*, 2025.
24. ZUO B, WU J, YANG Y. ChatGPT as a Questionnaire Tool for Tourism Research: Advantages and Limitations[J/OL]. *Tourism Science*, 1-11[2025-02-24]. <https://doi.org/10.16323/j.cnki.lykx.20241205.001>.
25. Zhao, H., Wang, Q., Zhou, T., et al. Research on willingness of medical students to use virtual simulation experiment platform based on the UTAUT model. *Disease Prevention and Control Bulletin*, 1-5. [Online]. Available: <https://doi.org/10.13215/j.cnki.jbyfkztb.2401039>.

26. WANG C. Study on the Willingness of Chinese University Students to Learn AI Painting Tools Based on an Extended UTAUT Model[J/OL]. Packaging Engineering, 1-13 [2025-02-24]. <http://kns.cnki.net/kcms/detail/50.1094.TB.20241210.1330.002.html>.
27. Meyer M A, Booker J M. Eliciting and analyzing expert judgment: a practical guide[M]. Society for industrial and applied mathematics, 2001.
28. BAUM S D, GOERTZEL B, GOERTZEL T G. How long until human-level AI? Results from an expert assessment[J].
29. Technological Forecasting and Social Change, 2011, 78(1): 185-195.Olson K. An examination of questionnaire evaluation by expert reviewers[J]. Field methods, 2010, 22(4): 295-318.
30. Zou Z, Mubin O, Alnajjar F, et al. A pilot study of measuring emotional response and perception of LLM-generated questionnaire and human-generated questionnaires[J]. Scientific reports, 2024, 14(1): 2781.
31. Kunst J R, Bierwiazzonek K. Utilizing AI questionnaire translations in cross-cultural and intercultural research: Insights and recommendations[J]. International Journal of Intercultural Relations, 2023, 97: 101888.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.