

Women's Preferences and Willingness to Pay for Artificial Intelligence Chatbots in Women's Health: A Discrete Choice Experiment Study

Jing Wang, Hewei Min, Tao Li, Jiaheng Li, Yang Jiang, Jingbo Zhang, Yibo Wu,
Xinying Sun

Submitted to: Journal of Medical Internet Research
on: October 08, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
---------------------------------	----------

Preprint
JMIR Publications

Women's Preferences and Willingness to Pay for Artificial Intelligence Chatbots in Women's Health: A Discrete Choice Experiment Study

Jing Wang¹; Hewei Min¹; Tao Li²; Jiaheng Li³; Yang Jiang⁴; Jingbo Zhang⁵; Yibo Wu¹; Xinying Sun¹

¹ School of Public Health, Peking University, Beijing, China Beijing CN

² Department of gynaecology and obstetrics, The Fourth Central Hospital of Baoding City, Hebei, China; Hebei CN

³ School of Basic Medical Sciences, Hebei University, Hebei, Baoding, China; Hebei CN

⁴ North China University of Science and Technology, Hebei, Tangshan, China; Hebei CN

⁵ Center for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China Beijing CN

Corresponding Author:

Xinying Sun

School of Public Health, Peking University, Beijing, China
38 Xueyuan Road, Haidian District, Beijing, China
Beijing
CN

Abstract

Background: Over 96% of adult women face health issues, with 70% experiencing conditions like infections. Mobile health education is increasingly popular but faces challenges in personalization and readability. Artificial intelligence chatbots provide tailored support, and a Discrete Choice Experiment can help in understanding user preferences to improve chatbot design.

Objective: This study explores the preferences of women towards artificial intelligence chatbots to improve health education communication, user experience, and health literacy.

Methods: A discrete choice experiment was conducted, identifying six main attributes of artificial intelligence chatbots: response accuracy, legibility, service cost, background information collection, information utility, and content provision. 957 participants from a hospital in Hebei Province participated, choosing between two hypothetical chatbots or opting for neither (a no-choice option). The conditional logit model was used to estimate user preferences.

Results: Including 957 participants for analysis, participants are more willing to choose a chatbot with 100% response accuracy($\beta=0.940$, $P<.001$), very easy to understand information($\beta=0.907$, $P<.001$), service fee of ¥0/month ($\beta=-0.095$, $P<.001$), information utility is very practical($\beta=1.085$, $P<.001$) and the provision of information content is disease knowledge($\beta=0.752$, $P<.001$). Whether or not to allow the collection of background information (only question-answering information) has no significant impact on women's choice preferences.

Conclusions: Artificial intelligence chatbots should focus on high response accuracy, clear content, free access, privacy protection, practical information, and disease knowledge to attract users and enhance health education.

(JMIR Preprints 08/10/2024:67303)

DOI: <https://doi.org/10.2196/preprints.67303>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org>



Original Manuscript

Women's Preferences and Willingness to Pay for Artificial Intelligence Chatbots in Women's Health: A Discrete Choice Experiment Study

Jing Wang^{a†}, Hewei Min^{a†}, Tao Li^b, Jiaheng Li^c, Yang Jiang^d, Jingbo Zhang^e, Yibo Wu^a □ Xinying Sun^{a*}

^a School of Public Health, Peking University, Beijing, China;

^b Department of gynaecology and obstetrics, The Fourth Central Hospital of Baoding City, Hebei, China;

^c School of Basic Medical Sciences, Hebei University, Hebei, Baoding, China;

^d North China University of Science and Technology, Hebei, Tangshan, China;

^e Center for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China;

* xysun@bjmu.edu.cn;

Tel: [+86 13691212050](tel:+8613691212050)

ORCID: [0000-0001-6638-1473](https://orcid.org/0000-0001-6638-1473)



Authors contributed equally to this work.

Abstract

Background

Over 96% of adult women face health issues, with 70% experiencing conditions like infections. Mobile health education is increasingly popular but faces challenges in personalization and readability. Artificial intelligence chatbots provide tailored support, and a Discrete Choice Experiment can help in understanding user preferences to improve chatbot design.

Objective

This study aims at exploring the preferences of women towards artificial intelligence chatbots to improve health education communication and user experience.

Methods

A discrete choice experiment was conducted, identifying six main attributes of artificial intelligence chatbots: response accuracy, legibility, service cost, background information collection, information utility, and content provision. 957 female participants from a hospital in Hebei Province participated, choosing between two hypothetical chatbots or opting for neither (a no-choice option). The conditional logit model was used to estimate user preferences.

Results

A total of 957 participants were included in the analysis. The results showed that participants preferred a chatbot with 100% response accuracy ($\beta = .940$, $P < .001$, 95%CI: 0.624–1.255), very

easy to understand information ($\beta = .907$, $P < .001$, 95%CI: 0.634–1.180), a service fee of ¥0/month ($\beta = -.095$, $P < .001$, 95%CI: -0.108– -0.082), practical information utility ($\beta = 1.085$, $P < .001$, 95%CI: 0.832–1.338), and provision of disease-related knowledge ($\beta = .752$, $P < .001$, 95%CI: 0.485–1.018). Whether or not to allow the collection of background information (only question-answering information) has no significant impact on women's choice preferences. Additionally, participants were willing to pay an additional ¥9.916 (95% CI: 6.843–12.292) for 100% response accuracy, ¥9.567 (95% CI: 6.843–12.292) for "very easy to understand" information, and ¥11.451 (95% CI: 8.704–14.198) for "very practical" information utility. Additionally, they were willing to pay ¥7.931 (95% CI: 4.975–10.886) for "knowledge of diseases" compared to "gender knowledge" (¥2.602, 95% CI: -0.551–5.756). The relative importance of the chatbot attributes indicated that information utility (28.12%) and response accuracy (24.37%) were the most influential factors in participants' preferences.

Conclusion

Artificial intelligence chatbots designed for female users should focus on high response accuracy, clear content, free access, privacy protection, practical information, and disease knowledge to attract users and enhance health education.

Keywords: women's health; artificial intelligence chatbots; discrete choice experiment; willingness to pay; health education

Introduction

According to the United Nations (2023), women make up 49.75% of the global population, yet their health needs are often overlooked[1]. In response, the World Health Organization (WHO) identified six key priorities for women's health in 2021, aiming to address the numerous challenges faced by women and girls in accessing adequate healthcare (WHO, 2023a)[2]. These priorities included ensuring universal access to sexual and reproductive health services and tackling noncommunicable diseases (NCDs), which remain a leading cause of preventable death among women globally (WHO). The agenda also focused on eliminating violence against women and preventing NCDs, including obesity. Furthermore, neglecting women's health could jeopardize the well-being of future generations.

Compared to other health conditions, issues related to women's health are generally more preventable and sensitive. Effective health education programs play a key role in preventing these conditions, while accurate health knowledge acquisition helps mitigate disease progression[3]. Combining systematic education with evidence-based knowledge dissemination not only benefits the women themselves but also has significant implications for society as a whole.

There are various methods to implement health education for women. These include regular physical examinations with structured educational components, distributing scientifically validated informational materials, and offering medical consultations integrated with knowledge-sharing sessions. Knowledge dissemination is also achieved through special television programs and radio segments[4] that translate complex medical information into accessible content. With the advancement of technology, the paradigm of health knowledge acquisition has been transforming. Mobile health education, based on applications and social media platforms, facilitates knowledge transfer while developing self-care skills among women[5-10]. In contrast to traditional methods, mobile health education offers greater accessibility, personalization, and cost-effectiveness, overcoming time and geographical barriers[11-13].

Mobile health education still faces challenges in pedagogical design and knowledge delivery. Educational effectiveness depends on content diversity, while knowledge retention requires information accuracy. For educational interventions targeting vulnerable groups, designing pedagogically sound content with actionable knowledge points is crucial. In this context, artificial intelligence (AI) chatbots emerge as a promising solution that may combine interactive education with personalized knowledge delivery[14]. AI, which comprises programs and algorithms that perform tasks traditionally requiring human intelligence, is increasingly applied in healthcare, with potential in both specialized and general settings[15,16]. AI chatbots, as one of the key AI applications, offer interactive platforms for delivering health information and services[17]. Using natural language processing (NLP)[18], these systems convert user inquiries into a machine-readable format, leveraging a broad spectrum of medical data to establish a knowledge base that supports health-related interactions[19]. AI chatbots could potentially offer appropriate health advice and facilitate health discussions, thereby supporting communication and patient empowerment[20]. With the expansion of mobile internet access, AI chatbots are being explored as a practical and accessible tool for women with limited healthcare access, enhancing patient-centered care and self-management[21]. Their use notably increased during the COVID-19 pandemic, demonstrating some positive support in various health domains[17].

Specifically, AI chatbots have the potential to restructure health education frameworks by providing tailored guidance, while establishing reliable knowledge repositories on public platforms[22,23]. For instance, **Rosie**, a health education chatbot, is a great example. Rosie aims to provide reliable health information to new mothers, helping them address various issues during pregnancy and postpartum. Through interactions with the community, Rosie has not only enhanced users' access to health information but also helped to narrow the health disparities between different racial groups to some extent. [24]. Maeda et al. conducted a randomized controlled trial with 927 women to assess the effects of a fertility education chatbot, finding it significantly improved fertility knowledge and preconception behaviors. The chatbot also reduced anxiety, demonstrating its potential to educate and empower women without increasing stress. [25]. Additionally, a review by Kim HK examined the broader effects of AI chatbots on women's health, including mental health, cancer self-care, and preconception intentions. The meta-analysis revealed that chatbot interventions significantly reduced anxiety, showing the positive impact of such technologies on physical, physiological, and cognitive health outcomes. This suggests that AI chatbots can serve as an effective tool in addressing various health concerns and providing digital therapeutic support to women[26].

However, despite these advancements, there has been insufficient exploration of how these AI tools can be customized for specific gynecological needs. AI chatbots that fully consider the needs and preferences of users may better enhance the effectiveness of interventions[27]. Previous studies have shown that people tend to care about the accuracy of medical chatbot responses and are concerned about whether their privacy is being violated[28-30]. A review of mental health chatbots indicated that patients generally find the usefulness and ease of use of these chatbots to be positive, but were less satisfied with the language expression and response quality of the chatbots. Some chatbots were considered slow in response, with answers that were often superficial, brief, and sometimes even confusing[31]. Moreover, another review revealed that the personalization of chatbots—such as in response results, content, user interfaces, delivery channels, and features—was an important factor in improving user satisfaction[32]. In summary, research on AI needs and preferences has mostly focused on the general population or adolescents, with some involvement in the development and research of mental health chatbots[27,31,33,34]. However, most studies have concentrated on English-language chatbots and lack in-depth investigation into preference differences among various demographic groups, which could potentially limit the benefits that different populations can derive from these technologies.

This study aims to explore Chinese women's preferences regarding the attributes of AI chatbots. To improve user experience and the accuracy of AI chatbot information, it is crucial to understand user preferences. A suitable method for measuring these preferences is the Discrete Choice Experiment (DCE), a robust survey technique that presents respondents with multiple choices among hypothetical treatments[35]. The choices made by respondents are influenced by the alternatives presented. By offering options that consist of various attributes at differing levels and repeatedly asking participants to select their preferred option or the one that maximizes utility[36], we seek to understand how users may value different aspects of AI chatbot use in women's health. Quantifying these preferences enables the identification of preferred attributes and provides insight into how women perceive and value AI chatbots for their health needs[37].

Methods

Study Design

In this study, we investigated women's preferences for attributes of the AI chatbots using the discrete choice experiment (DCE) method. DCE is a questionnaire-based stated preference approach to study people's preferences for certain characteristics or conditions by asking subjects to make choices about hypothetical scenarios or alternatives they prefer[38,29], DCE can be used to obtain quantitative data about people's preferences, gain in-depth insights and predictive power, and is often applied in market research, social sciences, medicine and other fields. Selected attributes in the DCE scheme can contain price attributes for different price levels to estimate WTP[40].

Development of the Discrete Choice Experiment

In this study, to ensure that the selected attributes comprehensively reflect user preferences for AI chatbots and have broad representativeness, we followed a systematic process involving a preliminary literature review, expert consultations, and a pre-survey.

First, we conducted a preliminary literature review by searching multiple academic databases (such as Google Scholar, Web of Science, PubMed, etc.) using relevant keywords like "AI chatbots," "user preferences," and "digital health," focusing on studies published between 2015 and 2022. We selected studies that focused on female participants, including both quantitative and qualitative research that explored the characteristics of AI chatbots and users' responses to them. This ensured that the literature reviewed was relevant to the female population, providing a solid theoretical foundation for identifying potential attributes that could influence user choices[28, 30, 41-45]. Based on this review, we moved to the expert consultation phase.

We conducted three rounds of expert consultations, with each round having specific objectives and outcomes. In the first round, we brought together a multidisciplinary team of 10 experts from fields such as law, ethics, artificial intelligence, and health data management. During this round, we presented the initial set of attributes identified from the literature review and asked the experts to review and comment on their relevance, clarity, and comprehensiveness. Their feedback led to the removal of certain attributes that were deemed too general or not actionable. In the second round, the attributes were refined based on the feedback from the first round, and the experts were asked to further discuss the attributes' levels, ensuring that they were clear and easy to understand for the target audience. This round also involved developing concrete examples for some attributes to help experts understand their real-world applications. In the third and final round, the experts were asked to validate the revised attributes and provide final input on their importance and whether they sufficiently reflected user preferences. The outcome of this round was the identification of six core attributes, with final adjustments made to the descriptions and levels to make them as actionable and

understandable as possible.

After the expert consultations, we proceeded with a pre-survey to test the applicability of the selected attributes within the target group. The survey was conducted with 100 women, who were recruited from a hospital in Hebei Province. The purpose of this pre-survey was to gauge how well the attributes were understood, ranked, and valued by potential users. Through this survey, we collected feedback on the clarity of the attribute descriptions and the relative importance of each attribute. The results indicated that some attribute descriptions were unclear or too complex, requiring further revision. Based on the pre-survey feedback, we modified the wording of the attributes to ensure that they were easily understandable and accurately captured the participants' preferences.

Through these three comprehensive validation steps—literature review, expert consultation, and pre-survey—we ultimately identified six key attributes that would guide the design of the AI chatbot: (1) response accuracy, (2) legibility, (3) service cost, (4) whether to allow background collection of information (question-answering information only), (5) information utility, and (6) provision of information content.

Each attribute and its level are shown in Figure 1.

Attribute	Attribute Level *
Response accuracy	60%; 70%; 80%; 90%; 100%
Legibility	Very easy to understand; Easier to understand; Harder to understand; Difficult to understand
Service cost	¥0/month; ¥5/month; ¥10/month; ¥15/month; ¥20/month
Whether to allow the collection of background information (only question-answering information)	Yes; No
Information Utility	Very practical; More practical; Less practical; Very impractical
Provision of information content	Knowledge of diseases; Daily health care; Knowledge of maternity; Gender knowledge

* Refers to the different levels or variations of each attribute used in the discrete choice experiment.

Figure 1: Attributes and levels of the chatbot.

Based on the identified attribute levels, we created choice sets containing various attribute levels for respondents to select from through orthogonal tests. Given that there are 2 to 5 attribute levels for each of the 6 attributes in our study, employing a full factorial design would result in 3,200 combinations ($5 \times 4 \times 5 \times 2 \times 4 \times 4 = 3,200$), which is clearly impractical for generating such a large number of choice sets. Consequently, we utilized a fractional factorial design to determine the optimal number of choice sets. Based on the principles of orthogonality, balance, and minimal overlap, 25 proposals were ultimately generated. Selecting one of them as a fixed reference, along with the remaining 24 options, forms a total of 24 selection sets. Each selection set contains three conceptual options ("choose option A"; "choose option B"; "choose neither"). Randomly divide all selection sets into 4 groups, each group consisting of 7 selection sets (6 random selection sets, 1 repeated selection set). Respondents selected their preferred combination of attributes for the AI chatbots in each scenario or opted for none, thereby minimizing bias resulting from forced choices.

According to the sample size formula for discrete choice experiments (DCE): $n = (500c)/(ta)$, where "c" represents the largest number of levels for an attribute, "t" denotes the number of choice sets in a block, and "a" indicates the number of alternatives. For this study, the values were as follows: "c" was set to 5, "t" was set to 6, and "a" was set to 2. Therefore, the minimum required sample size for this study was determined to be 209 participants. Figure 2 shows an example of a DCE program.

Please choose your preferred option from Option A and Option B. If you don't like either option, choose "None of them".

Attribute	Option A	Option B
Response accuracy	70%	100%
Legibility	easier to understand	harder to understand
Service cost	¥10/month	¥0/month
Whether to allow the collection of background information (only Q&A information)	No	No
Information Utility	Very practical	Very practical
Provision of information content	Knowledge of diseases	Knowledge of diseases

☐ Option A
 ☐ Option B
 ☐ None of them

Figure 2: An example scenario of the choice-based conjoint in the questionnaire.

Study population and data collection

This study began in December 2022 and involved a questionnaire survey of women at a hospital in Hebei Province, lasting for 2.5 months. The recruitment period started on December 20, 2022 and ended on February 15, 2023. Convenience sampling was employed, with gynecologists recruiting eligible female patients from both the gynecology outpatient department and the gynecology inpatient ward. Research staff explained the purpose of the study to eligible women and invited them to participate. Research information was collected through electronic questionnaires, which were divided into two sections: the first section gathered basic demographic information, including gender, age, education level, and usual residence, while the second section assessed preferences regarding AI chatbots. The electronic questionnaire was distributed via the wjx platform (<https://www.wjx.cn/>). After patients agreed to participate and completed the informed consent form, they filled out the questionnaire online by scanning a QR code. The inclusion and exclusion criteria are as follows:

Inclusion criteria:

1. Age ≥ 18 years.
2. Sex: Female.
3. Possesses the nationality of the People's Republic of China.
4. Own a smartphone and have the ability to use WeChat, including following WeChat public accounts.

5. Ability to complete web-based questionnaires independently or with assistance from other.
6. Possess basic literacy skills to enable normal communication and interaction.
7. Voluntary participation in the study, with willingness to complete an informed consent form.

Exclusion Criteria:

1. Individuals with severe cognitive impairments or mental health conditions.
2. Individuals who are participating in other similar research projects.
3. Individuals who are unwilling to cooperate.

A total of 1,281 women were invited to participate, and 1,216 women agreed to participate, yielding a participation rate of 94.9%. The final sample size of 957 participants was determined after excluding unqualified questionnaires based on the exclusion criteria. This process is depicted in Figure 3, which illustrates the flowchart of the study recruitment and analysis.

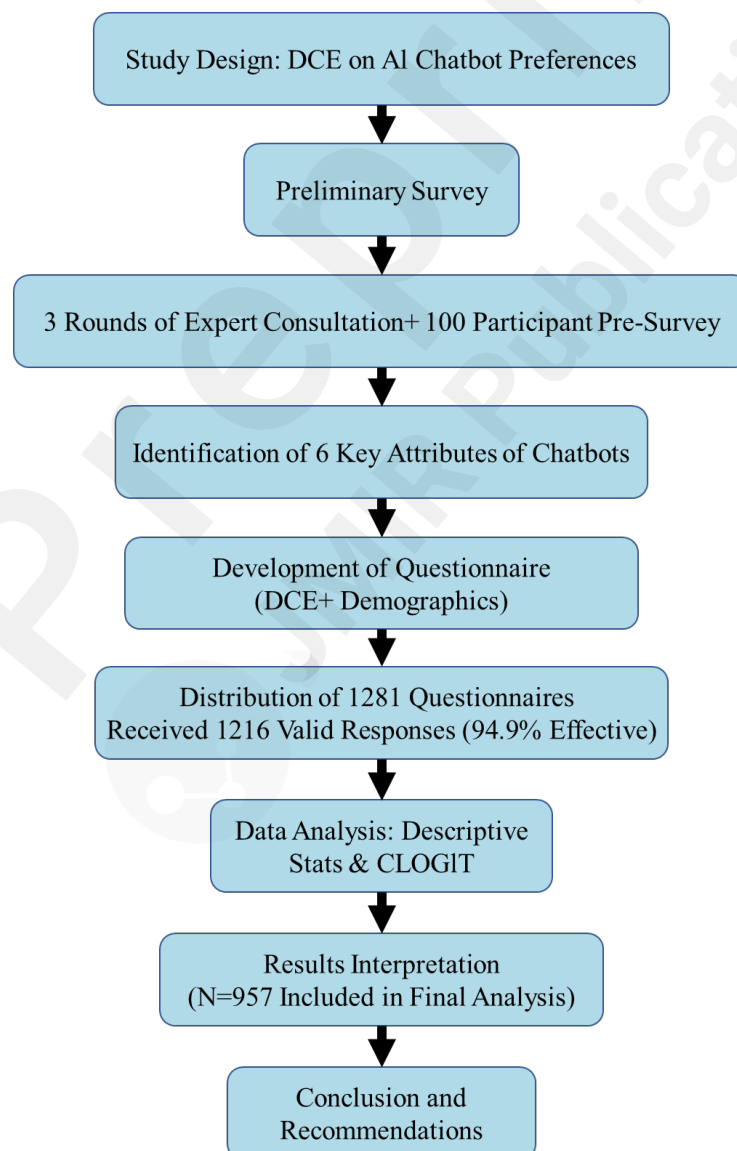


Figure 3. Flowchart of Study Process

Ethical Considerations

The study has been approved by the Ethics Committee of the Baoding No. 4 Central Hospital (approval number: 2022013). All study subjects have voluntarily participated and signed informed consent forms, ensuring that they are fully aware of the study's objectives, methods, and any potential risks. This consent was obtained prior to participation, in accordance with ethical standards for human subject research.

As the study involves human subjects, appropriate ethical oversight was conducted, and all research activities were approved by the Ethics Committee.

To ensure privacy and confidentiality, all study data are anonymized and de-identified to protect the identity of the participants. The data will be stored securely and will only be accessible to authorized personnel involved in the research.

No compensation is provided to the participants for their involvement in this study.

To maintain the confidentiality and anonymity of the participants, no identifiable images or data of individual participants will be included in the manuscript or supplementary materials.

Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics 26,0 (IBM Corporation, Armonk, NY, USA) and Stata 15.0. Descriptive statistics of demographic variables were performed by frequency counts (composition ratios). Conditional logit models (CLOGIT) were used to quantify the relative levels of attribute preferences used by the AI chatbots via Stata 15.0. Different levels of each attribute were dummy-coded and one of the levels was set as the reference level. In a Discrete Choice Experiment study, the calculated results provide important statistical information. These results include coefficients, p-values, standard errors, and 95% confidence intervals (95% CI). We also calculated the study participants' payment willingness (WTP) for different attribute levels of the AI chatbots to more intuitively reflect the strength of the respondents' attribute preferences for the AI chatbots.

Results

Characteristics of Respondents

The socio-demographic characteristics of the study sample are summarized in **Table 1**.

The age distribution showed that the largest group was between 26 and 35 years old (45.04%), followed by those aged 36 to 45 (29.26%). For further details on participant demographics, please refer to Table 1.

Table 1: General characteristics of the subjects(n=957)

Items	Number of cases (%)
Occupation	

	Employed	726 (75.86)
	Student	27 (2.82)
	Retired	11 (1.15)
	Unstable Occupation/Freelancer/Unemployed/Not in Employment	193 (20.17)
Location		
	Urban	633 (66.14)
	Rural	324 (33.86)
Age		
	18-25	159 (16.61)
	26-35	431 (45.04)
	36-45	280 (29.26)
	>45	87 (9.09)
Education level		
	Junior high school or lower	266 (27.80)
	Special school or senior high school	211 (22.05)
	Junior college or higher	480 (50.16)
Marital Status		
	Single	115 (12.02)
	Married (including first marriage, remarried, or remarried after divorce)	823 (86.00)
	Divorced	16 (1.67)
	Widowed	3 (0.31)
Income*		
	≤1000	66 (6.90)
	1001-3000	338 (35.32)
	3001-5000	386 (40.33)
	≥5000	167 (17.45)

* Per capita monthly household income

The Percent Importance of AI Chatbot Attributes

The percent importance reflects the extent to which each attribute influences the differences in preferences. It is calculated based on the range of preference weights across each attribute's levels, which provides a measure of its impact on decision-making. The sum of all percent importance is 100%, and a higher value corresponds to a greater influence on preferences. As shown in and Figure 4, the information utility of the chatbot is considered the most important attribute (28.123%), while cost and whether background information collection are the lowest, at 2.462% and 2.074%, respectively.

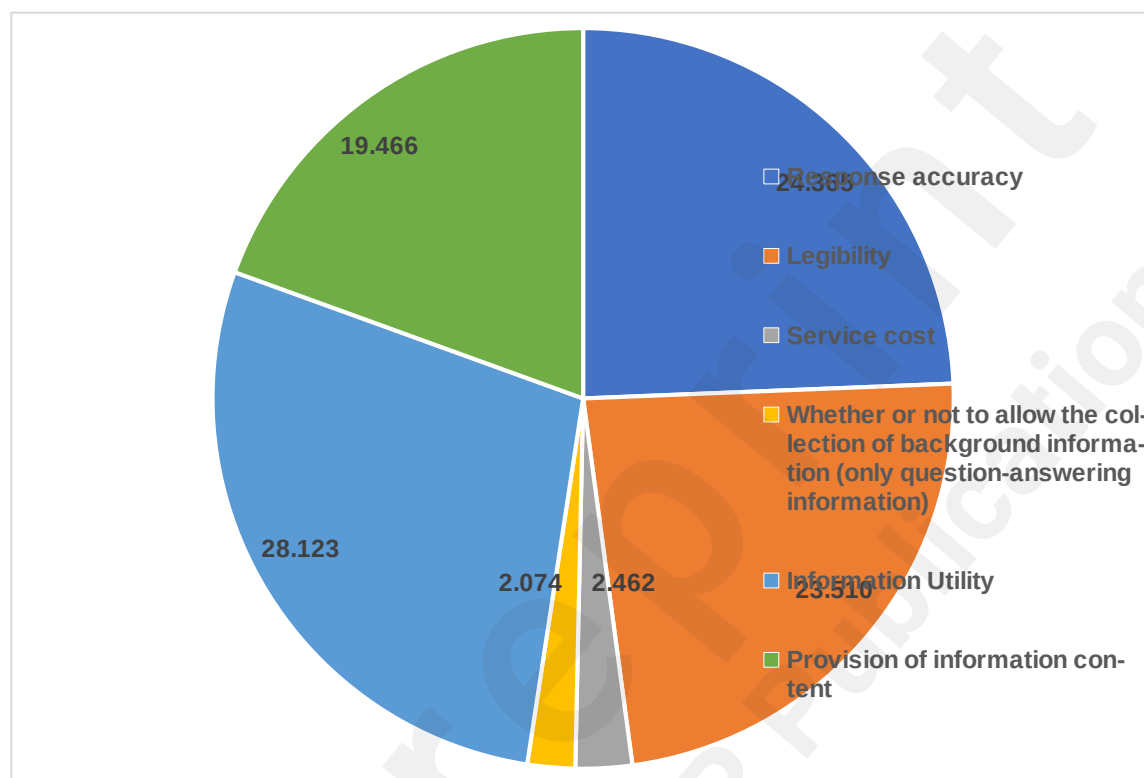


Figure 4. The Percent Importance of AI Chatbot Attributes. (N = 957)

DCE Results

The results of the CLOGIT model analysis indicate that the participants exhibited a preference for attributes such as 100% response accuracy, very easy to understand readability, ¥0/month service cost, very practical information utility, and provision of information content as knowledge of the disease in the AI chatbots. Table 2 presents detailed information on the preference attributes and their coefficients.

Table 2: CLOGIT results of participants' preferences for attributes of AI chatbots

Attributes and Levels	Coefficient t	P-value	Standard error	95% CI
Response accuracy				
60% ^a				
70%	.561	.001	0.168	0.232–0.891

80%	.511	.002	0.167	0.184–0.838
90%	.637	<.001	0.161	0.321–0.953
100%	.940	<.001	0.161	0.624–1.255
Legibility				
Harder to understand ^a				
Difficult to understand.	.207	.205	0.163	-0.113–0.528
Easier to understand	.725	<.001	0.137	0.456–0.993
Very easy to understand.	.907	<.001	0.139	0.634–1.180
Service cost	-.095	<.001	0.007	-0.108– -0.082
Whether or not to allow the collection of background information (only question-answering information)				
Yes ^a				
No	.080	.37	0.090	-0.095–0.254
Information Utility				
Very impractical ^a				
Less practical	.180	.241	0.153	-0.121–0.481
More practical	.584	<.001	0.164	0.262–0.907
Very practical	1.085	<.001	0.129	0.832–1.338
Provision of information content				
Gender knowledge ^a				
Knowledge of Maternity	.247	.108	0.153	-0.141–0.447
daily health care	.664	<.001	0.137	0.396–0.932
Knowledge of Diseases	.751	<.001	0.136	0.485–1.018

^a: Reference level

WTP Results

Willingness to pay (WTP) is a valid indicator of how much money a person is willing to sacrifice to choose one diagnostic attribute level over another. Table 3 presents the participants' willingness to pay for different attributes of the AI chatbot. The findings indicate that participants are willing to pay more for a women's health AI chatbot that provides accurate answers, clear content, does not collect

background information, has high information availability, and focuses on disease-related knowledge.

Table 3: Participants' willingness to pay for AI chatbots (WTP^a)

Attributes and Levels	WTP (CNY ^b)	95%CI
Response accuracy		
60%	REF	
70%	5.924	2.628–9.220
80%	5.389	2.034–8.744
90%	6.721	3.637–9.806
100%	9.916	6.843–12.292
Legibility		
Harder to understand	REF	
Difficult to understand.	2.188	-1.154–5.529
Easier to understand	7.646	4.883–10.409
Very easy to understand.	9.567	6.843–12.292
Whether or not to allow the collection of background information (only question-answering information)		
Yes	REF	
No	0.839	-1.017–2.695
Information Utility		
Very impractical.	REF	
Less practical	1.900	-1.278–5.078
More practical	6.167	2.701–9.633
Very practical	11.451	8.704–14.198
Provision of information content		
Gender knowledge	REF	
Knowledge of Maternity	2.602	-0.551–5.756

Daily health care	7.006	4.049–9.963
Knowledge of Diseases	7.931	4.975–10.886

^a: willingness to pay

^b: Chinese Yuan

Discussion

This study utilized a Discrete Choice Experiment (DCE) to explore women’s preferences for artificial intelligence (AI) chatbots. The results revealed significant variations in user preferences across different chatbot attributes, particularly for information usefulness, response accuracy, readability, and content provision, whereas cost and data collection were of relatively lower importance.

The scalability, accessibility, ease of use, and rapid information dissemination of AI chatbots offer supplementary benefits to public health efforts[46], addressing issues such as capacity constraints, social distancing requirements, and misinformation[47]. Research suggests that people sometimes prefer interacting with AI chatbots over doctors in certain contexts, as these chatbots can respond more quickly, provide high-quality feedback[48], and offer empathetic interactions. The application of AI in medicine is expanding across various domains, including medical image analysis[49], drug interaction detection[50], high-risk patient identification[51], and medical record coding[52]. In healthcare education, tools like ChatGPT have been used to facilitate personalized learning, encourage critical thinking, and support problem-based learning[53]. The widespread use of AI has made chatbots popular for accessing health information.

The study found that information utility is one of the most valued attributes of AI chatbots among participants. Participants appeared to prefer chatbots that provide practical and relevant information, aligning with the findings of Kim et al. (2024) in their systematic review and meta-analysis on women’s health[26], which suggested that AI chatbots may positively impact women’s health by reducing anxiety and depression, promoting healthy behaviors, and offering health education. Although the review highlights that the provision of practical and relevant information appears to be a key factor in improving health outcomes for women, the current study found that the P-value of the 'less practical' level did not reach significance, indicating that the practicality of information only has a significant impact on user choices when it is very high or very low. This suggests that users expect the information provided by chatbots to have practical application value. Therefore, when designing chatbots, developers should consider focusing on providing practical information that can meet users' actual needs and be immediately applied to their daily lives.

The results also indicated that response accuracy and readability significantly influence participants’ preferences. Participants seemed to favor AI chatbots with high accuracy and comprehensible content, which aligns with the systematic review by Aggarwal et al. (2023)[54]. Aggarwal et al. included 15 empirical studies on AI chatbots that facilitate health behavior changes, including interventions like healthy lifestyles, smoking cessation, and medication adherence. While some studies demonstrated the efficiency of AI chatbots, mixed results were observed regarding their feasibility, acceptability, and usability[55,56]. These findings suggest that improving chatbot accuracy and readability is crucial for effectively promoting behavior change. Accurate health information is likely key to building user trust[57], and this study further confirms that participants valued accurate responses. Readability also plays a significant role in users' choices, particularly among those with low health literacy or unfamiliarity with complex medical terms, as easily

understandable information enhances comprehension and adherence[58-60]. In this study, the P-values for the "difficult to understand" attribute were not statistically significant, suggesting that the readability of AI chatbot information might not substantially influence user preferences within certain thresholds. This may indicate that users prefer levels of readability that are "easier to understand" or "very easy to understand." When information reaches a level deemed difficult to comprehend, users might be disinclined to engage further with the AI chatbot. These findings suggest that AI chatbots should aim to optimize language and interface design based on accurate content to ensure comprehensibility and cater to diverse women's needs.

Regarding content preferences, participants showed a greater interest in disease knowledge, consistent with the systematic review by Alyasiri et al. (2024), which categorized AI applications across healthcare domains such as patient support and education[61]. AI chatbots appear to provide personalized health education and disease prevention information, offering content relevant to users' needs. The preference for pregnancy and childbirth knowledge did not seem to be significant, indicating that these types of information do not have a clear advantage in user selection preferences. This may be because women are more concerned with the prevention, diagnosis, and treatment of diseases themselves, and may not necessarily have a direct need for information related to pregnancy and childbirth as much as disease knowledge. However, it is important to note that the high preference for disease knowledge could also be influenced by the fact that participants were recruited from a gynecology unit of a hospital, where individuals are likely to have one or more health conditions and a stronger interest in learning about diseases. We cannot conclusively assume that disease knowledge would rate as highly in a non-hospital sample, and this should be taken into consideration when interpreting the findings. While emphasizing the importance of providing information on disease knowledge, it is crucial not to overlook the demand for content in other areas. Women's health is a multifaceted field that includes knowledge related to gender, pregnancy, childbirth, daily healthcare, and various diseases. Therefore, when developing chatbots, it is essential to consider all these aspects comprehensively to address the diverse needs of users.

On the topic of service cost, participants appeared to prefer free AI chatbots. This preference is consistent with other systematic reviews[62], which noted that free or low-cost services are more easily accepted by users, particularly when economic pressures are high. In a fast-paced era, the accessibility of free AI chatbots allows users to seek answers anytime and anywhere, increasing their reliance and usage. The preference for free services spans all age groups and socioeconomic backgrounds, suggesting that financial factors may strongly influence the choice for cost-free AI tools.

Finally, data collection did not significantly influence participants' preferences. However, Kim et al. (2024) emphasized the importance of privacy protection in AI applications for women's health[26]. While data collection can help improve service quality, ensuring data security remains crucial in healthcare settings. Users generally value the protection of personal health data, and limiting data collection to non-identifiable question-and-answer content can help alleviate concerns about data misuse or breaches.

This study provides valuable insights but has several limitations. Firstly, it employed convenience sampling and only included gynecological outpatients from a hospital in China. Therefore, the representativeness of the sample is limited. Since these female patients were surveyed in a medical setting, they are likely to be more interested in learning about diseases. As a result, the findings may not fully reflect the preferences and needs of a broader female population regarding women's health chatbots. Therefore, future research should test these chatbot attributes in different cultural and social contexts to assess whether these preferences are universal or context-dependent. Secondly, the

assumptions in the questionnaire might not fully reflect real-world preferences. Demographic factors such as age, education, and health literacy can influence how users engage with AI chatbots. Therefore, future studies should consider including a more diverse sample to capture a wider range of preferences. Finally, other variables, such as trust in technology and healthcare system perceptions, were not explored in this study but may impact user preferences. Expanding research to include these factors and diverse populations will help refine chatbot design to better meet the needs of different groups. Future studies should explore these aspects to provide a more comprehensive understanding of AI chatbot preferences across various contexts.

Conclusion

This discrete choice experiment highlights key preferences for AI chatbots in healthcare, indicating that users prioritize high response accuracy (100%), legibility, free services, and the practical utility of information, particularly in relation to disease knowledge. These findings suggest that the development of AI chatbots should emphasize clarity, accessibility at no cost, user privacy protection, and the relevance of health content. By incorporating these attributes, user engagement can be enhanced, the dissemination of health education can be improved, and the effectiveness of AI tools in promoting disease management can be maximized.

List of abbreviations

NLP: natural language processing; DCE: Discrete Choice Experiment; WTP: Willingness to Pay; CNY: Chinese Yuan

Ethics approval and consent to participate

The study has been approved by the Ethics Committee of the Baoding No. 4 Central Hospital (approval number 2022013), and all study subjects have signed informed consent forms to ensure that they clearly understand the purpose and process of the study and that participate voluntarily.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Key Research and Development Plan Project [grant numbers 2022YFC3600904].

Authors' contributions

YBW and HWM conceived and designed the study. JW, HWM, YJ and JBZ reviewed and revised the manuscript. YBW provided theoretical guidance and methodological expertise. HWM designed the questionnaire. TL and JHL are responsible for the study implementation. All authors provided revisions and approved the final manuscript.

Acknowledgments

We are grateful to the professionals and students who provided enormous support in the recruitment and data collection of this study, and the participants who took the time to cooperate with the survey.

References

- [1] The United Nations. Accessed November 28, 2023.
<https://population.un.org/wpp/Download/Standard/Population> (open in a new window).
- [2] The World Health Organization. "6 Priorities for Women and Health." Accessed November 28, 2023a.
<https://www.who.int/news-room/spotlight/6-priorities-for-women-and-health>.
- [3] Hibbard JH, Pope CR. Women's roles, interest in health and health behavior. *women Health*. 1987;12(2):67-84. doi: 10.1300/J013v12n02_05. PMID: 3424849.
- [4] Yang X, Yang F, Fu W, et al. The current status of community mental health services in three northern areas of China. *Family Medicine and Community Health* 2014;2:doi: 10.15212/FMCH.2014.0104
- [5] Lin F, Sun Q, Peng Z, et al. The innovations in China's primary health care reform: Development and characteristics of the community health services in Hangzhou. *Family Medicine and Community Health*. 2015;3:doi: 10.15212/FMCH.2015.0135
- [6] Carter M, Burley V, Nykjaer C, Cade J. Adherence to a Smartphone Application for Weight Loss Compared to Website and Paper Diary: Pilot Randomized Controlled Trial. *J Med Internet Res* 2013;15(4):e32. URL: <https://www.jmir.org/2013/4/e32>. DOI: 10.2196/jmir.2283.
- [7] Cheng Z, Guo S. Automatic Text Summarization for Public Health WeChat Official Accounts Platform Base on Improved TextRank. *J Environ Public Health*. 2022 Jul 13;2022:1166989. doi: 10.1155/2022/1166989. PMID: 35874881; PMCID: PMC9300291.
- [8] Chen X, Zhou X, Li H, Li J, Jiang H. The value of WeChat application in chronic diseases management in China. *Comput Methods Programs Biomed*. 2020 Nov;196. 105710. doi: 10.1016/j.cmpb.2020.105710. Epub 2020 Aug 14. PMID: 32858284.
- [9] Jiang Y, Liu F, Guo J, Sun P, Chen Z, Li J, Cai L, Zhao H, Gao P, Ding Z, Wu X. Evaluating an Intervention Program Using WeChat for Patients With Chronic Obstructive Pulmonary Disease: Randomized Controlled Trial. *J Med Internet Res*. 2020 Apr 21;22(4):e17089. doi: 10.2196/17089. PMID: 32314971; PMCID: PMC7201319.
- [10] Sun M, Yang L, Chen W, Luo H, Zheng K, Zhang Y, Lian T, Yang Y, Ni J. Current status of official WeChat accounts for public health education. *J Public Health (Oxf)*. 2021 Sep 22;43(3):618-624. doi: 10.1093/pubmed/fdz163. PMID: 31974552.
- [11] Safran Naimark J, Madar Z, Shahar DR. The impact of a Web-based app (eBalance) in promoting healthy lifestyles: randomized controlled trial. *J Med Internet Res*. 2015 Mar 2;17(3):e56. doi: 10.2196/jmir.3682. PMID: 25732936; PMCID: PMC4376125.
- [12] Lapp B, Zakus D. Factors associated with the utilization of institutional and home birth services among women in Ethiopia: A scoping review. *Family Medicine and Community Health* 2016;4:doi: 10.15212/FMCH.2016.0125

- [13] Kao CK, Liebovitz DM. Consumer Mobile Health Apps: Current State, Barriers, and Future Directions. *PM R*. 2017 May;9(5S):S106-S115. doi: 10.1016/j.pmrj.2017.02.018. PMID: 28527495.
- [14] Manickam P, Mariappan SA, Murugesan SM, Hansda S, Kaushik A, Shinde R, Thipperudraswamy SP. Artificial Intelligence (AI) and Internet of Medical Things (IoMT) Assisted Biomedical Systems for Intelligent Healthcare. *biosensors (Basel)*. 2022 Jul 25;12(8):562. doi: 10.3390/bios12080562. PMID: 35892459; PMCID: PMC9330886.
- [15] Hendrix N, Hauber B, Lee CI, Bansal A, Veenstra DL. Artificial intelligence in breast cancer screening: primary care provider preferences. *J Am Med Inform Assoc*. 2021 Jun 12;28(6):1117-1124. doi: 10.1093/jamia/ocaa292. PMID: 33367670; PMCID: PMC8200265.
- [16] Miller DD, Brown EW. Artificial Intelligence in Medical Practice: The Question to the Answer? *Am J Med*. 2018 Feb;131(2):129-133. doi: 10.1016/j.amjmed.2017.10.035. Epub 2017 Nov 7. PMID: 29126825.
- [17] Wilson L, Marasoiu M. The Development and Use of Chatbots in Public Health: a Scoping Review. *JMIR Hum Factors*. 2022 Oct 5;9(4):e35882. doi: 10.2196/35882 . pmid: 36197708; pmcid: pmc9536768.
- [18] Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, Chartash D. How Does ChatGPT Perform on the United States Medical Licensing Examination? the Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Med Educ*. 2023 Feb 8;9:e45312. doi: 10.2196/45312. PMID: 36753318; PMCID: PMC9947764.
- [19] Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, Ayoub W, Yang JD, Liran O, Spiegel B, Kuo A. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin Mol Hepatol*. 2023 Jul;29(3):721-732. doi: 10.3350/cmh.2023.0089. Epub 2023 Mar 22 . pmid: 36946005; pmcid: pmc10366809.
- [20] Liu J, Wang C, Liu S. Utility of ChatGPT in Clinical Practice. *J Med Internet Res*. 2023 Jun 28;25:e48568. doi: 10.2196/48568. PMID: 37379067; PMCID: PMC10365580.
- [21] Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, Normando E, Meinert E. The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review. *J Med Internet Res*. 2020 Oct 22;22(10):e20346. doi: 10.2196/20346. PMID: 33090118; PMCID: PMC7644372.
- [22] Chung K, Cho HY, Park JY. A Chatbot for Perinatal Women's and Partners' Obstetric and Mental Health Care: Development and Usability Evaluation Study. *JMIR Med Inform*. 2021 Mar 3;9(3):e18607. doi: 10.2196/18607. PMID: 33656442; PMCID: PMC7970298.
- [23] Tawfik E, Ghallab E, Moustafa A. A nurse versus a chatbot – the effect of an empowerment program on chemotherapy-related side effects and the self-care behaviors of women living with breast Cancer: a randomized controlled trial. *BMC Nurs*. 2023 Apr 6;22(1):102. doi: 10.1186/s12912-023-01243-7. PMID: 37024875; PMCID: PMC10077642.
- [24] Mane HY, Channell Doig A, Marin Gutierrez FX, Jasczynski M, Yue X, Srikanth NP, Mane S, Sun A, Moats RA, Patel P, He X, Boyd-Graber JL, Aparicio EM, Nguyen QC. Practical Guidance for the Development of Rosie, a Health Education Question-and-Answer Chatbot for New Mothers. *J Public Health Manag Pract*. 2023 Sep-Oct 01;29(5):663-670. doi: 10.1097/PHH.0000000000001781. PMID: 37478093; PMCID: PMC10372746.
- [25] Maeda E, Miyata A, Boivin J, Nomura K, Kumazawa Y, Shirasawa H, Saito H, Terada Y. Promoting fertility awareness and preconception health using a chatbot: a randomized controlled trial. *Reprod Biomed Online*. 2020 Dec;41(6):1133-1143. doi: 10.1016/j.rbmo.2020.09.006. Epub 2020 Sep 11. PMID: 33039321.

- [26] Kim HK. The Effects of Artificial Intelligence Chatbots on Women's Health: A Systematic Review and Meta-Analysis. *Healthcare (Basel)*. 2024 Feb 23;12(5):534. doi: 10.3390/healthcare12050534. PMID: 38470645; PMCID: PMC10930454.
- [27] Maenhout L, Peuters C, Cardon G, Compennolle S, Crombez G, DeSmet A. Participatory Development and Pilot Testing of an Adolescent Health Promotion Chatbot. *Front Public Health*. 2021 Nov 11;9:724779. doi: 10.3389/fpubh.2021.724779. PMID: 34858919; PMCID: PMC8632020.
- [28] Fan X, Chao D, Zhang Z, Wang D, Li X, Tian F. Utilization of Self-Diagnosis Health Chatbots in Real-World Settings: Case Study. *J Med Internet Res*. 2021 Jan 6;23(1):e19928. doi: 10.2196/19928. PMID: 33404508; PMCID: PMC7817366.
- [29] Chang IC, Shih YS, Kuo KM. Why would you use medical chatbots? interview and survey. *Int J Med Inform*. 2022 Sep;165:104827. doi: 10.1016/j.ijmedinf.2022.104827. Epub 2022 Jun 30. PMID: 35797921.
- [30] Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digit Health*. 2019 Aug 21;5:2055207619871808. doi: 10.1177/2055207619871808. PMID: 31467682; PMCID: PMC6704417.
- [31] Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review. *J Med Internet Res*. 2021 Jan 13;23(1):e17828. doi: 10.2196/17828. PMID: 33439133; PMCID: PMC7840290.
- [32] Xue J, Zhang B, Zhao Y, Zhang Q, Zheng C, Jiang J, Li H, Liu N, Li Z, Fu W, Peng Y, Logan J, Zhang J, Xiang X. Evaluation of the Current State of Chatbots for Digital Health: Scoping Review. *J Med Internet Res*. 2023 Dec 19;25:e47217. doi: 10.2196/47217. PMID: 38113097; PMCID: PMC10762606.
- [33] Mariamo A, Temcheff CE, Léger PM, Senecal S, Lau MA. Emotional Reactions and Likelihood of Response to Questions Designed for a Mental Health Chatbot Among Adolescents: Experimental Study. *JMIR Hum Factors*. 2021 Mar 18;8(1):e24343. doi: 10.2196/24343. PMID: 33734089; PMCID: PMC8080266.
- [34] Kim, Mirae & Oh, Jaedong & Kim, Doha & Shin, Jungwoo & Lee, Daeho. (2024). Understanding User Preferences in Developing a Mental Healthcare AI Chatbot: A Conjoint Analysis Approach. *International Journal of Human-Computer Interaction*. 1-9. 10.1080/10447318.2024.2353450.
- [35] Laba TL. Using Discrete Choice Experiment to elicit patient preferences for osteoporosis drug treatments: where to from here? *Arthritis Res Ther*. 2014. 16(2):106. doi: 10.1186/ar4501. Epub 2014 Mar 10. PMID: 25167089; PMCID: PMC4060568.
- [36] von Wedel P, Hagist C. Physicians' preferences and willingness to pay for artificial intelligence-based assistance tools: a discrete choice experiment among german radiologists. *BMC Health Serv Res*. 2022 Mar 26;22(1):398. doi: 10.1186/s12913-022-07769-x. PMID: 35346183; PMCID: PMC8959781.
- [37] Noben CYG, Stammen LA, Vaassen S, Haeren R, Stassen L, Mook WV, Essers B. Discrete choice experiment on educating value-based healthcare. *postgrad Med J*. 2021 Aug;97(1150):515-520. doi: 10.1136/postgradmedj-2019-137190. epub 2020 Aug 12. PMID: 32796111.
- [38] Ryan M, Bate A, Eastmond CJ, Ludbrook A. Use of discrete choice experiments to elicit preferences. *Qual Health Care*. 2001 Sep;10 Suppl 1(Suppl 1):i55-60. doi: 10.1136/qhc.0100055... PMID: 11533440; PMCID: PMC1765744.
- [39] Louviere JJ, Lancsar E. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. *Health Econ Policy Law*. 2009 Oct;4(Pt 4). 527-46. doi: 10.1017/S1744133109990193. PMID: 19715635.

- [40] von Wedel P, Hagist C. Physicians' preferences and willingness to pay for artificial intelligence-based assistance tools: a discrete choice experiment among german radiologists. *BMC Health Serv Res*. 2022 Mar 26;22(1):398. doi: 10.1186/s12913-022-07769-x. PMID: 35346183; PMCID: PMC8959781.
- [41] Battineni G, Chintalapudi N, Amenta F. AI Chatbot Design during an Epidemic Like the Novel Coronavirus. *Healthcare (Basel)*. 2020 Jun 3;8(2):154. doi: 10.3390/healthcare8020154. PMID: 32503298; PMCID: PMC7349073.
- [42] Qian Chen, Yeming Gong, Yaobin Lu, Jing Tang, Classifying and measuring the service quality of AI chatbot in frontline service, *Journal of Business Research*, Volume 145, 2022, Pages 552-568, ISSN 0148-2963, <https://doi.org/10.1016/j.jbusres.2022.02.088>.
- [43] Chen, Ja-Shen & Le, Tran-Thien-Y & Florence, Devina. (2021). Usability and responsiveness of artificial intelligence chatbot on online customer experience in e-retailing. *International Journal of Retail & Distribution Management*. 49. 1512-1531. 10.1108/IJRDM-08-2020-0312.
- [44] Hua Jiang, Yang Cheng, Jeongwon Yang, Shanbing Gao, AI-powered chatbot communication with customers: Dialogic interactions, satisfaction, engagement, and customer behavior, *Computers in Human Behavior*, Volume 134, 2022, 107329, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2022.107329>.
- [45] Ashfaq, Muhammad & Yun, Jiang & Yu, Shubin & Loureiro, Sandra. (2020). I, Chatbot: Modeling the Determinants of Users' Satisfaction and Continuance Intention of AI-Powered Service Agents. *Telematics and Informatics*. 54. 101473. 10.1016/j.tele.2020.101473.
- [46] Pearce C. Chronic disease in a digital health environment. *Family Medicine and Community Health* 2018;6:doi: 10.15212/FMCH.2017.0144
- [47] Amiri P, Karahanna E. Chatbot use cases in the Covid-19 public health response. *J Am Med Inform Assoc*. 2022 Apr 13;29(5):1000-1010. doi: 10.1093/jamia/ocac014. PMID: 35137107; PMCID: PMC8903403.
- [48] Ayers JW, Poliak A, Dredze M, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med*. 2023;183(6):589-596. doi:10.1001/jamainternmed.2023.1838
- [49] Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digit Med*. 2023 Jul 6;6(1):120. doi: 10.1038/s41746-023-00873-0. PMID: 37414860; PMCID: PMC10326069.
- [50] Dunn P, Cianflone D. Artificial intelligence in cardiology: Exciting but handle with caution. *Int J Cardiol*. 2023 Oct 1;388:131117. doi: 10.1016/j.ijcard.2023.06.008. Epub 2023 Jun 9. PMID: 37302418.
- [51] Zhong NN, Wang HQ, Huang XY, Li ZZ, Cao LM, Huo FY, Liu B, Bu LL. Enhancing head and neck tumor management with artificial intelligence: integration and perspectives. *Semin Cancer Biol*. 2023 Oct;95:52-74. doi: 10.1016/j.semcancer.2023.07.002. Epub 2023 Jul 18. PMID: 37473825.
- [52] Li W, Zhang Y, Chen F. ChatGPT in Colorectal Surgery: A Promising Tool or a Passing Fad? *Ann Biomed Eng*. 2023 Sep;51(9):1892-1897. doi: 10.1007/s10439-023-03232-y. Epub 2023 May 10. PMID: 37162695.
- [53] Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare (Basel)*. 2023 Mar 19;11(6):887. doi: 10.3390/healthcare11060887. PMID: 36981544; PMCID: PMC10048148.
- [54] Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial Intelligence-Based Chatbots for Promoting Health Behavioral Changes: Systematic Review. *J Med Internet Res*. 2023 Feb 24;25:e40789. doi: 10.2196/40789. PMID: 36826990; PMCID: PMC10007007.

- [55] Caranfa JT, Bommakanti NK, Young BK, Zhao PY. Accuracy of Vitreoretinal Disease Information From an Artificial Intelligence Chatbot. *JAMA Ophthalmol*. 2023 Aug 3:e233314. doi: 10.1001/jamaophthalmol.2023.3314. Epub ahead of print. PMID: 37535363; PMCID: PMC10401388.
- [56] Diamandis EP. Chatbot GPT can be grossly inaccurate. *Clin Chem Lab Med*. 2023 Jul 31. doi: 10.1515/cclm-2023-0765. Epub ahead of print. PMID: 37501646.
- [57] Hirose T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic Accuracy of Differential-Diagnosis Lists Generated by Generative Pretrained Transformer 3 Chatbot for Clinical Vignettes with Common Chief Complaints: a pilot study. *Int J Environ Res Public Health*. 2023 Feb 15;20(4). 3378. doi: 10.3390/ijerph20043378. PMID: 36834073; PMCID: PMC9967747.
- [58] Moons P, Van Bulck L. Using ChatGPT and Google Bard to improve the readability of written patient information: a proof-of-concept. *Eur J Cardiovasc Nurs*. 2023 Aug 21:zvad087. doi: 10.1093/eurjcn/zvad087. Epub ahead of print. PMID: 37603843.
- [59] Yurdakul OV, Kilicoglu MS, Bagcier F. Evaluating the reliability and readability of online information on osteoporosis. *Arch Endocrinol Metab*. 2021 Nov 1;65(1):85-92. doi: 10.20945/2359-3997000000311. Epub 2020 Nov 9. PMID: 33166440; PMCID: PMC10528702.
- [60] Denecke K, Borycki EM, Kushniruk AW. What Can We Learn from Quality Requirements in ISO/TS 82304-2 for Evaluating Conversational Agents in Healthcare? *Stud Health Technol Inform*. 2022 Nov 3;299:245-250. doi: 10.3233/SHTI220992. PMID: 36325870.
- [61] Younis HA, Eisa TAE, Nasser M, Sahib TM, Noor AA, Alyasiri OM, Salisu S, Hayder IM, Younis HA. A Systematic Review and Meta-Analysis of Artificial Intelligence Tools in Medicine and Healthcare: Applications, Considerations, Limitations, Motivation and Challenges. *Diagnostics (Basel)*. 2024 Jan 4;14(1):109. doi: 10.3390/diagnostics14010109. PMID: 38201418; PMCID: PMC10802884.
- [62] Grassini, E., Buzzi, M., Leporini, B. et al. A systematic review of chatbots in inclusive healthcare: insights from the last 5 years. *Univ Access Inf Soc* (2024). <https://doi.org/10.1007/s10209-024-01118-x>