Requirement analysis for social acceptance of AI medical interview - from view of quality in use

Natsuko Noda
Shibaura Institute of Technology
Tokyo, Japan
nnoda@shibaura-it.ac.jp

Shin'ichi Fukuzumi
RIKEN
Tokyo, Japan
shin-ichi.fukuzumi@riken.jp

Abstract— **The purpose of this study is to clarify the requirements that must be met for AI-based medical interview to be accepted by humans and society. In this study, we set up personas of users who have various ideas, and create usage cases of AI medical interview service based on the personas. From the created usage cases, we extract and analyze the feelings and thoughts of the personas in response to each question and usage flow of the AI medical interview service. From this analysis, we clarify requirements on AI medical interview services referring to the concept of quality in use in ISO/IEC 25010:2011. Using this standard, we were able to extract the requirements of "usability," "reliability," and "acceptability" that have been not enough considered in the past. In this paper, we introduce the requirements acquired from the analysis.**

Keywords— quality in use, AI systems, persona, requirement analysis

I. Introduction

Currently, AI technology is expected to be used in the medical field in a wide variety of ways, such as medical image diagnosis by AI-based image processing and medical record analysis by natural language processing. As one of the medical applications of AI, AI-based medical interview services are being implemented to reduce opportunities for human-to-human contact in order to prevent infection and to improve operational efficiency in hospitals.

However, it is difficult to draw a line between the responsibilities of the doctors and the AI-based services. In the case of the first visit to the hospital, the relationship between the user and the AI-system is simple because there is no judgment by the doctor. In the case of the second visit, the diagnosis and treatment by the doctors that the user has visited before will occur, and that makes the relationship between the user and the AI complex and blurs the lines of responsibility between AI and doctors. Therefore, it is necessary to make requirements for the acceptance of such AI diagnosis services in society.

To make such requirements, it is necessary to consider not only the analysis from the physician's point of view, but also the system that has contact with the patient and others [1]. And as pointed out in [2], it is necessary to analyze not only from the viewpoint of the service developer but also from the viewpoint of the users.

The purpose of this study is to clarify the requirements that must be met for AI-based medical interview to be accepted by humans and society.

We set up personas of users with a variety of possible ideas, and create usage cases for an AI medical interview service based on the personas. In this study, patients are limited to first-time patients; that means these personas have the first contact with the AI medical interview service. Then, from the created usage cases, we extract and analyze the feelings and thoughts of the personas in response to each question and the flow of use of the AI medical interview service. Finally, we extract the requirements on AI medical interview services referring to the concept of quality in use in ISO/IEC 25010:2011.

II. Extraction of feelings and thoughts for AI medical interview

In this section, we first assume a free AI medical interview service [3] and create user personas for it. We create not only just one persona but several personas. Although we survey just one specific AI medical interview service in this study, our goal is to obtain requirements for general AI-based medical services. Those services should be used by various users, that cannot be represent by one persona. Thus, we create several personas that could cover broader range of those users. We assume that each persona has a symptom, comes to a hospital outpatient clinic for the first time, uses the AI medical interview service in the hospital, and receives information on possible symptoms and recommended departments. Next, we analyze how the user feels and thinks about the AI interview service based on the assumed usage case. Table 1 shows some examples of questions asked by the AI medical interview service and the users' feelings and thoughts toward it.

Table 1 Examples of questions of AI medical interview and extracted feelings and thoughts

Question by the service	Feelings and/or thoughts	Persona
Please indicate your age and gender.	Since they use AI, can't they just use my patient registration card and skip this question?	No.1
Please indicate your age and gender.	If I were genderless, what could I answer?	(many)
Please provide your occupation and address.	I'm not comfortable typing the answers to such private questions.	No.2
Please provide your occupation and address.	I can answer this kind of private question without any hesitation with an AI.	(many)
Are you mentally depressed?	It's uncomfortable to be asked such question when I'm depressed because of my injury at a critical time.	No.3
Are you mentally depressed?	If I talk to the AI about mental issues, can the AI understand it?	No.3

Have you stopped going out on a daily basis?	I'm uncomfortable being asked such a misguided question when I'm bringing my child here because he has a fever.	No.6
Do you have any fear of weight gain or dietary restrictions?	This is a question that has not been applied to the patients I support. Difficult to support.	No.7
Do you have any dietary restrictions?	I'm here because I have no appetite, and this question is offensive.	No.8

In this study, we call this set of question and corresponding feeling/thought pairs as usage data. Note that there can be more than one feeling/thought for a single question by AI, so the set of the pairs of question and feeling/thought in the usage data contains more than one pair of question and feeling/thought for the same question. Each pair of question and feeling/thought, i.e. each usage data is numbered for identification. Note that this number is just for identification and the order of the numbers has no meaning.

III. CLASSIFICATION OF FEELINGS, THOUGHTS AND IMPORTANCE

As a preliminary step for clarifying the requirements for AI medical interview services, we classify the extracted usage data. To extract the requirements, we classify the usage data with high importance in the medical diagnosis and those with low importance.

A two-axis matrix diagram is used to classify the data from two perspectives: the perspective of "extracted feelings and thoughts" and the perspective of "questions" from which the feelings and thoughts are extracted.

- Horizontal axis: Extracted feelings and thoughts. In the horizontal axis, the criterion is whether or not the extracted feelings and thoughts are unique to systems and services using AI. For example, feelings and thoughts such as "It's easy to answer private questions with the AI." can be regarded as unique to AI-based services. On the other hand, feelings and thoughts such as "If I were genderless, what could I answer?" may be general thought to all medical interviews. This categorization is used to classify important usage data regarding feelings and thoughts that should be considered in the context of AI medical interview service.

- Vertical axis: Questions of the extraction source. On the vertical axis, the criterion is whether the question from which feelings and thoughts are extracted is easy to answer in the medical interview. For example, questions that deal with information on demographic attributes may be easy to answer. On the other hand, a question such as "have you been in a dense space or a poorly ventilated room recently?" is slightly less easy to answer, because it deals with information on the patient's factual behavior, which may depend on the user's memory and feelings.
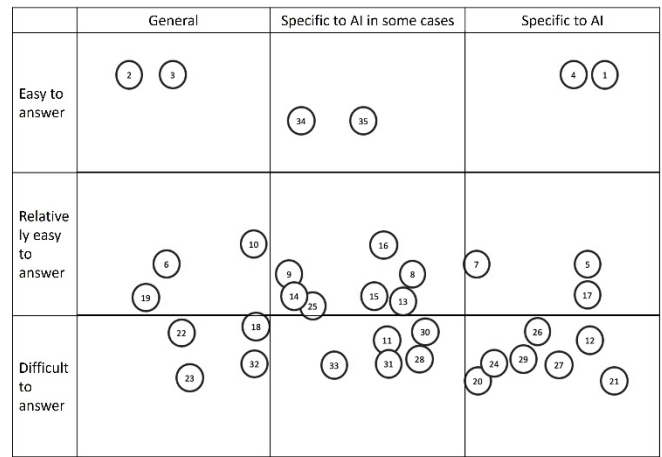

Fig. 1 Classification table of usage data

Based on the above classification method, a classification table of usage data is created, as shown in Figure 1. Each circle shows a usage data and the number in it means the identification number of the usage data.

The classification is based on a two-axis matrix analysis from two perspectives: the perspective of "strength of relevance of AI in extracted feeling and thought" of the usage data, and the perspective of "easiness of answering question." On the horizontal axis, the degree of specificity of "feelings and thoughts" as AI interview is classified. and on the vertical axis, the degree of ease of answering "questions" is classified.

According to this classification, the lower right the data is in the table, the more important it is considered to be for requirement acquisition for AI medical interview services from the following reasons. First, the data is in the right side is strongly related to the AI services. Because we try to extract requirements for AI medical interview services, not for medical interview services generally, the right side data is more important. Secondly, to be accepted widely in society, services should have high quality in use. However, difficult questions to answer that the services ask contribute to making the services harder to use. Therefore, the lower down the table the data is, the more carefully it has to be analyzed.

IV. REQUIREMENTS ACQUISITION

In this study, we refer to the quality in use incorporated in ISO/IEC 25010:2011 [4] in order to clarify the requirements for AI medical interviews from the perspective of use by patients, attendants, etc.

Based on the classification table shown in Figure 1, we find requirements for AI medical interview services. First, we examine the "feelings and thoughts" of the usage data and extract the issues for AI medical interviews. The issues are compared with the elements of the quality model of quality in use, and the quality elements related to the issues are identified. Then, the requirements to be satisfied by AI medical interviews are extracted for these quality elements. Figure 2 shows the extracted requirements, categorized according to the importance of the usage data.

Table 2 Extracted requirements

	From high importance usage data	From medium importance usage data	From low importance usage data
Usability requirements	- A number of metaphors and onomatopoeias should be available to convey the pain. - Use UI and background music that does not cause anxiety. - Use an objective method of answering the question. - Similar or peripheral questions should be grouped together. - Consider the surrounding environment and make the UI easy to handle private information. - Make the UI concise and easy to use.	- Provide a good tutorial or guide.	- The total time from start to result output should be shortened. - Provide a rationale for the results.
Reliability requirements	- The output results should be comparable to the doctor's diagnosis results. - To cover all diseases. - Indicate the need for questions. - Include a phase to ask for medical history and medication history. - Provide strict protection of information. - As an infection control measure, use voice UI and other methods to reduce the number of contacts.	- Assume errors in patient judgment and input of responses. - Identify the patient's unrecognized symptoms and cover the symptoms. - Make a comprehensive assessment of Covid-19 based on occupation, commuting method, and whether or not he/she participates in events.	- The response method should be such that it results in accurate input of information.
Acceptability requirements	- Tell the user that the purpose is screening. - It should be like spoken, not mechanical. - Observing the patient's state of tension in real time (for example, by measuring the heart rate), change the system's behavior. - Understand the characteristics of the patient and the background of the visit.	- The presentation of questions that assume input support by a caregiver should also be prepared. - To be linked with other systems for the prevention of the spread of Covid-19 infection (e.g. cocoa). - Clearly inform the user of the intended use of the input information.	- Consideration should be given to various ideas of gender to the extent that it is not medically problematic.

As a result, we were able to capture the impact of "use" (quality in use) from the patient's perspective and the medical side's perspective, and incorporate them as requirements. In addition, since the system requirements for improving the quality were presented in relation to "use," they were not simply functional requirements, but system requirements from the user's perspective. Although it is difficult to express these things as product or system specifications, they can be effective requirements for acceptance in society.

V. Conclusion

The demand for the use of AI in the medical field is rapidly increasing. Against this backdrop, we examined the requirements for AI to be accepted in society, with AI medical interviews as the first target.

Referring ISO/IEC 25010:2011, we were able to capture the impact of "use" (quality in use) from both the patient's and the medical side's perspectives and incorporate it into the requirements. In the future, we will consider more diverse usage scenarios to make the requirements more practical.

REFERENCES

[1] Fukuzumi, S. , et al. Extraction of new guideline items for ELSI-like perspectives in AI-based services, Proc. 32nd National Conference of the Japanese Society for Artificial Intelligence (2018). (in Japanese)

[2] Ema, Yusa; Nagakura, Katsue; Fujita, Takusen. Trust in Medical AI from Physician Survey, Proc.34th National Conference of the Japanese Society for Artificial Intelligence (2020). (in Japanese)

[3] Ubie AI consultation Ubie Corporation, https://ubie.app (referred Oct. 30, 2021)

[4] Fukuzumi, S., Wada, N. and Hirasawa, N.: "Quality in Use -Issues and proposal-", proceedings of IWESQ2020 (2020)