Adjusting Clarifying Questions to User's Profile

LIRAN NETA, University of Twente, The Netherlands

Literature reports different needs of users with high domain knowledge (experts) or little domain knowledge (novices). User-centered systems accommodate these differences. User-centered product search, particularly clarifying questions (CQs), can improve user performance and satisfaction in conversational product-search systems. Therefore, this research aims to enhance a conversational product search system with user-centered CQs. The research looked to figure out if providing CQs that match the user's level of knowledge results in a better search experience for the user. Our findings show that users' perceptions of the system's effectiveness can be slightly influenced by matching the type of CQs to their level of domain competence. For this research, we decided upon two types of CQs - Advanced and Basic. Advanced CQs include specific measurements, such as "12 to 18 inches", while Basic CQs include vague sizes, such as "small, medium, big". Overall, experts preferred the Advanced chatbot, while no significant distinction was found between the Basic and Advanced chatbots among novices. However, further elaboration on unfamiliar terms can be helpful for novices. Recommendation, given by the chatbot, on various capacities (such as RAM size) based on the user's purpose of usage seem to be valuable as well for the users.

Additional Key Words and Phrases: Information-seeking systems, Clarifying Questions, User-Centered search, User Profiling, Usability analysis

1 INTRODUCTION

Nowadays, searching for products online is a common activity. Online store customers often utilize search boxes or facets to find products with desired features. However, when users have specific requirements, they may need to look through multiple pages or refine their search queries multiple times to find what they want. Conversational product search is a method of searching, where users interact with technology in a natural conversation and receive responses [11]. This emerging field of research and usage is becoming increasingly popular because it allows users to search with complete sentences rather than keywords, making the process more tailored, convenient, and optimized [2, 7]. Technologies such as voice, natural language processing, and machine learning are bringing a more human-like interaction to consumer search experiences [11].

One of the key aspects of conversational search involves a mixed-initiative approach where users and the system take turns in leading the conversation. Users initiate the search by describing their information need and can request additional information at any point. The system can also take initiative by asking clarifying questions to better understand the user's intent [11]. Clarifying questions have become increasingly important in modern information-seeking systems, as they improve the search experience and usability by clarifying the user's needs [14] in a user-centered approach [8, 9]. Both users and the system benefit from these questions, as they result in improved retrieval performance and increased user satisfaction [11].

Although there has been progress in clarifying the users intent, studies suggest that asking too many CQs without considering the user's profile is detrimental to the product search experience [13]. To achieve this, selecting appropriate responses based on the conversation's context is crucial [1]. Research indicates that profiling users as novices or experts in a particular domain can help tailor the selection of CQs to their needs [5, 10]. Another study [12] supported the concept of utilizing non-technical or vague terminology when interacting with novice users. However, there is no research that studies the impact of adapting CQs to user profiles.

To fill this gap, in this research, we propose to improve user-centered conversational product search by providing CQs that match the user's profile (novices or experts). It is necessary to examine how targeted CQs influence their conversational search experience: *How does adjusting the type of Clarifying Questions (non-technical, technical) based on the user's profile (novices, experts) affect how the user perceives the system's efficiency?*

Basic, non-technical CQs were asked as a vague questions, such as "how big do you want your laptop screen to be - small, medium or large?". The technical equivalent was "how many inches do you want your laptop screen to be (between 12-18 inches)?". We considered four cases:

(1) basic (non-technical) CQs for novice users
(2) advanced (technical) CQs for novice users
(3) basic CQs for expert users
(4) advanced CQs for expert users

For each use case, we evaluated how the users perceived the system's efficiency. We defined efficiency as how easy to understand and answerable are the CQs that the user is being asked. The efficiency was evaluated qualitatively with in-depth online surveys.

By analyzing the results using inductive coding, we found that adjusting the type of CQs (Basic or Advanced) based on the user's profile (novice, expert) has a greater impact on how expert users perceive the system's efficiency. The results show that experts prefer the Advanced rather than the Basic chatbot, whereas no apparent distinction exists between the two chatbots among novices. We also found that novices want the chatbot to explain unfamiliar terms. The participants from both groups also suggested that the chatbot could recommend the required capacities of the laptop (such as RAM size) based on what tasks they need the laptop for.

The research contributes to the development of user-centered search systems that adapt the conversation to meet the needs and preferences of the user. Additionally, the research aims to improve the effectiveness of chatbots by understanding what the user wants it to do. For example, the chatbot can recommend capacities based on the user's initial description of how they plan to use their laptop and can explain terms that are unfamiliar to them.

TScIT 37, July 8, 2022, Enschede, The Netherlands

2 RELATED WORK

Clarifying questions is becoming more common in modern information seeking systems, with the goal of focusing on user needs [14]. In general, studies in the domain of user-centered product search have shown that better results are obtained when the search process is focused on the user needs [8, 9]. A particular way of adjusting the product search process to the user needs is done by understanding their degree of domain knowledgeability [5, 10]. According to Rowley [10], questions can be gathered into six broad areas, and one of them is the user profile. More particularly, Rowley states that "users range from being subject domain novices and computer novices all the way to subject experts and computer experts. The degree of knowledgeability of the [. . .] user and the domain experience should be reflected in the design of the user interface prompts, alerts and help facilities" [10]. In other words, the way the system works should be aligned with the user's domain knowledge. There is another study that suggests this categorizing of users into these two groups of novices and experts [6]. There is already evidence that user profiling should be considered when designing a conversational search system. For example, a study [12] indicated that missing domain knowledge may be a reason people use vague, non-technical terms. In other words, novices are more likely to use non-technical terms (e.g., big screen, fast computer) in their conversational search than specific and technical ones (16 inch, 32GB RAM). Having said that, there is still no clear indication that user profiling based on their level of domain knowledge (experts and novices) can increase the usability of a conversational search system.

System usability means how effectively, efficiently, and satisfactorily users may use a system to achieve their goals in a specific use context [4]. Brooke [4] describes effectiveness as the ability of users to complete tasks using the system and the quality of output of those tasks, while efficiency is focused on the resources required to perform the tasks. Satisfaction is the general reaction of the user who is using the system. In the context of CQ, we define efficiency as how easy to understand and answerable are the CQs that the user is being asked. If the CQs are easy to understand and to answer, in the way they are phrased, then it reduces the mental load of the user and increases the speed of performing the task. We define effectiveness as the relevance of the CQs – is the system asking the user for the right information? Otherwise, the system will not provide the product that is matching the user's needs. Such a case makes the system less effective. Last, we define satisfaction as the reaction of the user to the experience of using the system.

Overall, the literature suggests that asking clarifying questions by following the user-centered approach can lead to better use of the system and user satisfaction. However, there is still room for exploring how user profiling can be incorporated into the search process to improve the system further, make it more usable.

3 METHODS OF RESEARCH

In this research, we designed and tested CQs for two user profiles: novices (low domain knowledge) and experts (high domain knowledge). Therefore, we built two versions of a product search chatbot that asked two types of CQs: (1) used non-technical terms to ask about product requirements and (2) used precise, technical terms to ask for more specific product requirements (see Figure 1a and Figure 1b respectively).

We also took into account the cases of "misprofiling" (when novices are being asked advanced CQs, and experts receives the basic CQs). It was important to analyze the difference between the results under this condition, compared to the results under the "correct profiling" condition, where the CQs are adjusted to the users level of expertise. This analysis helped us explain whether the adjustment of CQs has a significant impact on the conversation or not.

This led to a 2x2 design with the following four conditions:

(1) basic CQs are given to novice users
(2) advanced CQs are given to novice users
(3) basic CQs are given to expert users
(4) advanced CQs are given to expert users

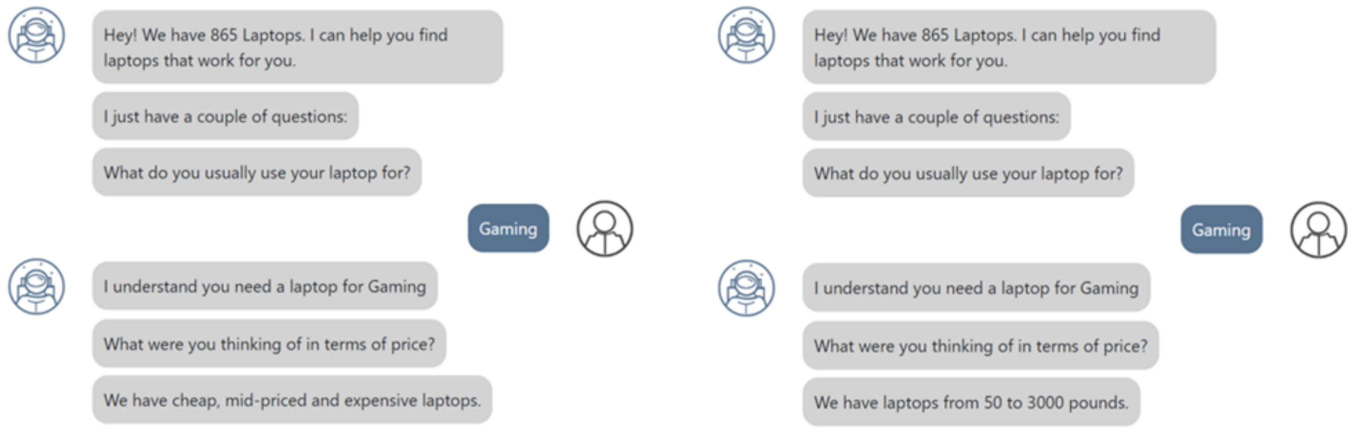
The method of research that was used in this study involved testing the perceived efficiency of chatbot websites with basic and advanced clarifying questions (CQs), using a sample of 24 participants (12 novices and 12 experts), with each condition being tested by six participants. Furthermore, this study has received clearance by the institute's ethics committee prior to conducting the study.

3.1 Participants

The study involved two distinct groups: novices and experts in the field of laptops. The participants did not receive any compensation for their participation. To identify participants for each group, we employed a self-assessment approach. At the beginning of the survey, participants were asked to rate their level of knowledge in laptops on a scale of 1 to 7. Our objective was to recruit individuals who either rated themselves as 1 or 2, indicating low expertise, or as 6 or 7, indicating high expertise. This approach allowed us to focus on individuals with clear distinctions in their perceived knowledge. Respondents who rated themselves as 4, which falls in the middle, were excluded due to the ambiguity of their expertise level. We decided to include some participants who rated themselves as 3 or 5. A subset of respondents who rated themselves as 5 had backgrounds in Computer Science or related fields, which leads to the assumption that this scale question suffered from the moderate responding bias. More statistics about the distribution of participants over the various cases could be found in Table 1 In addition, 10 of the 12 experts were males, while 8 of the 12 novices were females.

Table 1. Statistics of the distribution of the participants over the different conditions of the experiment (the type of chatbot in brackets)

	Novices (Basic)	Novices (Adv.)	Experts (Basic)	Experts (Adv.)
Females (No.)	5	0	3	2
Males (No.)	1	6	3	4
Age M (SD)	35.7 (11.6)	22.8 (1.7)	23.7 (1.9)	26.8 (4.7)
Domain knowledge M (SD)	2.0 (0.8)	6.2 (0.7)	2.2 (0.7)	5.0 (0.0)



(a) Basic Chatbot (b) Advanced Chatbot

Fig. 1. An example of the Basic (a) and the Advanced (b) chatbots.

3.2 Procedure

The participants received the URL address of the chatbot website based on their condition group. The aim of the study was to measure the perceived efficiency of chatbots when using non-technical (vague) terms versus technical terms. To test the perceived efficiency of the chatbot websites, the participants were given the task of searching for a laptop that fit their needs by interacting with one of the chatbots and answering the CQs. The CQs in both chatbot types (basic CQs and advanced CQs) were of the same order and asked about the same product aspects, but were presented in a different style, either using vague and non-technical terms such as 'small' and 'fast' or using specific technical terms such as 'size in inches' and 'RAM-size in GB'. The full list of CQs asked by the chatbots is shown in appendix A. After the conversational search interaction, the participants filled in a survey. The survey consisted of open questions that measured how the user perceived the system in terms of its efficiency. Participants were asked to provide feedback on several aspects of the system:

- Understanding the chatbot questions: Participants' feedback on how well they understood the questions asked by the chatbot.
- Relevance of terms used by the chatbot: Participants' opinions on whether the terms used by the chatbot were appropriate and made sense.
- Familiarity with terms used by the chatbot: Participants' level of familiarity with the specific words or phrases used by the chatbot.
- General impression of the chatbot: Participants' overall opinions and feelings about the chatbot's performance and interaction.
- Effect of the chatbot's knowledge level on the conversation: Participants' observations and comments on how the chatbot's level of knowledge affected the conversation.

3.3 Analysis

These answers were later analyzed using the method of inductive coding, which involved identifying themes or patterns that emerged from the qualitative data itself.

Once the data was collected through the survey, inductive coding was employed to gain an understanding of how users perceived the CQs. First, the data was read to gain a general understanding of the content. Next, codes or labels were identified that captured the key ideas or concepts present in the data, for each of the aspects of the system the participants were asked to provide feedback on (i.e., "relevance of terms", "familiarity with terms", etc.). These codes were descriptive and based on the actual content of the data. Based on these codes, patterns in the answers were found. Finally, these patterns were used to draw conclusions and make interpretations about the data.

To answer the main research question, several comparisons were made in the final part of the study. These comparisons were made based on the patterns. The first comparison was within the groups of users (novices and experts) themselves, to understand the impact of adjusting the CQs to the users (against the case of mis-profiling of a user). Next, each group was compared to determine for which group the clarifying question adjustment was more efficient from the users' perspective. Based on these results, conclusions or recommendations has been made.

4 RESULTS ANALYSIS

Overall, both chatbots (Basic and Advanced) were easy to comprehend by both groups (novices and experts), as shown in Table 2. Table 2 shows minor differences in how novices and experts perceived the understandability of the chatbots, which needs to be investigated with a larger sample size and statistical tests in future work. Participants indicated that they encountered difficulties in understanding at least one question. The challenges primarily arose from encountering technical terminology that was unfamiliar to

Table 2. The average rating (and standard deviation) given by users regarding their ability to understand the questions posed by the chatbot (on a scale from 1-7)

	Basic Chatbot M (SD)	**Advanced Chatbot** M (SD)
Novice	6.3 (0.7)	6.2 (1.0)
Expert	6.3 (1.1)	6.5 (0.8)

them and lacking knowledge about specific aspects of the product (more common for novices), as well as ambiguous language expressed by the chatbot. Figure 2 and Figure 3 visualize the count of questions that were difficult and easy to answer for novices and experts. A summary of the main points mentioned by the novices and experts is shown in Table 3.

In the following sections, we explore in more details the effects of the chatbot's content (what was asked) and the language (how was the question asked), to better understand how the users perceived the chatbot's efficiency. We start with the Basic chatbot and continue with the Advanced one.

4.1 Basic chatbot

4.1.1 Language. The Basic chatbot was easy to conversate for both the experts and the novices, who felt as if it was a real person talking to them. The language was clear and concise for them, and the questions were asked in a simple manner. The novices did not mention any problem with the way the questions were asked.

Half of the experts who used the basic chatbot explicitly mentioned that the vague options provided by the chatbot (e.g., "small, large") were helpful for some of the questions, since it gave them space for interpretation and mimic a real-life scenario of choosing a laptop, when usually we do not know the exact size. However, it was not the case for all the expert respondents. Some of the experts (2 out of 6) and one novice found the same questions difficult to answer because of lack of exact values, which the Basic chatbot was missing by its use of less technical and therefore more ambiguous language.

4.1.2 Content. Some novices felt that the chatbot asked them exactly the questions that should be asked when purchasing a laptop. On the other hand, 5 out of 6 novices mentioned that they do not know what RAM is, therefore they had a hard time understanding the question related to the RAM size.

The experts did not face the same issues, except for one participant who felt that the questions about the RAM and hard-drive storage were redundant, since they are dependent on the way the user is using the laptop. Both the experts and novices mentioned that questions about hardware specifics (processor, GPU maker, etc.) and general laptop layout (keyboard, screen attributes, etc.) were missing.

4.2 Advanced chatbot

4.2.1 Language. Similarly to the experience of the Basic chatbot's users, both the experts and novices found the language used by the Advanced chatbot easy, straightforward, clear, and concise. Some participants in each group noted that the chatbot's precise measurements (such as "laptops with storage between 64 GB and 2000 GB") provided as specifications during the question prompts contributed to the ease of use.

4.2.2 Content. However, they also had a hard time understanding the question talking about the RAM, because they were not sure what the needed capacity is to meet their needs. Some of them claimed the same thing regarding the hard drive storage size. It might be a result of lack of knowledge and unfamiliarity with the terms that were questioned about.

The experts experienced no difficulty at all and felt that the questions were all relevant and on point. One of them even spotted the problem for the novices, caused by the difficult term, and suggested a way to resolve it by having a short explanation of these terms and having references (e.g., 8GB of RAM "is ideal to run word and browsing [the internet]", whereas 32GB "is ideal for running multiple heavier programs, running heavy software and gaming"). In addition, some respondents from both groups claimed that the chatbot should be capable of asking follow-up questions in response to the purpose given by the respondent. As an example: the chatbot should induce the RAM and Hard-Drive storage needed for the laptop based on the "purpose" initially stated by the user. Like the Basic chatbot, both the novices and expert respondents claimed that the chatbot should ask about more hardware specifics and other laptop's attributes, such as processor, operating system etc.

4.3 The Effect of the chatbot's level of expertise

In the survey, participants were asked about their opinion regarding the influence of the chatbot's perceived knowledge level on the conversation. The responses varied, encompassing both positive and negative perspectives on how the chatbot's knowledge level affected the conversation. Both the novice and expert users rated both chatbots' level of knowledge around 5 (on a scale of 1 to 7). On average, novices rated the chatbots as much more knowledgeable than themselves, while experts rated them equally knowledgeable, as shown in Table 4.

In the following sections, we explore in more details the effects of the chatbot's assumed expertise on the users. We start with the Basic chatbot and continue with the Advanced one.

4.3.1 Basic chatbot. The impact of the Basic chatbot's expertise level varied among novices. Figure 4 demonstrates the count of negative, positive or neutral effect reported by the respondents as a result of the expertise of the Basic chatbot. Some found the chatbot to be easy to work with and to have an adequate knowledge, while others did not see the chatbot to have enough knowledge and suggested that the chatbot provide explanations for unfamiliar terms and offer recommendations based on the user's specific use case, such as suggesting an appropriate RAM size.

Among the experts, opinions differed as well. Some experts mentioned that the chatbot was appropriately proficient since it asked them the relevant questions for the task of finding a laptop, which they found to be useful as users. Others believed that the chatbot primarily targeted individuals with limited knowledge in the laptops field and assigned the chatbot an average level of domain expertise

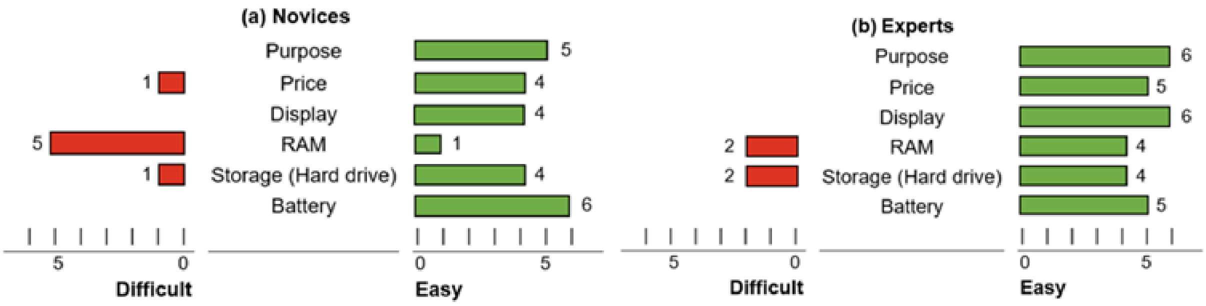


Fig. 2. A visualization showing which of the questions of the Basic chatbot were difficult to understand for the users (red), and which were easy to understand (green).

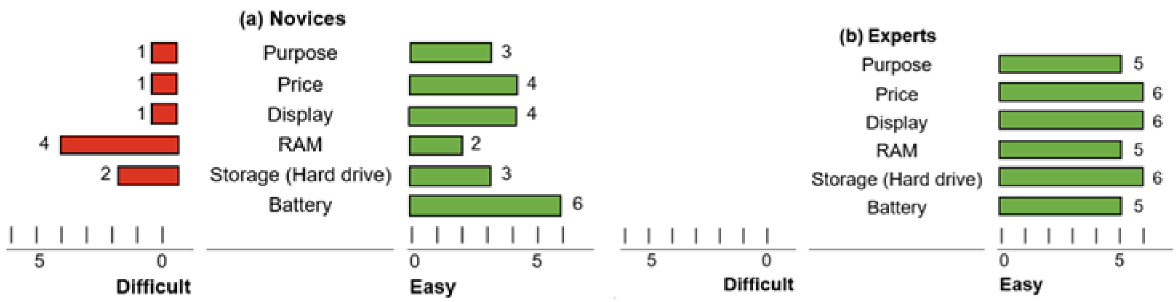


Fig. 3. A visualization showing which of the questions of the Advanced chatbot were difficult to understand for the users (red), and which were easy to understand (green).

Table 3. The main points given by both expert and novice participants regarding the questions asked by the chatbots (both Basic and Advanced)

	Basic chatbot	Advanced chatbot
Novices	- Clear language - Easy to understand - Formulate good - Language like a real person - The basic (and relevant) questions to ask when purchasing a new laptop - RAM is not well-known, therefore difficult to answer the question - Missing product aspects, such as processor, keyboard and mouse layout, and further attributes of the laptop	- Simple, concise and straightforward language - It is easy to understand because the chatbot mentions the range that exists, such as "between 10 to 18 inches" - Unsure what RAM, Hard-Drive storage and display size values are suitable - RAM is not well-known - Missing product aspects
Experts	- Clear language - Easy to understand - Formulate good - Language like a real person - Simple - Helpful language due to room for interpretation - Difficult due to ambiguous terms ("small", "big"), where the specific values are missing - Could have been a bit more formal - Missing product aspects, such as processor, GPU maker, keyboard and screen attributes, and further attributes of the laptop	- Easy language - Concise and straightforward questions - Clear language - The terminology was easy to understand - Relevant questions (covered the basic requirements of buying a laptop) - Technical terms might be unknown to novices - Could have been a bit more formal - Missing product aspects

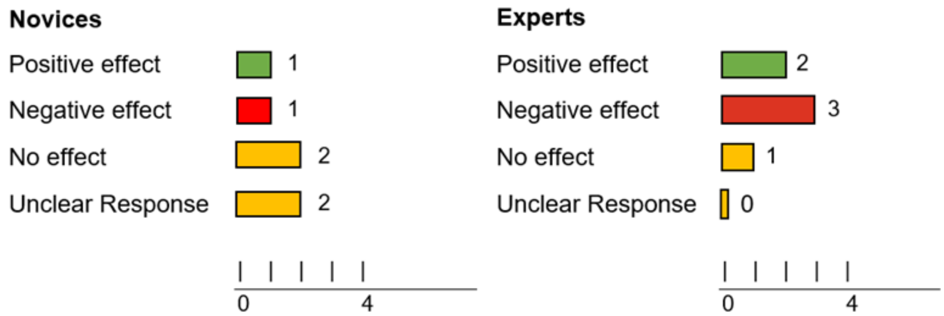


Fig. 4. A visualization showing the various number of participants who felt positive, negative or no-effect as a result of the Basic chatbot's level of expertise.

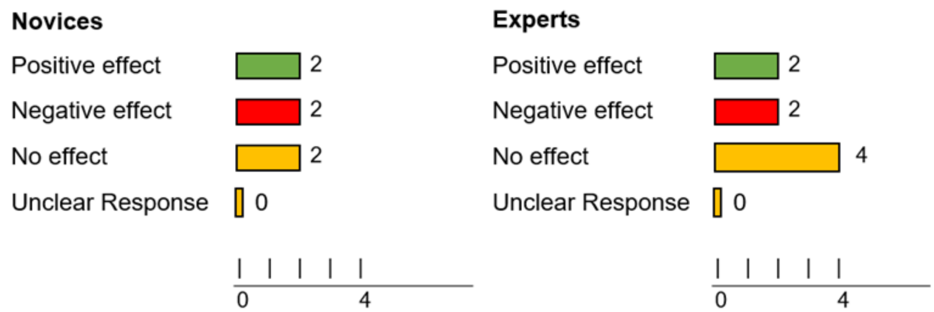


Fig. 5. A visualization showing the various number of participants who felt positive, negative or no-effect as a result of the Advanced chatbot's level of expertise.

Table 4. Users' average rating (and standard deviation) of the chatbot's degree of expertise (on a scale of 1-7).)

	Basic Chatbot	**Advanced Chatbot**
	M (SD)	M (SD)
Novice	5.2 (1.2)	5.0 (0.6)
Expert	5.2 (1.2)	5.3 (0.9)

(around 4, on a scale from 1-7). These participants recommended that the chatbot delves deeper into hardware specifications and provide more detailed explanations about the terms and their respective impacts, aligning with the suggestions made by the novices.

4.3.2 Advanced chatbot. Figure 5 demonstrates the count of negative, positive or neutral effect resulted from the expertise level of the Advanced chatbot, as reported by the respondents. Novices, much like those who used the Basic chatbot, felt that the chatbot does not have enough expertise, since it does not provide additional assistance, such as understanding unfamiliar terms and recommending appropriate capacities (such as RAM, storage, etc.) based on the user's initially stated purpose. Nonetheless, some users found the provided ranges and values to be sufficiently helpful during the conversation. The experts also did not rate the chatbot as an expert in the domain of laptops (5.3 on a scale from 1-7). However, like the novices, they felt that it covered all the fundamental requirements and found the precise ranges and values for specifications to have a positive effect on the conversation, or at least accelerate the process of finding a laptop. According to one of the respondents, this specification of terminology demonstrated the chatbot's trustworthiness: "as the chatbot used specific terminology I left the conversation feeling like I had received competent, trustworthy advice on what laptop to buy". However, the same quoted respondent mentioned that the chatbot should ask more questions to be fully trusted. Overall, it seems that the expert users were satisfied with the Advanced chatbot's level of knowledge.

5 DISCUSSION

This research focused on improving user-centered conversational product search by providing clarifying questions (CQs) that match the user's profile (novices or experts). Our results indicated that adjusting the type of CQs based on the user's profile had a greater impact on how expert users perceived the system's efficiency. Experts preferred the Advanced chatbot, while no significant distinction was found between the Basic and Advanced chatbots among novices. In fact, some novices even noted that the Advanced questions were helpful because they provided more detailed specifications. The latter was an interesting finding, and could be related to the Anomalous State of Knowledge (ASK) theory [3], which suggests that if someone searches, there is an underlying need to fill a knowledge gap. In our study, novices maybe felt the need to learn about laptops, and as a result were interested in seeing language or information that is slightly above their own knowledge level. Furthermore, our research primarily focused on the content generated by the chatbot that users see, rather than their own language when interacting with the chatbot. Therefore, it does not contradict the study [12] which indicated that novices tend to prefer using vague language.

Both novices and experts were able to understand most of the questions asked by the chatbots, indicating that the chatbots' language and content were generally clear and comprehensible, regardless the type of CQ (Basic or Advanced). However, participants faced difficulties in understanding certain questions, primarily due to encountering technical terminology that was unfamiliar for them and lacking knowledge about specific aspects of the product, especially for novices. Some novices mentioned that they did not know what is the needed capacity of some product aspects, such as RAM size. We assume it stems from their unfamiliarity of the terms and lack of knowledge in the field, as experts did not report this problem. Our research focused on vague versus technical CQs. Both types of questions included the same product aspects in it, but asked the question differently (using vague terms such as "big, small" or using exact measurements such as "8GB to 32GB"). Based on the results, we comprehend that further importance should be given to the product aspects themselves (RAM, Hard-Drive, etc.) and the ability of the chatbot to explain them to the users who do not know them (i.e., novices). Therefore, the results confirm the relevance of this research, pointing on the importance of adjusting the type of CQs to the user's profile. This supports other studies that indicated that profiling users as novices or experts in a particular domain can help tailor the selection of CQs to their needs [5, 10].

We therefore advise to adjust the system by matching the type of CQs to the user's profile. Novices may benefit from receiving CQs that explain unfamiliar product aspects, allowing them to learn and engage with the chatbot more efficiently. We recommend to explain the novices product aspects which are shown to be harder to understand, such as RAM and Hard-Drive Storage. An example for a CQ in this context: "What RAM size do you prefer? Bigger RAM size improves your laptop performance when running many tasks simultaneously, useful for heavy software such as Video-Editing and Gaming". Experts, on the other hand, should be presented with CQs that cater to their higher level of knowledge and specific requirements, saving the effort to read unnecessary information which is more relevant to novices. However, it might be useful to present both novices and experts the recommended product aspects sizes, based on the purpose of use.

Moreover, both novices and experts perceived no clear distinction in expertise level between the Basic and Advanced chatbots. This finding supports the idea that the perceived level of knowledge depends on the way the chatbot assists users in understanding unfamiliar terms and go the extra mile by offering recommendations based on the ongoing conversation.

Effectiveness is one of the components that determine the usability of a system [4]. We defined the chatbot's effectiveness as the relevance of the CQs it asks. According to the users, most of the questions asked by the chatbot were relevant for the purpose of finding a laptop. The questions which were found irrelevant were mostly those who contained unfamiliar terms for the users. This finding support the notion that the chatbot asked the CQs mostly effectively, having an impact on its general usability.

Adjusting the CQs based on the user's profile also carries risks. If novices receive basic questions, there is a possibility that they might not receive the necessary guidance or detailed information they require to make informed decisions about the product. This possibility was seen in the results of this study, when novices mentioned that some product aspects are missing. On the other hand, if experts face advanced questions, we speculate that there is a possibility of overwhelming them with unnecessary technical specifications than what they have expected (such as the exact RAM size) and making the conversation overly complex, resulting in a less efficient search experience. However, our data from the experiment does not support the last argument. In addition, misclassifying users also poses a risk. Assigning basic CQs to experts or advanced CQs to novices can lead to misunderstandings and difficulties. In our experiment, experts felt that the basic questions are too ambiguous and miss the specific values, while novices struggled to understand and engage with advanced questions.

To mitigate these risks, it is essential to accurately determine users profiles using reliable methods such as short quiz, rather than relying solely on self-assessments. This ensures that the chatbot selects the appropriate level of questions that align with user's knowledge and expertise.

Overall, the study contributes to the development of user-centered search systems by adjusting the selection of CQs based on user profiles. The findings highlight the importance of considering user's knowledge and familiarity with terminology when designing conversational search systems. Providing explanations for unfamiliar terms seems to be relevant to provide the appropriate chatbot based on the user's profile. In addition, offering recommendations on the various sizes (such as RAM size) based on the user's specific needs seems needed for all the users, regardless their expertise level.

The research has several limitations that should be considered. The study could be replicated with a larger sample to validate our qualitative findings with quantitative statistics. Additionally, the study focused on experts and novices, but the process for categorizing the participants to either experts or novices was based on self-assessments. That led to some surprising results, where individual experts were not aware of terms which novices knew about. Future similar research should involve a better method to find relevant participants for the study, e.g., using a quiz or an interview to assess the level of knowledge more reliably.

6 CONCLUSIONS

In this research, we performed a qualitative analysis that provides insights into this impact. We did so by dividing users into two groups - experts and novices. We evaluated how users perceive the system's efficiency by asking each group about their experience of using the different chatbots (Basic and Advanced).

To answer our research question, we can see that adjusting the type of CQs (non-technical, technical) based on the user's profile has a greater impact on how expert users perceive the system's efficiency. The results show that experts prefer the Advanced rather than the Basic chatbot, whereas no apparent distinction exists between the two chatbots among novices.

In general, both novices and experts recommended to have more CQs about the laptop's hardware specifications. Furthermore, they wish the chatbot could recommend them on what capacities are required to meet their needs (e.g., recommending how much RAM size a gaming laptop requires).

For further work, we suggest to conduct a quantitative research, with significant amount of participants. In addition, it may be beneficial to provide a brief explanation of the more technical terms to novice users (in the CQs themselves) and assess its impact on their perception of the system's efficiency. Another research, that seems to be relevant based on the participants' responses, should explore the impact of recommending the users the required capacity (such as RAM size) based on their initial purpose.

REFERENCES

[1] Haya Al-Thani, Tamer Elsayed, and Bernard J. Jansen. 2022. Improving conversational search with query reformulation using selective contextual history. *Data and Information Management* (2022), 100025. https://doi.org/10.1016/j.dim.2022.100025

[2] Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, and W. Bruce Croft. 2019. Asking Clarifying Questions in Open-Domain Information-Seeking Conversations. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3331184.3331265

[3] Nicholas Belkin. 2014. Anomalous States Of Knowledge As A Basis For Information Retrieval. *Canadian Journal of Information Science* (11 2014), 133–143.

[4] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (11 1995).

[5] Daniel Felix, Christoph Niederberger, Patrick Steiger, and Markus Stolze. 2001. *Feature-oriented vs. Needs-oriented Product Access for Non-Expert Online Shoppers.* Springer US, Boston, MA, 399–406. https://doi.org/10.1007/0-306-47009-8_28

[6] Jiyin He, Pernilla Qvarfordt, Martin Halvey, and Gene Golovchinsky. 2016. Beyond actions: Exploring the discovery of tactics from user logs. *Information Processing Management* 52, 6 (2016), 1200–1226. https://doi.org/10.1016/j.ipm.2016.05.007

[7] Gurmeet Manku, James Lee-Thorp, Bhargav Kanagal, Joshua Ainslie, Jingchen Feng, Zach Pearson, Ebenezer Anjorin, Sudeep Gandhe, Ilya Eckstein, Jim Rosswog, Sumit Sanghai, Michael Pohl, Larry Adams, and D. Sivakumar. 2022. ShopTalk: A System for Conversational Faceted Search. In *Proceedings of SIGIR eCom '22*. ACM, New York, NY, USA. https://sigir-ecom.github.io/ecom22Papers/paper_3793.pdf

[8] Andrea Papenmeier, Daniel Hienert, Firas Sabbah, Norbert Fuhr, and Dagmar Kern. 2022. UNDR: User-Needs-Driven Ranking of Products in ECommerce. In *Proceedings of ACM SIGIR Workshop on eCommerce (SIGIR eCom'22)*. ACM, New York, NY, USA. https://sigir-ecom.github.io/ecom22Papers/paper_6975.pdf

[9] Andrea Papenmeier, Alfred Sliwa, Dagmar Kern, Daniel Hienert, Ahmet Aker, and Norbert Fuhr. 2020. 'A Modern Up-To-Date Laptop' - Vagueness in Natural Language Queries for Product Search. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3357236.3395489

[10] Jennifer Rowley. 2000. Product search in e-shopping: A review and research propositions. *Journal of Consumer Marketing* 17 (02 2000). https://doi.org/10.1108/07363760010309528

[11] Ivan Sekulić, Mohammad Aliannejadi, and Fabio Crestani. 2021. Towards Facet-Driven Generation of Clarifying Questions for Conversational Search. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3471158.3472257

[12] Stephanie Solt. 2015. Vagueness and Imprecision: Empirical Foundations. *Annual Review of Linguistics* 1, 1 (2015), 107–127. https://doi.org/10.1146/annurev-linguist-030514-125150 arXiv:https://doi.org/10.1146/annurev-linguist-030514-125150

[13] Jie Zou, Evangelos Kanoulas, and Yiqun Liu. 2020. An Empirical Study of Clarifying Question-Based Systems. arXiv:2008.00279 [cs.IR]

[14] Jie Zou, Aixin Sun, Cheng Long, Mohammad Aliannejadi, and Evangelos Kanoulas. 2023. Asking Clarifying Questions: To benefit or to disturb users in Web search? *Information Processing Management* 60, 2 (2023), 103176. https://doi.org/10.1016/j.ipm.2022.103176

A LIST OF QUESTIONS

Table 5. The list of questions asked by the chatbots (Basic and Advance)

Product aspect	Basic chatbot	Advanced chatbot
Purpose	What do you usually use your computer for?	What do you usually use your computer for?
Price	What were you thinking of in terms of price? We have cheap, mid-priced and expensive laptops	What were you thinking of in terms of price? We have laptops from 50 to 3000 pounds
Display size	How big should the display be? Our laptops have either small, medium or large display size	How big should the display be? Our laptops have displays between 10 to 18 inches
Hard Drive Storage	How much hard drive storage do you need? Our laptops have either small, medium or big hard drive storage size	How much hard drive storage do you need? We have laptops with storage between 64 GB and 2000 GB (2TB)
RAM	What RAM size do you need? Our laptops have either small, medium or big RAM size	What RAM size do you need? Our laptops have between 1GB and 32GB RAM
Battery	What are your requirements on battery life? Our laptops have either short, intermediate or long lasting battery	What are your requirements on battery life? We have laptops that last between 6 and 20 hours