Patient Education and Counseling 105 (2022) 1488–1494

Contents lists available at ScienceDirect

Patient Education and Counseling

journal homepage: www.elsevier.com/locate/pec

Ask Rosa – The making of a digital genetic conversation tool, a chatbot, about hereditary breast and ovarian cancer

Elen Siglen[a,*], Hildegunn Høberg Vetti[a,b], Aslaug Beathe Forberg Lunde[c], Thomas Akselberg Hatlebrekke[d], Nina Strømsvik[e,f], Anniken Hamang[g], Sigrid Tronsli Hovland[h], Jill Walker Rettberg[i], Vidar M. Steen[j,k], Cathrine Bjorvatn[a,c,k]

[a] Western Norway Familial Cancer Center, Haukeland University Hospital, Bergen, Norway
[b] Faculty of Health Studies, VID Specialized University, Bergen, Norway
[c] Department of Research and Development, Haukeland University Hospital, Bergen, Norway
[d] Helse Vest IKT, Bergen, Norway
[e] Faculty of Health and Social Sciences, Western Norway University of Applied Sciences, Bergen, Norway
[f] Northern Norway Familial Cancer Center, University Hospital of North-Norway, Tromsø, Norway
[g] Department of Medical Genetics, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway
[h] Department of Medical Genetics, Oslo University Hospital, Oslo, Norway
[i] Digital culture, University of Bergen, Bergen, Norway
[j] Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway
[k] Department of Clinical Science, University of Bergen, Bergen, Norway

A R T I C L E I N F O

Article history:
Received 17 February 2021
Received in revised form 17 September 2021
Accepted 18 September 2021

Keywords:
Genetic counseling
Genetic information
Genetic communication
Information tool
Support tool
Patient educationChatbot
Chatbot technology
Artificial intelligence
Hereditary cancer
Breast cancer
Ovarian cancer

A B S T R A C T

Objective: We aimed at developing a pilot version of an app (Rosa) that can perform digital conversations with breast or ovarian cancer patients about genetic BRCA testing, using chatbot technology, to identify best practices for future patient-focused chatbots.
Methods: We chose a commercial chatbot platform and participatory methodology with a team of patient representatives, IT engineers, genetic counselors and clinical geneticists, within a nationwide collaboration. An iterative approach ensured extensive user and formal usability testing during the development process.
Results: The development phase lasted for two years until the pilot version was completed in December 2019. The iteration steps disclosed major challenges in the artificial intelligence (AI)-based matching of user provided questions with predefined information in the database, leading initially to high level of fallback answers. We therefore developed strategies to reduce potential language ambiguities (e.g. *BRCA1* vs *BRCA2*) and overcome dialogue confusion. The first prototype contained a database with 500 predefined questions and 67 corresponding predefined answers, while the final version included 2257 predefined questions and 144 predefined answers. Despite the limited AI functionality of the chatbot, the testing revealed that the users liked the layout and found the chatbot trustworthy and reader friendly.
Conclusions: Building a health chatbot is challenging, expensive and time consuming with today's technology. The users had a positive attitude to the chatbot, and would use it in a real life setting, if given to them by health care personnel.
Practice implications: We here present a framework for future health chatbot initiatives. The participatory methodology in combination with an iterative approach ensured that the patient perspective was incorporated at every level of the development process. We strongly recommend this approach in patient-centered health innovations.

1. Introduction

The increasing implementation of genetic testing in breast and ovarian cancer patients has revealed that BRCA-mutations accounts for 2–5% of breast cancer diagnoses [1,2] and 15–20% of all ovarian cancer diagnoses [2,3]. Having a pathogenic variant in one of these

* Corresponding author at: Western Norway Familial Cancer Center, Haukeland University Hospital, Jonas Lies veg 61, 5021 Bergen, Norway.
E-mail address: elen.siglen@helse-bergen.no (E. Siglen).

https://doi.org/10.1016/j.pec.2021.09.027
0738-3991/

genes determine a treatment path [4,5]. Learning that your breast cancer may be hereditary is distressing to some, especially the younger and those with less perceived social support [6]. Genetic counseling, referred to as the process of helping people understand and adapt to the medical, psychological and familial implications of genetic contributions to disease [7], may not be available within the timeframe of being diagnosed with cancer and scheduled for surgery or other treatment. Genetic testing in this timeframe, often referred to as Treatment Focused Genetic Testing (TFGT) [5], is nevertheless generally well accepted by breast and ovarian cancer patients [5,6], and knowing your BRCA status as a breast or ovarian cancer patient is on the verge of being inevitable.

The uptake of pre-symptomatic testing in BRCA-families is in the range of 30–70% for first-degree relatives, depending on gender, age, offspring and time [8]. With second-degree relatives the uptake drops, leaving us with a significant amount of at-risk relatives that never seek genetic counseling. This is a paradox as the patients accepting TFGT report undergoing genetic testing not only for themselves, but to provide their relatives with genetic information [5]. Ensuring that patients are adequately informed and ready to make decisions about their own health, including the implications for their relatives, is a major current challenge. During genetic counseling, this challenge is addressed. Traditionally a patient attends genetic counseling prior to having a genetic test and again when the test result is ready. In TFGT traditional genetic counseling is rarely achievable within most health care systems. Still, the patients need, and expect comprehensive information. Different approaches are therefore practiced to ensure the patients' informed consent before genetic testing in this setting. Some centers offer rapid genetic counseling [9], others provide a brief consultation with a clinician [2], or written information [10].

There are alternative ways to provide patients with easy access to reliable, streamlined, yet personalized, genetic information. A chatbot is a virtual assistant, designed to perform human-like digital conversations with a user [11], often about a given topic. It contains a database of predefined questions and predefined answers organized in dialogues, using artificial intelligence (AI) to match the user's question with a predefined question in the dialogues. Once matched, the corresponding predefined answer is provided. In health care, various chatbots have been made to provide veryfied medical information to the public on the web [12]. The chatbots are highly scalable, easy to use and available on demand, and may reach users in times and places where other modes cannot [13]. The complex nature of health information demands medical chatbots to be built in close partnership with health care personnel and patients [12,14]. This opens a potential for providing genetic information to the public in a safe and efficient manner. They will not replace face-to-face genetic counseling, but offer relevant high-quality information, and if trained by health care personnel, chatbots may be safer than Google [12]. However, chatbots as a tool in healthcare has yet to be robustly assessed, and clear guidance on development and evaluation is lacking [14].

Several chatbots are in use by genetic counselors, with Gia (Genetic Information Assistant) as the best known in America. Gia facilitates patient consent, provides follow up after genetic test results, and provides a sharing tool to help discuss the results with at risk family members. Focus groups with patients who had used Gia revealed that most patients enjoy the ability to explore genetic information at their own pace, and on their own time [15]. Chatbots may ensure easy access to correct medical information, but they will not be able to offer psychosocial support with today's technology. Providing information via a chatbot service about the testing process and the consequences of carrying a pathogenic genetic variant may therefore at first sound insensitive. However, the alternative may be resorting to internet searches, with the risk of misunderstandings and faulty information. For genetic counseling services, chatbots therefore hold the potential to reduce workload [16], reallocate time for more highly skilled patient care, improve accessibility of our services [17] and possibly augment family communications through empowering the patient [15].

1.1. Objective

Genetic testing has become an integral part of breast cancer diagnostics, treatment and follow-up, thereby increasing the need for genetic information and counseling services. We therefore aimed to design and develop a pilot version of an app-based digital conversation tool (the Rosa chatbot) as a tailor-made accessible and reliable source of information about hereditary breast and ovarian cancer, and to share our experiences of obstacles and best practices in this process.

1.2. Patient involvement

We decided to make the chatbot a nationwide resource through a national collaboration, and conducted workshops with patient representatives and staff from four departments of medical genetics in Norway, to map the content that was needed. We invited potential users and health carers to participate in discussions to uncover and describe the actual needs and find solutions, as well as conducting product testing.

2. Material and methods

We chose to use a commercially available platform. The chatbot from Kindly supports Norwegian language, and uses machine learning (ML) and natural language processing (NLP) [18]. This choice allowed us to focus resources on building the information content rather than on technical development, and it makes the project more relevant to health institutions without extensive in-house AI support. It took two years to complete the pilot version of the chatbot (December 2017 – December 2019).

2.1. Building a chatbot – the first steps

As the first step of building a pilot version of Rosa, an experienced genetic counselor wrote dialogues for the chatbot database, to cover the planned content of information. To ensure the quality of the medical information given to the patients by Rosa, a clinical geneticist reviewed and approved every predefined answer. In addition, an educator ensured a professional tone of voice and optimized the structure and layout of the answers to avoid misunderstandings.

In the chatbot database, the combination of a predefined question and the corresponding predefined answer (outcome) is called a dialogue (see Fig. 1). A certain predefined question can only exist in one dialogue. If the AI-based matching of a user-provided question produces two or more matches with predefined questions, the chatbot does not know which outcome to provide and will chose at random, giving either a correct answer or a completely or partially wrong answer. If a user-provided question does not lead to a corresponding match in the database, a fallback answer will be provided, stating that the chatbot does not understand the question, and asking the user to retry using different words.

2.2. User interface and chatbot administration

A mobile application and a web Application Programming Interface (API) was developed to ensure communication between the patient and the Natural Language Processor (NLP) [18]. NLP is a subgroup of AI that assists computers to understand, interpret and manipulate human language. The mobile application ("app") works

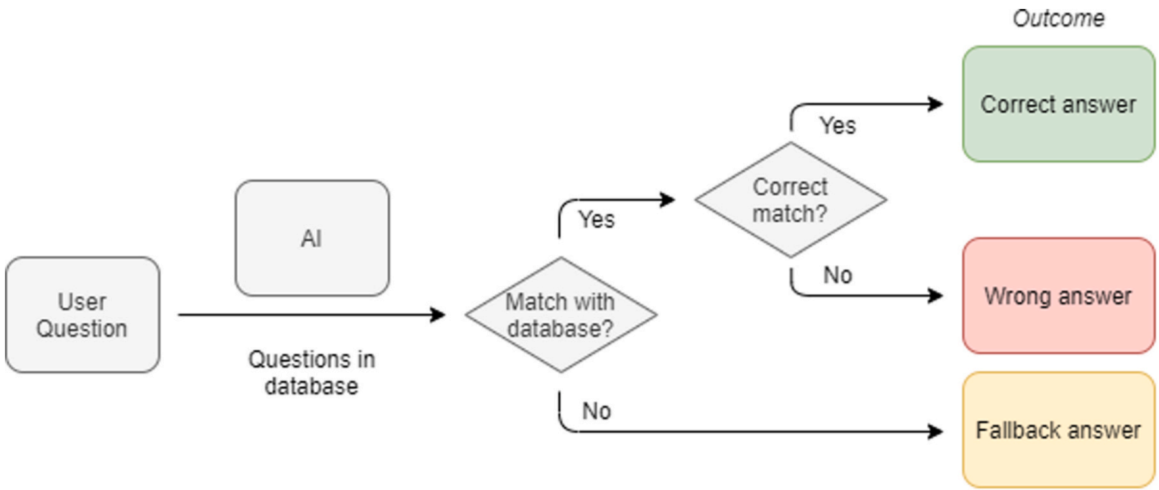

Fig. 1. Illustration of the chat process.

as the patient's interface, and was developed with Ionic [19]. A user interface framework was built on top of Angular, a development platform for building scalable web applications [20], and Cordova, a mobile development framework extending an application across more than one platform [21]. The application communicates with a web API that forwards the user provided question to the NLP, receives an answer from the NLP, and finally returns the answer to the application. Using a web API between the client and the NLP is chosen in order to keep the communication with the NLP provider anonymous. To make sure that only invited individuals access the application, it is protected with a password. Sensitive information like IP address, device identification, personal ID numbers etc. are removed before forwarding the user-provided questions to the NLP. To keep track of conversations between the users and the chatbot, a web application for administrators was implemented, using Angular. The conversations are displayed anonymously, and only chatbot administrators have access.

2.3. Design and testing

The pilot project is based on participatory methodology. This means that users take active part in all stages of the development process [22]. Participatory design is recommended by other developers of health chatbots, and should include user testing and formal usability testing by user representatives and health care personnel [23–25]. User testing refers to situations where users are testing a design in their own environment without interruption of a moderator or observer. User testing of a chatbot like Rosa will uncover missing answers, duplicate questions in different dialogues, and incomplete lists of questions associated with each answer. Identifying missing answers and incomplete lists of questions forms the basis for expanding the database. Formal usability testing refers to settings where users test a design through task performance under observation [23]. Usability has been identified as one of the factors determining the success of an application [26] and refers to the facility with which users can use a technological device to achieve a particular goal [27]. User testing and formal usability testing may be repeated until a satisfactory product has been developed, and they are an important component of health intervention development.

2.3.1. Initial workshops

The starting point of building Rosa was to conduct workshops with user representatives and medical genetic personnel throughout the country, to map the need and content of a hereditary breast cancer chatbot. All together six workshops were held, with 58 participants from four medical genetic departments in Norway providing genetic testing and counseling for hereditary breast and ovarian cancer. They were a mix of experienced genetic counselors, medical geneticists, administrative staff and laboratory staff. Two patients with verified BRCA-mutations who had previously attended genetic counseling and had undergone prophylactic surgery, reviewed the questions and answers continuously, and added questions they felt were lacking.

The first three workshops focused on delineating the scope and content of the chatbot, to come up with as many questions as possible for its' database. The collected questions were grouped thematically as many overlapped, and an experienced genetic counselor wrote the corresponding answers. We then shifted focus and organized workshops to help construct answers. This resulted in the first version of the Rosa prototype (see Table 1), consisting of information provided by experts.

2.3.2. User testing

We invited two health care personnel (genetic counselors or clinical geneticists) from four departments of medical genetics in Norway to participate, together with two patients with BRCA-mutations (patient representatives), and one research nurse. Of these 11 participants 4 had attended at least one of the previous workshops. They were given access to the Rosa prototype for four weeks with the instruction to conduct five separate chat sessions with Rosa during this period, preferably spending at least five minutes each time. They could ask Rosa any question they wanted about hereditary breast and ovarian cancer.

2.3.3. Formal usability testing

Eight new test persons working within the Western Norway Health Trust, all being unfamiliar with hereditary breast-and ovarian cancer, and two patient representatives, each solved a task in Rosa while being observed. One of the project staff (ES) registered how they used Rosa, answered their questions during the session, and received feedback in real time.

2.3.4. Iteration steps

We planned three iterations in this study; the initial workshops, followed by the user testing, and the formal usability testing steps. During the development process, we decided to add a fourth iteration with manual review of the performance of the chatbot. Every iteration refers to a significant expansion of the database, hence improving the chatbot functionality.

Table 1
The design and development of the chatbot Rosa included four iteration phases. Q – predefined question; A – predefined answer. *Fidelity level refers to how similar a training situation must be, relative to the operational situation, in order to train most efficiently [28].

	Iteration 1	Iteration 2	Iteration 3	Iteration 4
Method	Workshop	User test and manual evaluation	Formal usability test	Manual review of all answers.
No. of participants	60	9 (2 drop-outs)	9 (1 drop-out)	3
Revision level	Making prototype	Minor/major revisions	Major revisions	Major revisions
Fidelity level*	Low	Low	Low/moderate	Moderate/high
Product level	Prototype	Prototype	1.generation App	2.generationApp
Q&A in database	500/67 Ratio 7.5	850/100 Ratio 8.5	1250/111 Ratio: 11.3	2256/144 Ratio 15.7
Take home message	Nationwide collaboration is helpful to ensure successful implementation in health care	Few test persons per test, and short test periods. Repeat until satisfied	Plan thoroughly as this test is time consuming. Test persons are tolerant with new tools. Specifically ask for negative feedback.	Time consuming, but mandatory. Provides an overview of the dialogues, removes errors, and discover dialogue confusions. Makes the chatbot more robust.

2.3.5. Scoring of the performance

We manually evaluated the user-provided questions and corresponding answers that were produced during the user testing, to classify the outcomes as correct, wrong, or fallback answers (see Fig. 1). A wrong answer implied that the AI had been unsuccessful in the matching procedure. A fallback answer indicated missing or ambiguous information in the database.

3. Results

In the first iteration phase, we conduced six workshops throughout Norway with participating medical genetic personnel and patient representatives (see Table 1). The examination of the draft content of the chatbot database resulted in approval of 67 predefined answers, with 500 corresponding predefined questions (range: 1–23 predefined questions per predefined answer), giving a Q:A ratio of 7.5. Informal exploratory testing of the performance of the prototype gave an expected high fallback rate (43%).

Of the 11 individuals selected for participation in the user testing, 9 completed (1 clinical geneticists, 5 genetic counselors, 1 research nurse, and 2 patients with BRCA-mutation), and 2 dropped out (1 clinical geneticist who had previously attended one workshop and 1 genetic counselor). All of the 9 remaining test persons were experienced personnel in cancer genetics, 6 of them were new to chatbot technology, whereas 3 had attended one of the previous workshops and therefore had some insight.

The subsequent iteration step 2 included manual inspection and evaluation of all dialogues (Table 2). We found that 63% of the 822 user-provided questions were given a specified, predefined answer, while the remaining 37% got the standardized fallback answer (Table 2). Nearly 70% of the specified answers were rated as correct, still leaving a substantial fraction of wrong answers. In many cases, we observed that the wrong answers were related to problems of separating questions of almost similar wording, in particular the distinction between *BRCA1* and *BRCA2*. This observation prompted us to perform a focused re-examination of all predefined answers related to *BRCA1* vs *BRCA2*. Another challenge was related to the keyword functionality embedded in the chatbot. We had defined the Norwegian word for breast cancer ("brystkreft") and ovarian cancer ("eggstokkreft") as keywords in Kindly. If the chatbot does not understand a question containing a word defined as a keyword, a general answer regarding the keyword is given. This was not well accepted by several users, leading to a general review of keywords where among other the "breast cancer" and "ovarian cancer" keywords were deleted.

This iteration step resulted in 33 new dialogues together with expansion of number of questions, yielding a database of 100 predefined answers covering a variety of 850 predefined questions (Q:A ratio of 8.5) (Table 1, iteration 2).

7 health care personnel (1 drop-out) and 2 patient representatives took part in the formal usability testing (see Table 1,

Table 2
Overview of the user testing in iteration step 2.

Number of test persons:	9
Number of dialogues recorded:	71
Total testing time:	507 min
Testing time per dialogue (mean value):	7 min 11 s (range 3 s – 24 min)
Total testing time per test person (mean value):	56 min 20 s (range N/A)
Number of questions per dialogue (mean value):	12 (range 1 – 41)
Total number of questions asked:	822
Questions going to fallback answer:	303 (37%)
Questions that were answered:	519 (63%)
Correct answers:	359 (69%)
Wrong answers:	169 (31%)

Table 3
Overview of the answers provided by the test persons to questions about usability components.

Questions posed to participants	Answers provided by the participants			
What worked well?	Nice layout, easy to use (5)	Friendly and trustworthy answers (5)	Easy access to quality assured information. Fast replies (3)	Reader friendly Like the split chat bubbles (2)
What did not work well?	Too many wrong answers (3)	Too many fallbacks (2)	Struggle to find the right way to ask questions/extract the answers (3)	Subtitles in videos were hard to read (1)
How did you feel when given a wrong answer?	I do not mind, it provides information and gives me ideas of other things to ask. Must be limited to keep me in. (3)	Makes me frustrated. Makes the dialogue fragmented (2)	I rephrase myself/ try again with different words (1)	I bear with it. I know I am talking to a computer and that this is a test. Would annoy me if this was the end product (4)
How did you feel when given a fallback answer?	Frustrated. I don't know how to pose the question in the "right" way. Potentially stops the dialogue (3)	I don't mind, I rephrase myself/ try again with different words (3)	I bear with it. I know I'm talking to a computer and that I'm part of a test. (3)	
Do you miss anything? What?	User evaluation in the chatbot (thumb up/down-button on answers). Or an "alternative answer" button.(2)	Menu with contact information to genetic counselor (6)	Index / FAQ (4)	References/Provide the sources of information used (1)
Do you think Rosa fills a need?	Yes (3)	Yes, if it works (4)	Yes, if recommended by health care personnel (1)	Yes, if the chatbot answers tough/sensitive questions (1)

In all questions, the maximum alternatives of answers given by the participants were four. This was not a limitation by the researchers in the study, the answers merely didn't vary more. In parenthesis is the number of test persons providing the answer.

iteration 3). Each of them spent between 40 and 65 min (mean: 51 min) (data not displayed), exploring the functionality of the chatbot. All users liked the layout of the app and stated it was easy, intuitive to use and understandable, and that it had a friendly tone of voice (see Table 3). They were frustrated by not knowing what information was available in Rosa and how to ask the best questions in order to extract this information. All suggested including a FAQ functionality to get started.

We observed that some users tended to write long and explanatory questions. When asked about this observation they said this was to ensure that the chatbot understood their question correctly. Others chose to write questions in a keyword format or SMS-like style, similar to googling. When asked about this approach, they said they wanted to see what predefined answers that existed in the chatbot about particular topics.

Iteration 3 (see Table 1) led to a major revision of the chatbot, based on the experiences and feedback from the formal usability testing. The IT team was involved to embed a menu providing information about privacy, a presentation of Rosa and the research team, brief general information about hereditary breast and ovarian cancer, and a FAQ page. We also extended the chatbot Q&A database, reaching a total of 111 dialogues covering 1250 predefined questions (Q:A ratio of 11.3).

The second and third iteration steps disclosed a major challenge with a rather high rate of fallback answers and wrong answers. As a consequence, we added a fourth iteration (see Table 1) of the database that increased the number of dialogues to 144 and almost doubled the number of predefined unique questions to a total of 2256 with a Q:A ratio of 15.3 (range: 1–56 questions per answer). Preliminary performance testing after this revision indicated a fallback rate below 15%.

4. Discussion and conclusion

4.1. Discussion

In this project, we have built a pilot version of a chatbot for hereditary breast and ovarian cancer, with the future goal of using this digital tool as a supplement to traditional face-to-face genetic counseling. The process proved to be markedly more time consuming and challenging than anticipated, running for 2 years from the start of the practical work to the completion of the pilot version. In line with participatory methodology, the chatbot was created with user involvement at all levels, involving both health care personnel and patients that had signed up as patient representatives. This combination proved successful for defining the medical information content and predefined answers in the database, as well as checking the usability of the chatbot in a realistic setting. The process included workshops, user testing and formal usability testing, in parallel with a gradual improvement of the database content and user interface through four iteration steps. It is worth noticing that the database initially contained 67 predefined answers and 500 corresponding questions (Q:A ratio of 7.5), whereas the finalized pilot version has 144 predefined answers and 2256 corresponding questions (Q:A ratio of 15.7), as a response to the experience that was obtained during the development of Rosa.

The fallback rate decreased with expansion of the database, from 43% in the first prototype, to 13% in the fourth iteration. This trend is expected as each dialogue needs at least 20 related predefined questions for the embedded AI to work optimally [18]. Very few dialogues met this criterium in the prototype, whereas almost all did in the final app. There is a general agreement among companies delivering chatbot platforms that the acceptable level of fallback should be below 10%. This threshold is not documented in research. Future use and testing of Rosa will gradually increase the number of predefined questions in the database, leading to further drop of fallback. What is not addressed in chatbot literature is the acceptable level of wrong answers. Some amount of incorrect selection of answers, referred to as dialogue confusion, is inevitable with today's technology. It is therefore crucial to develop strategies to avoid misunderstandings and faulty advices.

The challenge with dialogue confusion became obvious after the user testing and subsequent evaluation of the dialogues in iteration 2, which disclosed that more than a third of the user-provided questions led to a general fallback answer, and among the questions that got a specific answer, about 30% did not properly answer the question asked. This implied that less than half of all user-provided questions were successfully matched with the predefined questions in the database. This problem was mainly related to inadequate distinction between answers and questions with almost similar wording for essential information content, e.g., *BRCA1* versus *BRCA2*. If the user asks "What is the risk of ovarian cancer with BRCA1 mutation?", Rosa may respond by matching it with a *BRCA2*-related dialogue, with only a single digit difference from *BRCA1*, which will be wrong. To make the chatbot able to differentiate questions at this level, the threshold for similarity of the wording would need to be almost 100%. However, this threshold will disable the AI function and leave the chatbot with providing only answers that match perfectly, leading to a high fraction of fallback answers. We solved this problem by merging predefined answers concerning *BRCA1* and *BRCA2*, and make joint dialogues.

Still, a certain number of wrong answers seems unavoidable with the present chatbot technology. To ensure that the user can detect such potentially harmful situations, we decided that all answers should include a full sentence that explain the context. As an example, the user-provided question "What is the risk of breast cancer?" will be answered by the chatbot as "For women carrying a *BRCA1* mutation the life time risk of breast cancer is….". This highlights the importance of high involvement by experienced personnel when building the content in a health information chatbot. Conversational design is very different from designing user interfaces, and experience with human to human conversations about the given topic is crucial in developing robust dialogues [25].

The keyword function of the chatbot needs to be used carefully, to reduce the number of general answers about a topic. If a user-provided question contains a word defined as a keyword and the chatbot fails to match the question in the database, the chatbot will respond with the corresponding keyword reply instead of the general fallback reply. Our definition of breast cancer and ovarian cancer ("brystkreft" and "eggstokkreft" in Norwegian, respectively) as keywords were done to avoid questions about these cancers to go to fallback. During the user test, several participants contacted the administrator saying that the breast/ovarian cancer replies were annoying, as they were given to any breast or ovarian cancer question Rosa failed to match. They would rather prefer the fallback answer. These keywords were therefore removed after this test, along with a general review of all keywords.

The formal usability testing demonstrated that all users liked the layout of the app and found it easy to use. All user testers felt frustrated not knowing what answers were available in the app and how to propose the right questions in order to receive these answers. They all looked for a FAQ to get themselves started. We noticed that several of them tended to write long and explanatory questions, as they assumed that this was necessary to make the chatbot understand their questions correctly. This indicates that many users have a high trust in the embedded AI. However, with the chatbot technology available today, long explanatory questions are actually counterproductive, increasing the risk of fallback or wrong answers. Others chose to write questions similar to googling, since they thought that approach could disclose all answers that existed about that particular topic. As chatbot technology is something most people are unfamiliar with, it is only natural to approach it the way you would approach either a human conversation, or a google search. However, as a chatbot provides the one most fitting answer to the user-provided question, this strategy will increase the fallback rate.

Iteration four expanded the dialogue database both in turns of multiple dialogues, and markedly increased the number of questions per dialogue. With a participatory methodology approach, building a chatbot is teamwork. With every iteration, there is a large amount of rebuilding work. High involvement by both end users and health care personnel throughout the process is vital, making validation of the predefined chatbot-provided answers by experts a crucial part of the process. Chatbots may be seen as depersonalized, cold and inhuman, however the motivation to try out health chatbots is high as they are considered time-saving and accessible [24]. This motivation ensures there is a future for chatbots. As humans of the twenty-first century, we are used to online availability. A chatbot provides the information we seek in an understandable manner, without the need of filtrating multiple hits as with internet/google searches.

4.1.1. Study limitations and future research

In the process of making a pilot version of the chatbot Rosa we chose an expert panel (the initial workshops) to select topics for the database. All together 58 clinical geneticists, experienced genetic counselors, administrative staff and laboratory staff contributed, based on their knowledge of what the patients normally ask for before, during and after genetic counseling for hereditary breast and ovarian cancer. We invited two patient representatives to serve as consultants in this process. They were both women with a BRCA mutation, who had undergone prophylactic surgeries. This setting made them also experts in the field, providing the patient's perspective at every stage. Ideally, patient representatives should have been present at every workshop, preferably different participants each time. To overcome this limitation we have initiated in depth interviews with patients given access to Rosa before, during and after genetic counselling and testing. The results from these interviews will be used in a fifth iteration, to produce a final app version, and thus mark the end of the pilot phase.

We strongly suggest future research about chatbots in health care to focus on reducing fallback answers, overcoming dialogue confusion and develop strategies for successful implementation in health care services.

4.2. Conclusions

Successful implementation of chatbots in health care calls for tools that correctly analyze the user's question to provide the correct answer in return [29]. We have documented the challenging and time-consuming process of building a chatbot that can serve as a source of information about hereditary breast and ovarian cancer. When AI reaches the level of being able to create adequate answers itself and recognize the nuances in written language that differentiate two similar questions, we expect to observe a paradigm shift in the use of chatbots in health care. In genetic counseling services, the patients will have the ability to prepare and educate themselves before meeting the genetic counselor, through the chatbot giving them correct information, in their own environment at their own pace and time. Both patients and genetic counselors may benefit from chatbots as counseling sessions may be even more personalized and tailored to the need of the patients. Chatbots may thus serve as the perfect companion to genetic counseling.

The beauty of a chatbot is in its simplicity. This simplicity is also a chatbot's main threat. The manual labour required to build a robust chatbot is still substantial. A chatbot does not have to provide a perfect human-like conversation; in fact, it should not be mistaken for a human, it should be valued for what it is, a support tool at your service providing accurate information in lay language. As stated by Powell [30] artificial intelligence used in health care must pass the implementation game rather than the imitation game.

4.3. Practice implications

See Table 4.

Table 4
Overall take home messages from the chatbot builders and the test persons in both tests.

Administrator's experiences	Users' experiences
Initiate nationwide or multicenter collaboration to increase likability of successful implementation.	It's reader friendly and understandable. They like the lay out.
Use participatory methodology to ensure that patient perspective is incorporated at every iteration.	It provides information in an understandable manner, without the need of sorting through multiple hits as with internet searches.
Plan testing thoroughly and ask specifically for negative feedback	Some level of fallback and wrong answers is accepted, however aspire to keep it as low as possible.
Schedule few test persons and short test periods. Repeat rather than expand.	It must include a FAQ or Menu that provides an overview of the chatbot content.
Conduct expert review of all dialogues to ensure correct outputs and reduce dialogue confusions.	It provides easy access to correct information. Available 24/7.
Ensure all answers are in full sentences, and include all information needed to understand the answer correctly, so that possible cases of dialogue confusions are detected by the user.	Will use it if recommended by health care personnel.

Funding and acknowledgements

The study is funded by the Norwegian Cancer Society, grant number 194750-2017. The authors would like to thank the four medical genetics departments in Norway contributing to the study, both with user testing and with composing questions and answers for the chatbot database.

CRediT authorship contribution statement

Elen Siglen: Investigation, Methodology, Data curation, Formal analysis, Visualization, Writing original draft, Writing review & editing, Hildegunn Høberg Vetti: Methodology, Data curation, Supervision, Writing original draft, Writing review & editing, Aslaug BF Lunde: Data Curation, Writing review & editing, Thomas Hatlebrekke: Resources, Software, Data curation, Writing review & editing, Nina Strømsvik: Resources, Writing review & editing, Anniken Hamang: Resources, Writing review & editing, Sigrid T Nergård: Resources, Writing review & editing, Jill W Rettberg: Writing original draft, Writing review & editing, Vidar M Steen: Conceptualization, Funding acquisition, Writing original draft, Writing review & editing, Cathrine Bjorvatn: Conceptualization, Funding acquisition, Project administration, Supervision, Writing original draft, Writing rewiew & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Sun J, Meng H, Yao L, Lv M, Bai J, Zhang J, et al. Germline mutations in cancer susceptibility genes in a large series of unselected breast cancer patients. Pers Med Imaging 2017;23:6113–9. https://doi.org/10.1158/1078-0432.CCR-16-3227%. (J Clinical Cancer Researc).
[2] Høberg-Vetti H, Bjorvatn C, Fiane BE, Aas T, Woie K, Espelid H, et al. BRCA1/2 testing in newly diagnosed breast and ovarian cancer patients without prior genetic counselling: the DNA-BONus study. Eur J Hum Genet 2016;24:881–8. https://doi.org/10.1038/ejhg.2015.196
[3] Pal T, Permuth-Wey J, Betts JA, Krischer JP, Fiorica J, Arango H, et al. BRCA1 and BRCA2 mutations account for a large proportion of ovarian carcinoma cases. Cancer 2005;104:2807–16. https://doi.org/10.1002/cncr.21536
[4] Tung NM, Boughey JC, Pierce LJ, Robson ME, Bedrosian I, Dietz JR, et al. Management of hereditary breast cancer: American society of clinical oncology, American society for radiation oncology, and society of surgical oncology guideline. J Clin Oncol 2020;38:2080–106. https://doi.org/10.1200/jco.20.00299
[5] Wright S, Porteous M, Stirling D, Lawton J, Young O, Gourley C, et al. Patients' views of Treatment-Focused Genetic Testing (TFGT): some lessons for the mainstreaming of BRCA1 and BRCA2 testing. J Genet Couns 2018;27:1459–72. https://doi.org/10.1007/s10897-018-0261-5
[6] Høberg-Vetti H, Eide GE, Siglen E, Listøl W, Haavind MT, Hoogerbrugge N, et al. Cancer-related distress in unselected women with newly diagnosed breast or ovarian cancer undergoing BRCA1/2 testing without pretest genetic counseling. Acta Oncol 2019;58:175–81. https://doi.org/10.1080/0284186X.2018.1502466
[7] National Society of Genetic Counselors' Definition Task F, Resta R, Biesecker BB, Bennett RL, Blum S, Hahn SE, et al. A new definition of genetic counseling: national society of genetic counselors' task force report. J Genet Couns 2006;15:77–83. https://doi.org/10.1007/s10897-005-9014-3
[8] Forde C, Brunstrom K, Woodward E, Bowers N, Pereira M, Wallace AJ, et al. Uptake of pre-symptomatic testing for BRCA1 and BRCA2 is age, gender, offspring and time-dependent. J Med Genet 2020;0:1–5. https://doi.org/10.1136/jmedgenet-2019-106544
[9] Wevers MR, Ausems MG, Verhoef S, Bleiker EM, Hahn DE, Brouwer T, et al. Does rapid genetic counseling and testing in newly diagnosed breast cancer patients cause additional psychosocial distress? Results from a randomized clinical trial. Genet Med 2016;18:117–26. https://doi.org/10.1038/gim.2015.50
[10] Nilsson MP, Törngren T, Henriksson K, Kristoffersson U, Kvist A, Silfverberg B, et al. BRCAsearch: written pre-test information and BRCA1/2 germline mutation testing in unselected patients with newly diagnosed breast cancer. Breast Cancer Res Treat 2018;168:117–26. https://doi.org/10.1007/s10549-017-4584-y
[11] Kearney E, Wojcik A, Babu D. Artificial intelligence in genetic services delivery: utopia or apocalypse? J Genet Couns 2020;29:8–17. https://doi.org/10.1002/jgc4.1192
[12] Greene A, Greene CC, Greene C. Artificial intelligence, chatbots, and the future of medicine. Lancet Oncol 2019;20:481–2. https://doi.org/10.1016/S1470-2045(19)30142-1
[13] Joerin A, Rauws M, Fulmer R, Black V. Ethical artificial intelligence for digital health organizations. Cureus 2020;12:e7202https://doi.org/10.7759/cureus.7202
[14] Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng YL, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020;22:e17158https://doi.org/10.2196/17158
[15] Nazareth S, Nussbaum RL, Siglen E, Wicklund CA. Chatbots & artificial intelligence to scale genetic information delivery. Chatbots Artif Intell Scale Genet Inf Deliv 2021;30:7–10. https://doi.org/10.1002/jgc4.1359
[16] L. Reis, C. Maier, J. Mattke, T. Weitzel Chatbots in Healhcare: Status quo, application scenarios for physicians and patients and future directions Proceedings of the 28th European Conference on Information Systems (ECIS) 2020 An Online AIS Conference,.June 15-17, 2020⟨https://aisel.aisnet.org/ecis2020_rp/163⟩.
[17] Schmidlen T, Schwartz M, DiLoreto K, Kirchner HL, Sturm AC. Patient assessment of chatbots for the scalable delivery of genetic counseling. J Genet Couns 2019;28:1166–77. https://doi.org/10.1002/jgc4.1169
[18] ⟨https://www.kindly.ai/en/products/⟩, 2021 (accessed 01 January 2021).
[19] ⟨https://ionicframework.com/docs⟩, 2021 (accessed 01 January 2021).
[20] ⟨https://angular.io/docs⟩, 2021 (accessed 01 January 2021).
[21] ⟨https://cordova.apache.org/docs/en/latest/⟩, 2021 (accessed 01 January 2021).
[22] Simonsen JE, Robertson T. Routledge International Handbook of Participatory Design. London: Routledge,; 2012.
[23] Merle C. User and usability testing - how it should be undertaken? Australas J Educ Technol 1995;11. https://doi.org/10.14742/ajet.2075
[24] 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;5:1–12. https://doi.org/10.1177/2055207619871808
[25] Cameron G, Cameron DW, Megaw G, Bond D, Mulvenna M, O'Neill S, et al. Best practices for designing chatbots in mental healthcare - a case study on iHelpr. Proceedings of the 32nd International BCS Human Computer Interaction Conference 2018;129:1–5. https://doi.org/10.14236/ewic/HCI2018.129
[26] Baharuddin R, Singh D, Razali R. Usability dimensions for mobile applications - a review. Res J Appl Sci Eng Technol 2013;5:2225–31. https://doi.org/10.19026/rjaset.5.4776
[27] Zapata BC, Fernández-Alemán JL, Idri A, Toval AIA. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015;39:1. https://doi.org/10.1007/s10916-014-0182-2
[28] Hays RT, Singer MJ. Simulation fidelity in training system design. Bridging the Gap Between Reality and Training. Springer-Verlag,; 1989.
[29] Reshmi S, Balakrishnan K. Implementation of an inquisitive chatbot for database supported knowledge bases. Sādhanā 2016;41:1173–8. https://doi.org/10.1007/s12046-016-0544-1
[30] Powell J, Me Trust. I'm a chatbot: how artificial intelligence in health care fails the turing test. J Med Internet Res 2019;21:e16222https://doi.org/10.2196/16222