Conversational Agents as an Aid for Cancer Survivors' Information Search

Barikisu Issaka
Advertising & Public Relations
Michigan State University
issakaba@msu.edu

Young Anna Argyris
Media & Information Michigan
State University
argyris@msu.edu

Pang-Ning Tan
Computer Science & Engineering
Michigan State University
ptan@msu.edu

Abstract

Cancer survivorship often lasts over 20 years and affects over 18 million Americans. As such, cancer survivors need high-quality information to self-manage their illness, but digital health technologies to aid cancer survivors' navigation through the wealth of information are still in their infancy. Simultaneously, cancer misinformation has grown exponentially on digital platforms, veering survivors from conventional therapies to untested alternative treatments. Towards this end, this conceptual paper explores the potential of conversational agents (generative artificial intelligence aids) designed to select and present credible information to aid in survivors' long-term journey of managing cancer. Specifically, mining cancer (mis)information from diverse sources and building knowledge graphs are suggested to structure trustworthy and relatable dialogues of conversational agents for cancer survivors.

Keywords: generative artificial intelligence; conversational agents; cancer survivorship; information search; misinformation; knowledge graph

1. Introduction

Cancer survivors actively seek information for managing their long-term cancer trajectory, ranging from diagnosis to recovery. In 2022, there are 18.1 million cancer survivors in the United States (US), representing roughly 5.4% of the population. With our aging society, the number of cancer survivors is projected to increase by 24.4% to 22.5 million by 2032 (Cancer Net, 2022).

Although there are multiple definitions of cancer survivorship (Marzorati et al., 2017), the National Coalition for Cancer Survivorship defines a cancer survivor as an individual from the moment of cancer diagnosis throughout their life experiencing different stages of survival. This aligns with the perspective of Dr. Mullan, who initially referred to himself and fellow cancer patients as resilient "survivors" regardless of their life expectancy (Marzorati et al., 2017). For this reason, several prior studies use the term cancer survivors interchangeably with cancer patients (e.g., Chua et al., 2021). Following this convention, we refer to individuals as cancer survivors from the moment they are diagnosed with cancer and undergo and complete their treatments. According to this definition, many cancer survivors (over 18%) spend 20 or more years after diagnosis (Cancer Net, 2022).

To manage this long-term trajectory, cancer survivors need unique information to track complicated treatment options, address the fear of recurrences and side effects from treatments, and initiate lifestyle changes. The rapid growth of health misinformation on digital platforms complicates their search for such information. Although technology can help cancer patients and their caregivers receive better support, these online platforms are also riddled with misinformation (Warner et al., 2022). Many sources offer treatment alternatives and recommendations that are medically inaccurate, incomplete, and dangerous (Gentile et al., 2018). Johnson et al.'s study (2022) on the accuracy of cancer treatment information on social media showed that of 200 social media posts that cancer experts reviewed, 32.5% contained misinformation, and 30.5% contained harmful information. Among articles containing misinformation, 76.9% were deliberately produced and disseminated by manufacturers of alternative medicine. The New York Times article entitled "Dr. Google is a Liar (Warraich, 2018)" states:

> *"Cancer is another big target for pushers of medical misinformation—many of whom are making money off alternative therapies…when cancer patients turn to alternative therapies…they are 2.5 times more likely to die. By exploiting people's fears, those who dissuade patients from getting evidence-based treatment have blood on their hands."*

Cancer survivors' dire need for information to manage their illness amid the growth of cancer misinformation calls for interventions. Among many interventions, we are interested in a generative artificial intelligence (AI) tool—i.e., conversational agents (CAs). The recent spotlight on CAs is primarily attributed to ChatGPT. A CA is defined as *an artificial*

URI: https://hdl.handle.net/10125/106840
978-0-9981331-7-1
(CC BY-NC-ND 4.0)

intelligence system that mimics human conversation, either using text or spoken language (Rheu et al., 2021). The global CA market is expected to grow by 22% annually between 2022 and 2025, reaching nearly US$14 billion by 2025 (Comes et al., 2021). CAs are now the most used AI technology in business organizations, and their adoption rates grew during the pandemic (Comes et al., 2021).

Although CA's potential in general healthcare, as an aid for clinical decision-making and fostering lifestyle changes, has been established, CA's capacity to aid cancer patients in their information search has yet to be fully realized. This void in research is manifested in ovarian and cervical cancers. Very few attempted to develop technological aids for these cancer patients compared to much attention to other cancer types (e.g., breast and colon cancers). Blindly creating a generic CA for all cancer types will not fulfill ovarian and cervical cancer survivors' particular needs, as each involves distinct treatment options and challenges.

Accordingly, the main objective of this study is to explore the potential of developing CAs as a self-management tool for ovarian and cervical cancer survivors. To attain this objective, we provide a comprehensive review of the literature on cancer survivors' information needs, obstacles to filling those needs, and the potential of CAs in fulfilling those needs. Further, we note the hurdles of using CAs as a self-management tool, including unintuitive interactions, awkward sequences of dialogues, and users' distrust of AI systems in general. Based on these analyses, we suggest the development of knowledge graphs constructed on cancer (mis)information mined from online sources (such as websites and social media).

In so doing, we apply the design science framework along the spectrum of "identify – develop – evaluate." Specifically, we elucidate how CAs' sequence in conversing with cancer survivors should be developed based on knowledge graphs so that the dialogues progress naturally and intuitively, fulfilling their information needs. These intuitive and relatable CAs will fulfill ovarian/cervical cancer survivors' information needs, veering them away from misinformation they are exposed to online. Finally, we conclude with this study's contributions to advancing theories.

2. Literature Review

2.1 Cancer Survivors' Information Needs and Search Behaviors

Numerous studies have been conducted to shed light on cancer survivors' information-seeking behaviors. A survey involving 3,300 colorectal cancer survivors revealed that > 20% of participants actively sought information and advice regarding diet and lifestyle, but 30% of these survivors were dissatisfied with the level of the information supplied in this domain (Department of Health-Quality Health, 2012).

Cancer survivors experience long-term health and psychosocial implications that distinguish their information needs from those of general patients (Rutten et al., 2005). Firstly, many cancer survivors clarify information they have obtained from healthcare providers (Wang et al., 2018). About 40% of cancer survivors did so outside of clinical settings, using the Internet, their friends and family, other patients, and support groups as their sources of information (Schook et al., 2014). Secondly, cancer survivors seek information on managing recurrence, hereditary passing to descendants, and any long-term consequences that may arise due to their illnesses and treatments (Kent et al., 2012). Thirdly, patients search for information to take charge of their treatment plans because many participants were concerned about the side effects of cancer treatments (Budenz et al., 2022). Finally, cancer survivors need information to cope with psychosocial difficulties, such as stress management, anxiety, and concern for family, friends, and partners (Lisy et al., 2019).

What further complicates the understanding of cancer survivors' information needs is their demand for divergent information depending on their cancer types. Participants preferred to look for information specific to their cancer types rather than general cancer information (Budenz et al., 2022). For instance, breast cancer survivors seek diet and nutrition advice due to the higher risks for recurrence and the resultant need for long-term monitoring and lifestyle changes (Chua et al., 2021).

2.2 Difficulty in Discerning Quality Information

The abundance of information available yet the lack of assistance in selecting credible content impedes cancer survivors' information seeking. Marchetti et al.'s study on skin cancer survivors (2022) demonstrated that, while health- and cancer-related information-seeking behaviors were prevalent, 21.6% of respondents experienced frustration, and 28.2% said that they had to spend a significant amount of effort finding needed information (28.2%).

The sheer quantity of available information sources, such as websites, blogs, books, podcasts, social networks, and mainstream news, exacerbates this frustration in finding reliable information. While young cancer survivors prefer technology-based information sources as varied as the Internet, wellness apps, or YouTube (Meneses, 2010), excessive information from a multitude of sources leads to information overload, hampering the comprehension of information for informed decision-making (Gentile et al., 2018). For instance, a study conducted in the US focusing on sun-safe behaviors and skin cancer survivors found that participants experienced high information overload (Jensen et al., 2020). Likewise, Hu et al. (2021) noted the absence of guidelines for authoritative, credible information amid extensive information channels contributed to stress among breast cancer survivors in China (Hu et al., 2021). This abundance of information in the absence of assistance results in cancer survivors' non-adherence to treatments (Eraslan et al., 2022). Patients who reported high levels of information overload exhibited elevated levels of depression and anxiety, which are counterproductive to treatment adherence (Eraslan et al., 2022).

2.3 Growing Cancer Misinformation

One significant consequence of experiencing information overload is susceptibility to encountering and grappling with misinformation. When an individual experiences an influx of information, it becomes increasingly more complex for an average person to discern credible content from non-credible ones without clear guidelines. A study examining information-seeking patterns among testicular, lung, and colorectal cancer patients demonstrated that most individuals relied on informal sources, such as online platforms, media outlets, and input from others, to obtain diet information. However, this reliance on informal sources led to encounters with misinformation, needing clarification on what constitutes a balanced diet for survivors (Beeken et al., 2016).

This vulnerability is critical because of the rapid spread of cancer misinformation on digital platforms. Misinformation regarding food and other natural sources, such as herbs, remains a common source of cancer misinformation. When cancer misinformation posts were broken down by topic on Pinterest, 39.6% discussed foods, including the "Mediterranean diet for breast cancer," that either purportedly cause, treat or prevent cancer. Others asked readers if they were using a specific breast cancer medication, suggesting that ten meals may "completely prevent breast cancer." Most of those food items had probiotics, without conclusive evidence of their efficacy among early-stage breast cancer patients. The use of turmeric was encouraged by other pins (Wilner et al., 2020). Warner et al. (2022) discovered that much of the false information cancer survivors encountered focused on nutrition and food, such as recommending herbs and foods, with about 40% of these claims making use of buzzwords such as "anti-cancer," "cancer-fighting," and "cancer cell killing." Regarding herbs, Zenone et al. (2020) observed widespread misinformation about the potency of Cannabinoids. Cancer survivors are now interested in this substance, which may also be the subject of health-related myths about it being a panacea for various ailments (Zenone et al., 2020).

Complementary and alternative medicine is also one of the most sought-after topics online among cancer survivors (Nguyen et al., 2013). Various medical healthcare techniques, products, and systems not typically considered part of traditional medicine are referred to as complementary and alternative medicine. A study that sought to identify the misinformation cancer survivors encounter online revealed that participants highlighted outdated cancer survival rates, inaccurate information about alternative treatments, and other breast cancer patients' experiences that did not align with their own experiences (Perrault et al., 2020). Other types of misinformation typically arise from spiritual and religious views that cast doubt on scientific findings (Peterson and Ivengar., 2022).

2.4 Technology Interventions for Cancer Management

One reason cancer survivors defied conventional therapies and utilized online information suggesting complementary and alternative medicine is the lack of assistance from their social circles and medical teams in search of credible information necessary for the self-management of cancer (Holmes et al., 2017). To address the challenges cancer survivors experience in their long journey to treat cancers, digital health technologies, including artificial intelligence (AI) tools, have emerged. Indeed, the field of cancer survivorship care has witnessed remarkable progress with AI-driven digital health technologies.

AI-based solution tools were first introduced for data analysis and management in cancer patients to predict recurrence, survivability, and adverse treatment reactions. For instance, Torrente et al. (2022) demonstrated the potential of data-driven AI solutions in estimating cancer-specific survival rates, stratifying patients by risks, and improving follow-up management across different cancer pathologies. Gangganayh et al. (2021) proposed a fully automated clinician-friendly AI-enabled database platform for breast cancer survival prediction. Their digital platform, iSurvive, encompasses a comprehensive range of features, including a database, digitized questionnaires, automated machine learning, and interactive visualizations. This platform facilitates efficient data collection and management to assist clinicians in treating cancers among their patients.

A second set of AI tools for digital health solutions is wearable technology for cancer survivors. Lynch et al. (2019) shed light on the potential of wearable technology as an affordable and scalable opportunity to promote active lifestyles among cancer survivors. A systematic review (Blount et al. 2021) synthesizes evidence from randomized controlled trials and controlled trials investigating the effects of wearable health technology-based physical activity interventions on physiological, cognitive, and emotional outcomes in breast cancer survivors. Their findings indicate that wearable health technology reduces sedentary behavior and increases moderate-to-vigorous intensity physical activity. This increase in physical activity is associated with improvements in cognitive function, attitude, and reduction in anxiety and worry, showcasing the efficacy of wearable health technology in enhancing the well-being of breast cancer survivors.

The third set of tools is mHealth applications that patients can install on their mobile devices to self-manage their cancer journey. Examining the landscape of mHealth self-management apps for breast cancer survivors, Kapoor et al. (2020) conducted a content analysis to assess the inclusion of key features derived from the chronic care model. These mHealth applications provide education for cancer survivorship and opportunities to network with other patients and healthcare providers. However, this study highlights the need for comprehensive mHealth resources, emphasizing the need for survivor and healthcare provider involvement in app development to address unmet needs.

Using these tools, patients have mentioned increased motivation in reporting their outcomes, leading to improved patient-reported outcomes. Moreover, these tools have demonstrated the ability to reduce fatigue and pain levels, enhance overall quality of life, and improve physical function. Rupert et al. (2013) developed "Cancer in the Family," an interactive tool fueled by AI to counsel women with increased risks of hereditary breast and ovarian cancer. Through a pilot evaluation, the study revealed the tool's effectiveness in enhancing patients' knowledge, cooperation in providing comprehensive family history, and initiation of a visit to healthcare providers for screenings for hereditary breast and ovarian cancer. Finally, AI-powered chatbots were proposed as information portals, enabling users to search using natural language. These interactive chatbots are crucial in facilitating information search, clarifying ambiguous information needs, and assisting users in formulating complex requests (Xiao et al., 2023).

2.5 Unattended Ovarian and Cervical Cancer

There is a gap in the literature on technological interventions for cervical and ovarian cancers. Prior studies have focused on specific cancer types, such as breast, prostate, testicular, and lung cancers (e.g., Rupert et al., 2013; Svoboda, 2020; Van Booven et al., 2021), but very few have paid attention to cervical and ovarian cancers. Cervical and ovarian cancer survivors have distinct information needs related to their illness. Examples are reproductive and menstrual history, sexual activities, and marital statuses (Stead et al., 2003). As noted earlier, the unique characteristics associated with each type of cancer determine the kinds of information individuals with that type of cancer require. As such, generic information about promoting physical health (e.g., active lifestyle and diet regime) is insufficient for fulfilling the information needs of cervical and ovarian cancer survivors.

In conclusion, there is a crucial need to bridge this prevailing gap by developing digital health technologies specifically designed to combat and effectively address the misinformation cancer survivors encounter during their search for accurate and reliable information. Such a tool will provide cancer survivors with the means to navigate the overwhelming landscape of information while counteracting the negative impacts of misinformation. Moreover, such aids are needed to address the challenges of ovarian and cervical cancer populations. By doing so, we can gain valuable insights that will contribute to developing tailored interventions to meet the informational needs of cancer survivors.

3. Conversational Agents

3.1 Definition and Potential of Conversational Agents for Health in General

To address the need mentioned above for technological aids, we propose the implementation of conversational agents (CAs). CAs encompass physical and virtual autonomous entities with reactive and proactive capabilities (Holz et al., 2009). In the literature, CAs are referred to using various terms, such as chatbots, voice assistants, voice bots, or intelligent personal assistants. CAs are not mere tools but software programs that serve as facilitators or assistants (Lieberman, 1997), effectively simulating human conversation and interacting with users employing natural language (Bittner et al., 2019).

Using natural language technologies, CAs engage users in text-based information-seeking and task-oriented dialogues across various applications. Their design allows for the acceptance and generation of natural language input and output, facilitating social conversations (De Keyser et al., 2019). Similarly, the language model ChatGPT, created by OpenAI, can assist individuals and communities in making educated decisions about their health since it can generate human-like writing based on massive volumes of data (Panch et al., 2019).

Not only have CAs, including chatbots, been proven effective in disseminating health-related information during global pandemics like COVID-19, but their ability to combat general misinformation has been documented (Almalki et al., 2020). This highlights their significant potential in fulfilling cancer survivors' information needs while tackling various forms of misinformation. By leveraging CAs, users can access content and services more personalized and intimately than traditional non-conversational self-service technologies (Sheehan et al., 2020).

CAs have demonstrated efficacy in inducing behavior changes and addressing physical fitness and mental health, which are common challenges among cancer survivors. Patients who followed CAs' dialogues reported a significant increase in positive attitudes toward physical health (Schulman et al. 2009). Similarly, CAs like Woebot, Wysa, and Joy have the potential to provide valuable assistance to individuals grappling with mental health challenges (Kretzschmar et al., 2019). Moreover, studies evaluating the efficacy of CAs in healthcare have shown a significant reduction in depression symptoms (Laranjo et al., 2018).

In summary, prior studies attest to the CAs' potential to guide cancer survivors in navigating the large pool of credible and non—credible information. Their personalized and intimate nature and effectiveness in driving behavior changes and supporting mental health make them valuable tools to help cancer survivors.

3.2 Users' Distrust in Generative AI Systems

The potential mentioned above of CAs has its share of challenges. These challenges include patient safety, integration with other technologies, and medico-legal and ethical concerns (Bickmore et al., 2018; Van Pinxteren et al., 2020). In addition, technical difficulties have been identified. Li et al. (2018) highlighted errors generated by the system when attempting to recognize natural speech and create responses to users' queries. Furthermore, the unintuitive interactions leave users to rely on trial and error to understand the functions of CAs (Bickmore et al., 2018).

More importantly, users' distrust of AI-based systems, such as CAs, often leads to the refusal to accept AI-based systems. Reasons for distrust of AI systems included fear of vulnerability (McLean et al., 2020), limited knowledge (Nadarzynski et al., 2019), resistance to technological change (Maier et al., 2019), and the lack of humanness (Van Pinxteren et al., 2020; Nadarzynski et al., 2019). CAs have yet to fully tap into the potential of human-derived communicative behaviors (Van Pinxteren et al., 2020).

3.3 Design Science Approach – Knowledge Graph to Model Cancer Survivors' Understanding of Their Illnesses

To address the above issues, we employ some of the principles of the design science approach (DSR) devised to solve and make sense of a given problem (Teixeira et al., 2017). The DSR approach consists of (i) identifying and defining the problem to solve, (ii) building a model of an artifact (an explicit representation of a chosen solution to the problem of interest), and (iii) demonstrating whether the artifact can solve the problem using the model (Teixeira et al., 2017).

According to the DSR approach, we identified the problem to solve (i.e., an intuitive and trustworthy CA to assist ovarian/cervical cancer survivors' information search) as shown in the section *Literature Review* and the section *Users' Distrust in Generative AI Systems*. According to the DSR approach, our second step is to build a model. To do so, we chose to build a knowledge graph using an unsupervised learning approach from a text corpus.

A knowledge graph represents semantic relationships through triplets extracted from a comprehensive dataset that captures real-world facts (Peng et al., 2023). By visualizing these triplets as a graph, with entities as nodes and relationships as edges, we create what is known as a knowledge graph.

We will use a knowledge graph to represent how cancer patients make sense of their illness, treatment options, side effects and recurrences, and access barriers (e.g., availability and costs of those treatments). Our justification for creating this mental model using a knowledge graph is three-fold. First, there is likely a gap between an ordinary cancer survivor's understanding of cancer and the knowledge in medical literature. This gap is likely because patients, without proper medical training, cannot understand all the content documented in medical literature. Simultaneously, medical literature may not contain the information patients need the most (e.g., access barriers, such as costs and availability of treatments). Thus, it is necessary to construct a representation of the cancer treatment trajectory from a patient's perspective.

Second, because this knowledge graph visually represents how concepts are connected in patients' mental models, this graph can be translated into a structure by which one can design intuitive and relatable user interactions. In other words, this knowledge graph will show what answers would be most appropriate and intuitive to a particular question by identifying how that question is positioned in the mental map of a patient.

Third, since knowledge graphs visualize relationships among concepts in patients' understanding of cancer, this visualization helps us identify why and how one falls prey to misinformation. This explicit, documented visual presentation of cancer patients' mental models will facilitate the effort to debunk misbeliefs such that an intervention can explicitly address the link between misconceptions.

Empirical evidence shows the potential of a similar approach in this context, albeit not knowledge graphs. Lee et al. (2022) harnessed the power of social media and language analysis to fulfill the unmet needs of ovarian cancer patients and caregivers. Leveraging initial postings from online health communities, they developed an automated model capable of classifying the diverse needs expressed by ovarian cancer patients and caregivers.

Knowledge graphs have also been successfully employed in numerous studies on misinformation. For instance, Chen et al. (2022) conducted extensive experiments using real-world datasets on diabetes and cancer, demonstrating the effectiveness of biomedical knowledge graph guidance. Furthermore, Koloski et al. (2022) leveraged knowledge graphs to enhance text representations of COVID-19 misinformation. Additionally, Mayank et al. (2022) proposed a combination of text encoding and knowledge graphs to demonstrate how the semantics of health-related misinformation are connected.

3.4. Methodology – Data Collection

Given the empirical support to the promise of knowledge graphs to solve our identified problem, we collected data about cancer survivors' understanding of their illness. Specifically, we obtained extensive coverage of cancer (mis)information that survivors seek, discuss, and are exposed to. Our approach involved gathering credible data from (i) credible sources such as reputable news and health websites and (ii) expert insights from health practitioners specializing in oncology. Various reputable sources have been tapped into, including *The Wall Street Journal*, *The New York Times*, *USA Today*, and *The Washington Post*. Additionally, we have collected content from online communities like *Cancer Survivors Network*, *Inspire* (Ovarian Cancer), *Smart Patients*, *Cervivor*, *Women's Health*, and *Cancer Today*.

To collect less reliable, unverified data, we engaged social media platforms like Facebook, Twitter, and Reddit. References to these groups allowed us to find the most common topics discussed among cancer patients, including unverified sources. These groups are formed organically among individual patients; thus, their group pages allowed us to collect their conversations in the most natural, uninterrupted ways. This way, we ensured that we collected both credible and non-credible information available online.

Our next step was manually labeling the collected content into two categories (credible vs. non-credible information). The criterion for discerning between the two was the source's credibility—whether the source of information under discussion was verified or not. As our manual labeling was complete, we began to engage healthcare providers (i.e., oncologists) to verify the accuracy of our labels. These healthcare practitioners will provide insights based on their professional experience, aiding in validating the accuracy of the information gathered. This iterative process ensures that the information used for training our model is grounded in expert judgment and reliable medical expertise. The collaboration with oncologists has already commenced, with interview sessions initiated to harness their domain expertise. Our plans engage a diverse array of healthcare providers who interact with patients, including nurses, nurse practitioners, nurse educators, and physicians' assistants in reputable healthcare institutions.

It is worth noting that all participants' ethical considerations and privacy have been paramount throughout this endeavor. An Institutional Review Board (IRB) approved our data collection before its commencement. This demonstrates our commitment to upholding the highest standards of ethical research practices.

3.5 Development of Knowledge Graphs

We will build our knowledge graph following the sequences below.

Data cleaning. Data cleaning is needed as a preprocessing step to alleviate such data quality issues by transforming the noisy or irregular word tokens in the text into their standardized forms. The data cleaning step can be implemented using a combination of automated text correction tools (Almeida et al., 2017) such as NeuSpell (Almalki et al., 2020) and hand-crafted filtering or transformation rules.

Information extraction. A knowledge graph can be formed by extracting a collection of triplets (s, r, o), where s and o are the entities/concepts described in a sentence, while r is a predicate that captures the relationship between the pair of entities/concepts. Given a text corpus, the extraction of the triplets typically involves the following subtasks (Wani et al., 2021): (1) Sentence segmentation (where the input text is split into sentences using existing Python NLP libraries such as nltk and spacy), (2) Entity extraction (which identifies the set of subjects s and objects o by applying word tokenization followed by part-of-speech tagging to each sentence in the text corpus) and (3) Relation extraction (which identifies the predicate r associated with each pair of entities based on the verb phrase obtained from the part-of-speech tags).

Graph refinement (Paulheim, 2017). The initial set of triplets found may be further processed to ensure the correctness and completeness of information. First, the irrelevant entities and spurious relations harvested from the unstructured text must be discarded. Furthermore, new relations must be added using link prediction methods to complete the graph (Kazemi and Poole, 2018).

Graph visualization. Finally, the triplets found in the previous step are provided to a visualization library such as Pyvis and networks to display the set of entities as nodes of the knowledge graph and relations as edges.

3.6 Iterative Developments of the CA's Scripts with User Testing

Once a knowledge graph is built, we will create CAs' scripts for conversing with users. A visualized example is shown in **Figure 1**. We will conduct multiple focus group interviews to develop the scripts with users' input iteratively.

Once the scripts are finalized, we will implement a rule-based dialog system to guide CAs' responses to users' inputs. The CA will be implemented in python, which has an extensive suite of open-source libraries such as ChatterBot for designing flexible chatbots with a machine learning backend. An initial draft of the look and feel of our CAs is presented in **Fig. 1**.

Once the CA development is complete, we will conduct controlled laboratory experiments to ensure that the CAs are perceived as relatable, trustworthy, and acceptable and whether they fulfill cancer survivors' information needs. This evaluation of CA is the final step of the DSR approach in which developers assess the efficacy of the proposed solution's capacity to address the problem of interest (Teixeira et al., 2017).

For all these user evaluations, we will recruit cancer survivors through various cancer survivor networks, such as *Cancer Survivors Network* and *Cervive*. Their participation will be voluntary, and we will guarantee the anonymity and privacy of the participants and the secure storage of their personal information.

Hello! How can I support you today? 😊

Hey chatbot! 🌸 I read an article on social media that said cannabis can help with my recovery after cervical cancer surgery. Is that true? 🌿

Hi there! 🤗 Thanks for reaching out. I hope your recovery is going well. According to reliable sources, cannabis does not aid in recovery after cervical cancer surgery. (More details will be added later.)

Oh, I see. I'm experiencing a lot of pain after my surgery. Since cannabis doesn't help, can I take any over-the-counter medicine for relief?

I'm not a doctor but opting for over-the-counter medicine may be a better choice than cannabis 🤗. However, it's crucial to consult your healthcare provider for pain management options. They can recommend suitable over-the-counter painkillers or prescribe medication tailored to your recovery. They'll ensure you receive the most appropriate and safe relief. 🩺⭐

Thank you for your guidance.

You're welcome! I'm always here to support you. Take care and feel better soon!

Figure 1. A conceptual example of a CA

4. Discussion

4.1 Contributions

Our study addresses cancer survivors' current information needs and their susceptibility to misinformation while seeking information for managing their cancer (Jensen et al., 2020). In so doing, we recognize the potential of a generative AI aid, specifically CAs, to fill this need. CAs can be implemented in an existing mobile application (e.g., a patient portal) as a part of a cancer self-management tool. Integrated into such an application, CAs will first provide credible information necessary to manage cancer and help redirect survivors from misinformation.

We acknowledge that CAs face challenges, notably rigid interactions, unintuitive sequences of dialogues between CAs and users, and unrelatable responses of CAs, which all result in users' distrust of them.

To alleviate these challenges, we followed the DSR approach by building a knowledge graph to represent the problem and developing and evaluating the final system with end users. This way, CAs will be able to provide sufficient answers that resonate with users' understanding of cancers in an intuitive and relatable manner.

Our work could contribute to advancing generalizable scientific knowledge and theories. Our study presents a DSR approach to alleviate users' distrust in generative AI systems. Specifically, we have argued that using knowledge graphs helps build cancer survivors' mental models of understanding their illnesses and treatments, thus informing the implementation of CAs. In doing so, our study will eventually advance the principles or frameworks for developing self-management tools for chronic illness.

4.2 Limitations and Suggestions for Future Research

Although our study makes valuable contributions, it is essential to acknowledge its limitations. One notable limitation is the absence of empirical research that elucidates insights into the diverse needs of cancer survivors and the specific challenges they encounter. Content analyses of existing cancer misinformation online can offer a deeper understanding of the severity of cancer misinformation and how it interferes with survivors' efforts for recovery. Furthermore, while we have demonstrated the potential of knowledge graphs, empirical data would provide a more robust foundation for their effective integration into CAs.

In light of these limitations, we recommend that future studies incorporate content analyses and empirical research to explore the multifaceted needs of cancer survivors and gain a deeper understanding of their challenges. This empirical approach would enable researchers to gather more detailed and nuanced information. Additionally, future investigators should prioritize collecting empirical data to inform the development and optimization of knowledge graphs within CAs, leading to more sophisticated and impactful implementations. By addressing these limitations and conducting empirical studies, researchers can advance the field and contribute valuable insights that inform the design and implementation of AI-based tools for supporting the self-management of chronic illnesses.

5. References

[1] Almalki, M., & Azeez, F. (2020). Health chatbots for fighting COVID-19: a scoping review. Acta Informatica Medica, 28(4), 241. Peng, C., Xia, F., Naseriparsa, M., & Osborne, F. (2023). Knowledge graphs: opportunities and challenges. *Artificial Intelligence Review*, 1-32.

[2] Almeida, H., Briand, A., & Meurs, M. J. (2017, September). Detecting Early Risk of Depression from Social Media User-generated Content. In *CLEF* (working notes).

[3] Beeken, R. J., Williams, K., Wardle, J., & Croker, H. (2016). "What about diet?" A qualitative study of cancer survivors' views on diet and cancer and their sources of information. *European Journal of Cancer Care*, 25(5), 774–783.

[4] Bickmore, T. W., Trinh, H., Olafsson, S., O'Leary, T. K., Asadi, R., Rickles, N. M., & Cruz, R. (2018). Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. *Journal of medical Internet research*, 20(9), e11510

[5] Bittner, E. A., Küstermann, G. C., & Tratzky, C. (2019). The facilitator is a bot: Towards a conversational agent for facilitating idea elaboration on idea platforms.

[6] Blount, D. S., McDonough, D. J., & Gao, Z. (2021). Effect of wearable technology-based physical activity interventions on breast cancer survivors' physiological, cognitive, and emotional outcomes: a systematic review. *Journal of Clinical Medicine*, 10(9), 2015.

[7] Budenz, A., Sleight, A. G., & Klein, W. M. (2022). A qualitative study of online information-seeking preferences among cancer survivors. *Journal of Cancer Survivorship*, 1-12.

[8] Cancer Net, (2022). https://www.cancer.net/survivorship/what-cancer-survivorship (last accessed on 09/04/23)

[9] Comes, S., Schatsky, D., & Chauhan, R. (2021, November 16). Conversational AI. Deloitte Insights. https://www2.deloitte.com/us/en/insights/focus/signals-for-strategists/the-future-of-conversational-ai.html

[10] Chen, C., Wang, H., Shapiro, M., Xiao, Y., Wang, F., & Shu, K. (2022). Combating Health Misinformation in social media: Characterization, Detection, Intervention, and Open Issues. arXiv preprint arXiv:2211.05289.

[11] Chua, G. P., Ng, Q. S., Tan, H. K., & Ong, W. S. (2021). Cancer survivors: What are their information-seeking behaviors?. *Journal of Cancer Education*, 36, 1237-1247.

[12] De Keyser, A., Köcher, S., Alkire, L., Verbeeck, C., & Kandampully, J. (2019). Frontline service technology infusion: conceptual archetypes and future research directions. Journal of Service Management, 30(1), 156-183.

[13] Department of Health-Quality Report (2012). https://www.ahrq.gov/sites/default/files/publications/files/2012nhqr.pdf (last accessed on 09/04/23)

[14] Eraslan, P., & Tufan, G. (2022). Cancer information overload may be a crucial determinant of adjuvant

aromatase inhibitor adherence. *European Review for Medical & Pharmacological Sciences*, 26(19).

[15] Ganggayah, M. D., Dhillon, S. K., Islam, T., Kalhor, F., Chiang, T. C., Kalafi, E. Y., & Taib, N. A. (2021). An artificial intelligence-enabled pipeline for medical domain: Malaysian breast cancer survivorship cohort as a case study. *Diagnostics*, 11(8), 1492.

[16] Gentile, D., Markham, M. J., & Eaton, T. (2018). Patients with cancer and social media: harness benefits, avoid drawbacks. *Journal of oncology practice*, 14(12), 731–736.

[17] Holmes, M. M., Bishop, F. L., & Calman, L. (2017). “I just googled and read everything”: Exploring breast cancer survivors’ use of the internet to find information on complementary medicine. *Complementary therapies in medicine*, pp. 33, 78–84.

[18] Holz, T., Dragone, M., & O’Hare, G. M. (2009). Where robots and virtual agents meet: a survey of social interaction research across milgram’s reality-virtuality continuum. International Journal of Social Robotics, 1, 83-93.Chicago.

[19] Hu, R. Y., Wang, J. Y., Chen, W. L., Zhao, J., Shao, C. H., Wang, J. W., ... & Yu, J. M. (2021). Stress, coping strategies and expectations among breast cancer survivors in China: a qualitative study. *BMC Psychology*, 9, 1–12.

[20] Jensen, J. D., Pokharel, M., Carcioppolo, N., Upshaw, S., John, K. K., & Katz, R. A. (2020). Cancer information overload: discriminant validity and relationship to safe sun behaviors. *Patient education and counseling*, 103(2), 309–314.

[21] Johnson, S. B., Parsons, M., Dorff, T., Moran, M. S., Ward, J. H., Cohen, S. A., ... & Fagerlin, A. (2022). Cancer misinformation and harmful information on Facebook and other social media: a brief report. *Journal of the National Cancer Institute*, 114(7), 1036-1039.

[22] Kapoor, A., Nambisan, P., & Baker, E. (2020). Mobile applications for breast cancer survivorship and self-management: A systematic review. *Health informatics* journal, 26(4), 2892–2905.

[23] Kazemi, S. M., & Poole, D. (2018). Simple embedding for link prediction in knowledge graphs. Advances in neural information processing systems, 31.

[24] Kent, E. E., Arora, N. K., Rowland, J. H., Bellizzi, K. M., Forsythe, L. P., Hamilton, A. S. & Aziz, N. M. (2012). Health information needs and health-related quality of life in a diverse population of long-term cancer survivors. *Patient education and counseling*, *89*(2), 345–352.

[25] Koloski, B., Perdih, T. S., Robnik-Šikonja, M., Pollak, S., & Škrlj, B. (2022). Knowledge graph informed fake news classification via heterogeneous representation ensembles. *Neurocomputing*, *496*, 208-226.

[26] Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., Singh, I., & NeurOx Young People’s Advisory Group. (2019). Can your phone be your therapist? Young people’s ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomedical informatics insights*, *11*, 1178222619829083.

[27] Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., ... & Coiera, E. (2018). Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, *25*(9), 1248-1258.

[28] Lee, Y. J., Jang, H., Campbell, G., Carenini, G., Thomas, T., & Donovan, H. (2022). Identifying language features associated with needs of ovarian cancer patients and caregivers using social media. *Cancer Nursing*, 45(3), E639-E645.

[29] Li, T. J. J., Labutov, I., Myers, B. A., Azaria, A., Rudnicky, A. I., & Mitchell, T. M. (2018). An end user development approach for failure handling in goal-oriented conversational agents. *Studies in Conversational UX Design*.

[30] Lieberman, H. (1997, March). Autonomous interface agents. In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems 67-74.

[31] Lisy, K., Langdon, L., Piper, A., & Jefford, M. (2019). Identifying the most prevalent unmet needs of cancer survivors in Australia: a systematic review. Asia-Pacific Journal of Clinical Oncology, 15(5), e68-e78.

[32] Lynch, B. M., Nguyen, N. H., Moore, M. M., Reeves, M. M., Rosenberg, D. E., Boyle, T., Vallance, J., Milton, S., Friedenreich, C., & English, D. R. (2019). A randomized controlled trial of a wearable technology-based intervention for increasing moderate to vigorous physical activity and reducing sedentary behavior in breast cancer survivors: The ACTIVATE trial. *Cancer*. 125(16). 2846-2855.

[33] Maier, S. B., Jussupow, E., & Heinzl, A. (2019). Good, bad, or both? Measurement of physician’s ambivalent attitudes towards AI.

[34] Marchetti, M. A., Sar-Graycar, L., Dusza, S. W., Nanda, J. K., Kurtansky, N., Rotemberg, V. M., & Hay, J. L. (2022). Prevalence and Age-Related Patterns in Health Information–Seeking Behaviors and Technology Use Among Skin Cancer Survivors: Survey Study. *JMIR* Dermatology, *5*(2), e36256.

[35] Marzorati, C., Riva, S., & Pravettoni, G. (2017). Who is a cancer survivor? A systematic review of published definitions. *Journal of Cancer Education*, *32*, 228-237.

[36] Mayank, M., Sharma, S., & Sharma, R. (2022, November). DEAP-FAKED: Knowledge graph based approach for fake news detection. In *2022 IEEE/ACM* International *Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 47–51). IEEE.

[37] McLean, G., Osei-Frimpong, K., Wilson, A., & Pitardi, V. (2020). How live chat assistants drive travel consumers’ attitudes, trust and purchase intentions: the role of human touch. *International Journal of Contemporary Hospitality Management*, 32(5), 1795-1812.

[38] Meneses, K. (2010). Development of the Fertility and Cancer Project: An Internet approach to help young cancer survivors. *Number 2/March 2010*, *37*(2), 191–197.

[39] Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial intelligence (AI)-led

chatbot services in healthcare: A mixed-methods study. *Digital health*, *5*, 2055207619871808.

[40] Nguyen, S. K. A., & Ingledew, P. A. (2013). Tangled in the breast cancer web: an evaluation of the usage of web-based information resources by breast cancer patients. *Journal of Cancer Education*, pp. *28*, 662–668.

[41] Panch, T., Mattie, H., & Atun, R. (2019). Artificial intelligence and algorithmic bias: implications for health systems. Journal of global health, 9(2).

[42] Paulheim, H. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic web, 8(3), 489-508.

[43] Peng, C., Xia, F., Naseriparsa, M., & Osborne, F. (2023). Knowledge graphs: opportunities and challenges. *Artificial Intelligence Review*, pp. 1–32.

[44] Perrault, E. K., Hildenbrand, G. M., McCullock, S. P., Schmitz, K. J., & Lambert, N. J. (2020). Online information seeking behaviors of breast cancer patients before and after diagnosis: from website discovery to improving website information. *Cancer treatment and research* communications, *23*, 100176.

[45] Peterson, E. & Iyengar, S. (2022). Partisan reasoning in a high stakes environment: Assessing partisan informational gaps on COVID-19. *Harvard Kenedy School Misinformation Review*, 3(2), https://doi.org/10.37016/mr-2020-96

[46] Rupert, D. J., Squiers, L. B., Renaud, J. M., Whitehead, N. S., Osborn, R. J., Furberg, R. D., Squire, C. M., & Tzeng, J. P. (2013). Communicating risk of hereditary breast and ovarian cancer with an interactive decision support tool. *Patient Education and Counseling*, 92(2). 188-196.

[47] Rheu, M., Shin, J. Y., Peng, W., & Huh-Yoo, J. (2021). Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human–Computer Interaction*, 37(1), 81–96.

[48] Rutten, L. J. F., Arora, N. K., Bakos, A. D., Aziz, N., & Rowland, J. (2005). Information needs and sources of information among cancer patients: a systematic review of research (1980–2003). *Patient education and* counseling, *57*(3), 250-261.

[49] Schook, R. M., Linssen, C., Schramel, F. M., Festen, J., Lammers, E., Smit, E. F., ... & Westerman,

[50] Stead, M. L., Brown, J. M., Fallowfield, L., & Selby, P. (2003). Lack of communication between healthcare professionals and women with ovarian cancer about sexual issues. British journal of cancer, 88(5), 666-671.

[51] Svoboda, E. (2020). RNA fragments signal ovarian cancer. Nature, 579(7800), S20-S21.

[52] Schulman, D., & Bickmore, T. (2009, April). Persuading users through counseling dialogue with a conversational agent. In *Proceedings of the 4th international conference on persuasive technology* (pp. 1-8).

[53] Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research*, *115*, 14-24.

[54] Teixeira, J., Patrício, L., Huang, K.-H., Fisk, R. P., Nóbrega, L., & Constantine, L. (2017). The MINDS Method: Integrating Management and Interaction Design Perspectives for Service Design. Journal of Service Research, 20(3), 240–258.

[55] Torrente, M., Sousa, P. A., Hernández, R., Blanco, M., Calvo, V., Collazo, A., ... & Provencio, M. (2022). An Artificial Intelligence-Based Tool for Data Analysis and Prognosis in Cancer Patients: Results from the Clarify Study. *Cancers*, *14*(16), 4041.

[56] Van Pinxteren, M. M., Pluymaekers, M., & Lemmink, J. G. (2020). Human-like communication in conversational agents: a literature review and research agenda. *Journal of Service Management*, *31*(2), 203-225.

[57] Van Booven, D. J., Kuchakulla, M., Pai, R., Frech, F. S., Ramasahayam, R., Reddy, P., ... & Arora, H. (2021). A systematic review of artificial intelligence in prostate cancer. Research and reports in urology, 31-39.

[58] Wang, T., Molassiotis, A., Chung, B. P. M., & Tan, J. Y. (2018). Unmet care needs of advanced cancer patients and their informal caregivers: a systematic review. *BMC palliative care*, 17(1), 1–29.

[59] Wani, I. Joshi, S. Khandve, V. Wagh and R. Joshi, "Evaluating deep learning approaches for covid19 fake news detection" in *International Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation*, Springer, pp. 153–163, 2021.

[60] Warner, E. L., Basen-Engquist, K. M., Badger, T. A., Crane, T. E., & Raber-Ramsey, M. (2022). The Online Cancer Nutrition Misinformation: A Framework of behavior change based on exposure to cancer nutrition misinformation. *Cance*r, 128(13), 2540–2548.

[61] Warraich, H. (2018, December 17). Dr. Google is a liar. The New York Times. https://www.nytimes.com/2018/12/16/opinion/statin-side-effects-cancer.html (last accessed on 09/04/23)

[62] Wilner, T., & Holton, A. (2020). Breast cancer prevention and treatment: Misinformation on Pinterest, 2018. *American Journal of public health*, *110*(S3), S300-S304.

[63] Xiao, Z., Liao, Q. V., Zhou, M., Grandison, T., & Li, Y. (2023, March). Powering an AI Chatbot with Expert Sourcing to Support Credible Health Information Access. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 2-18).

[64] Zenone, M., Snyder, J., & Caulfield, T. (2020). Crowdfunding cannabidiol (CBD) for cancer: hype and misinformation on GoFundMe. *American Journal of Public Health*, 110(S3), S294-S299.