

DESIGN AND EVALUATION OF A CONVERSATIONAL AGENT FOR  
SUPPORTING DOMESTIC VIOLENCE SURVIVORS

by

Abdulrahman Aldkheel

A dissertation submitted to the faculty of  
The University of North Carolina at Charlotte  
in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy in  
Computing and Information Systems

Charlotte

2024

Approved by:

---

Dr. Lina Zhou

---

Dr. Dongsong Zhang

---

Dr. Weichao Wang

---

Dr. Cheryl Brown

© 2024

Abdulrahman Aldkheel

ALL RIGHTS RESERVED

## ABSTRACT

ABDULRAHMAN ALDKHEEL. Design and Evaluation of A Conversational Agent for Supporting Domestic Violence Survivors. (Under the direction of DR. LINA ZHOU)

Domestic violence (DV) is widely recognized as a significant problem with detrimental impacts on the mental, physical, and socio-economic well-being of individuals, families, and the broader community. Despite various resources designated to support survivors, they may not be easily accessible or readily available. More importantly, DV survivors often hesitate to disclose their experiences to others and may even refrain from seeking assistance because of social, emotional, privacy, and cultural concerns. As providing immediate support in response to DV is crucial for the physical and psychological well-being of survivors, they need timely assistance from non-judgmental first responders.

Technological advancements, particularly in automated Conversational Agent (CA), are progressing rapidly. CAs are gaining attention as a promising tool for providing counseling and support, addressing the above-mentioned challenges faced by survivors using traditional support resources. The primary objective of this dissertation research is to design, develop, and evaluate a CA-based solution that assists DV survivors in accessing services more easily, increasing their awareness, and enhancing the available support. To this end, we first identify the meta requirements and design principles of CA for survivors by interviewing DV professionals, then design and develop SafeHaven, a CA-based prototype for supporting DV survivors by following the design principles and meeting the specified meta requirements, and finally evaluate the effectiveness and perception of SafeHaven by conducting user experiments.

Our findings suggest that CAs should empathize with survivors' experiences and provide them with meaningful informational, tangible, and emotional support, prioritizing their safety and

maintaining transparent, private, and trustworthy communication. We evaluated SafeHaven with DV survivors, their friends and family, and DV professionals, in terms of emotional, informational, and instrumental support it provided, among other measures. In addition, we compared the CA with a traditional search engine and emerging ChatGPT. The experiment results show that SafeHaven outperformed both search engine and ChatGPT in offering emotional, informational, and instrumental support, high information quality, and user trust.

This dissertation research identifies the meta-requirements and design principles for designing a CA for DV survivors. This is the first study that assesses the effectiveness of CAs in assisting individuals who experience DV and/or provide support to DV survivors, and providing tailored, context-sensitive support to them. In a broader sense, the outcomes of this research will be instrumental in guiding the future development of CA-based support systems for DV survivors. This dissertation highlights the transformative potential of CAs for DV survivors, as well as significant implications for DV organizations, support groups, and CA developers, proposing novel avenues and strategies for enhancing the anonymity and accessibility of the support for survivors.

## ACKNOWLEDGMENT

I would like to express my profound gratitude and thanks to my advisor, Dr. Lina Zhou, for her invaluable guidance and support throughout my PhD journey. I have participated in several research projects under Dr. Zhou's supervision. Such participation contributed immensely to my academic growth. Under her guidance, I have acquired a wealth of knowledge and critical thinking skills. During my research, Dr. Zhou provided detailed guidance and was readily available whenever I needed assistance. Under her supervision, I have cultivated a strong sense of commitment and dedication to my research. Most of all, I am truly grateful for her patience throughout my research and projects, demonstrating the importance of staying thoughtful and meticulous in my academic endeavors.

My appreciation also extends to the members of my dissertation committee, Dr. Dongsong Zhang, Dr. Chao Wang, and Dr. Cheryl Brown, for their insightful feedback on my dissertation research, which has significantly improved the quality of my dissertation.

I also would like to thank and express my gratitude to my beloved family, including my parents, Dr. Mohammed Aldkheel and Dr. Kifah Almazroua, my brothers Abdulaziz and Abdullah, and my wife Lama Almuhaideb and our newborn Abdulelah for their love, encouragement, and continuous support throughout my academic journey. I am deeply grateful for everything they have done to assist me in the successful attainment of my PhD degree.

In addition, I would like to extend my thanks to my friends and colleagues in the KAIM research group: Marran Aldossari, Hamad Alsaleh, Zhe Fu, Jaewan Lim, Zhihui Liu, and Kanlun Wang whose support and encouragement contributed greatly to my research's success.

Lastly, I would like to express my heartfelt appreciation to the Kingdom of Saudi Arabia and King Saud University for providing me with the opportunity to pursue my doctoral degree and for their consistent support during my academic studies.

## TABLE OF CONTENTS

LIST OF TABLES .....	X
LIST OF FIGURES .....	XI
1 INTRODUCTION .....	1
1.1 Motivations .....	1
1.2 Scope of Research.....	4
1.3 Structure of Dissertation .....	8
2 BACKGROUND AND RELATED WORK.....	9
2.1 Background.....	9
2.2 Non-Technology Mediated DV Intervention.....	12
2.2.1 DV Survivors' Self-Disclosure .....	12
2.2.2 DV Intervention.....	14
2.3 Technology-Mediated DV Intervention.....	15
2.3.1 DV Survivor's Self-Disclosure .....	15
2.3.2 DV Intervention.....	18
2.3.3 Technology Support .....	21
2.4 Conversational Agents .....	32
2.4.1 Fundamentals .....	32
2.4.2 Trust and Self-disclosure.....	39
2.4.3 Perceived Anonymity .....	41
2.4.4 DV Intervention.....	42
3 REQUIREMENTS AND ROLES OF CA-BASED DV INTERVENTION.....	51
3.1 Introduction.....	51
3.2 Background and Related Work.....	53
3.2.1 DV Survivors.....	54
3.2.2 DV Intervention: A Technology-Based Solution.....	55
3.2.3 Design Principles of CAs .....	58
3.3 Research Methodology .....	60
3.3.1 Interviews .....	63

3.3.2	Participants .....	63
3.3.3	Procedure.....	64
3.3.4	Data Analysis .....	65
3.4	Results.....	65
3.4.1	Survivors' Needs and Goals .....	66
3.4.2	CA Design Guideline .....	77
3.4.3	The Use of CAs: Opportunities, Challenges and Design Suggestions.....	109
4	SAFEHAVEN: DESIGN AND IMPLEMENTATION .....	123
4.1	Introduction.....	123
4.2	SafeHaven – A Dialog Flowchart and Key Features and Functions .....	124
4.3	Detailed Design.....	128
4.3.1	Greeting and Assuring Safety .....	129
4.3.2	Screening and Abuse Identification .....	131
4.3.3	Safety Plan Development .....	133
4.3.4	Survivors' Stories.....	135
4.3.5	Referral Resources .....	136
4.4	Implementation .....	181
4.5	Hypotheses Development .....	183
5	EVALUATION OF SAFEHAVEN .....	192
5.1	Methods.....	192
5.1.1	Task Design.....	192
5.1.2	Research Instruments .....	195
5.1.3	Procedure.....	199
5.1.4	Cognitive Walkthrough.....	202
5.1.5	Participants Recruitment .....	203
5.1.6	Ethical Consideration .....	209
5.1.7	Data Analysis .....	210
5.2	Quantitative Analysis Results .....	210
5.2.1	SafeHaven Vs. Search Engine.....	212
5.2.2	SafeHaven vs. ChatGPT.....	220
5.3	Qualitative Analysis Results .....	229
5.3.1	Cognitive Walkthrough Results .....	229
5.3.2	Interviews Results .....	254



6	DISCUSSION AND CONCLUSION .....	311
6.1	Main Findings .....	311
6.1.1	Emotional Support.....	313
6.1.2	Usability .....	314
6.1.3	Anonymity and Trustworthiness .....	315
6.2	Research Contributions .....	316
6.3	Implications.....	318
6.4	Limitations and Future Work.....	320
6.5	Conclusion .....	322
	REFERENCES .....	323
	APPENDICES .....	347
	APPENDIX A: INTERVIEW PROTOCOL FOR STUDY 1 .....	347
	APPENDIX B: SURVEY INSTRUMENTS .....	352
	APPENDIX C: SCREENING TOOL FOR DV .....	356
	APPENDIX D: INTERVIEW PROTOCOL FOR STUDY 2 .....	357
	APPENDIX E: COGNITIVE WALKTHROUGH: AN ILLUSTRATION .....	359

## LIST OF TABLES

Table 2.1: A review of studies on the technology-based Interventions for DV .....	27
Table 2.2: An overview of the characteristics of CAs used in DV support.....	48
Table 2.3: A list of functions of reviewed CAs .....	50
Table 3.1: CA design principles informing the IS literature.....	60
Table 3.2: An overview of the CAs' meta requirements .....	78
Table 3.3: Design principles based on the identified meta requirements .....	108
Table 5.1: Descriptive statistics from demographic and general survey responses (N = 36) .....	205
Table 5.2: Descriptive statistics and statistical comparisons of the effectiveness between SafeHaven and search engine .....	211
Table 5.3: Descriptive statistics and statistical comparisons of the effectiveness between SafeHaven and ChatGPT .....	211
Table 5.4: Descriptive statistics and statistical comparisons of user perceptions of information quality between SafeHaven and search engine .....	216
Table 5.5: Descriptive statistics and statistical comparisons of user perceptions of information quality between SafeHaven and ChatGPT .....	224
Table 5.6: Challenges, Tasks, and Recommendations for Improving User Experience with SafeHaven .....	231

## LIST OF FIGURES

Figure 2.1: The architecture of a CA-based dialog system.....	34
Figure 2.2: Finite state-based system.....	35
Figure 2.3: Frame-based system .....	35
Figure 2.4: An agent-based system.....	35
Figure 2.5: The ChatGPT training process [167].....	37
Figure 3.1: The design research cycle [234], [238] .....	62
Figure 3.2: Summary of interviews outcomes .....	65
Figure 3.3: Conversation guideline- Avoiding reiterating the question .....	85
Figure 3.4: Conversation guideline- Using closed-ended and open-ended questions .....	86
Figure 3.5: CA feature design- offering emotional support.....	93
Figure 3.6: CA feature design- providing local resources .....	97
Figure 3.7: Conversation recommendations: Anthropomorphizing the CA.....	121
Figure 3.8: Major phases of the CA's conversation flow.....	121
Figure 4.1: Flowchart of a dialogue session .....	125
Figure 4.2: Welcoming message.....	130
Figure 4.3: Assuring safety .....	130
Figure 4.4: Abuse screening .....	132
Figure 4.5: Abuse identification .....	132
Figure 4.6: General safety tips.....	133
Figure 4.7: Developing a safety plan .....	134
Figure 4.8: Personalized safety plan .....	135
Figure 4.9: An example of the survivor's story.....	136
Figure 4.10: Resources module.....	137
Figure 4.11: Find a shelter .....	139
Figure 4.12: List of nearby shelters (Domesticshelters.org).....	139
Figure 4.13: Find a family therapist.....	141
Figure 4.14: List of family therapists (aamft.org).....	141
Figure 4.15: Find a courthouse .....	144
Figure 4.16: List of courthouses locations (WomensLaw.org).....	144
Figure 4.17: Find a doctor/hospital.....	146

Figure 4.18: Find a doctor (Sharing the location and providing the specialty) .....	146
Figure 4.19: List of doctors (the US News Health, <a href="https://health.usnews.com/">https://health.usnews.com/</a> ).....	147
Figure 4.20: Financial resources .....	150
Figure 4.21: Find a food pantry .....	156
Figure 4.22: List of nearby food pantries (findhelp.org) .....	156
Figure 4.23: Low-cost healthcare .....	160
Figure 4.24: List of health centers (findahealthcenter.hrsa.gov) .....	160
Figure 4.25: What is domestic violence? .....	163
Figure 4.26: The Power and Control Wheel .....	165
Figure 4.27: Gathering evidence.....	168
Figure 4.28: Reporting abuse .....	170
Figure 4.29: Forms of technology abuse.....	176
Figure 4.30: Phone safety tips.....	177
Figure 4.31: Risk assessment- Questions.....	178
Figure 4.32: Risk assessment- Risk score and level .....	179
Figure 4.33: Priority setting .....	180
Figure 4.34: Priority setting - summary of priorities .....	181
Figure 4.35: The process flow in interacting with the API.....	182
Figure 5.1: Task procedure .....	202
Figure 5.2: Boxplots of emotional support of SafeHaven and a search engine .....	213
Figure 5.3: Boxplots of informational and instrumental support of SafeHaven and a search engine.....	214
Figure 5.4: Boxplots of usability of SafeHaven and a search engine .....	217
Figure 5.5: Boxplots of the perceived anonymity of the SafeHaven and a search engine .....	219
Figure 5.6: Bar charts of perceived trust of SafeHaven and a search engine .....	220
Figure 5.7: Boxplots of emotional support of SafeHaven and ChatGPT .....	222
Figure 5.8: Boxplots of informational and instrumental support of SafeHaven and ChatGPT ..	223
Figure 5.9: Boxplot of usability of SafeHaven and ChatGPT .....	226
Figure 5.10: Boxplots of perceived anonymity of SafeHaven and ChatGPT .....	228
Figure 5.11: Bar charts of perceived trust of SafeHaven and ChatGPT .....	229
Figure 5.12: Question formatting issue (Highlighted in red rectangle) .....	235

Figure 5.13: Responses display issue (Highlighted in red rectangle) .....	236
Figure 5.14: Multiple zip code inputs issue (Highlighted in red rectangle) .....	239
Figure 5.15: Missing city and state information issue (Highlighted in red rectangle) .....	239
Figure 5.16: Navigation issue (Highlighted in red arrow) .....	244
Figure 5.17: Inefficient responsiveness to specialized tips (Highlighted in red rectangle) .....	244
Figure 5.18: Inefficient handling of specialized tips (Highlighted in red rectangle) .....	248
Figure 5.19: Excessive use of emojis (Highlighted in red arrows) .....	252
Figure 5.20: Language sensitivity issue (Underlined in red) .....	254
Figure 5.21: A summary of strengths of SafeHaven .....	255
Figure 5.22: A summary of SafeHaven issues with suggestions for improvement .....	268

## 1 INTRODUCTION

Throughout this chapter, we outline the motivations of this dissertation research. We also present the research questions, the scope, and the outcomes of this dissertation research. Additionally, this chapter discusses the research contributions.

### 1.1 Motivations

Domestic Violence (DV) is a major problem that affects the public health for individuals, impacting more than 35% of women and 28% of men in the United States who have experienced physical violence, sexual violence and stalking by an intimate partner [1]. Compared to physical and sexual violence, emotional abuse is even more common, in which approximately half of women have been exposed to this form of DV over the course of their lifetime [1]. Moreover, the vast majority of DV incidents are not visible to society as those cases are not reported to the police [2], [3], [4]. Despite that self-disclosure might help survivors to receive positive social reaction [5], they are often hesitant to disclose their stories due to barriers like feeling of stigma, shame, guilt, fear of losing children, confidentiality, and fear of revenge or not being believed [6], [7], [8], [9]. A study [5] found that negative social responses to disclosure of the assault, along with a total lack of support, led to the emergence of PTSD symptoms through maladaptive coping strategies and a decreased sense of control over the recovery process. In the same context, a substantial body of empirical and clinical evidence indicates that women who have been severely maltreated are socially isolated and have few people in their social circles who can provide them with support [10], [11], [12], [13]. Besides psychological and physical problems, studies have shown that survivors of DV are highly susceptible to economic hardships, such as

financial inadequacy and the loss of their jobs, due to the patterns of coercive control and threats that abusers exert over their victims [14], [15].

Social and cultural influences, coupled with a lack of knowledge, and tangible and emotional support, can present challenges for DV survivors in managing their issues and seeking assistance from the community [16], [17], [18]. According to the study [19], survivors of DV are reluctant to seek help from formal agencies (e.g., health services, law enforcement, community services) due to a lack of awareness, access issues, consequences of disclosure, lack of material resources, personal obstacles, and system deficiencies. Additionally, it is not uncommon for DV survivors to avoid asking for support from their family and friends due to feelings of embarrassment surrounding disclosing violence within their relationships [10]. Providing social and emotional support to women who have experienced DV is crucial to their psychological and behavioral well-being [20]. A lack of knowledge, emotional and tangible support may also result in the continuation of abuse, which can lead to both serious physical and psychological consequences. Thus, to ensure survivors can effectively cope with their incidents from the outset, it is critical to provide them with informational, instrumental, and emotional support.

Despite the aforementioned barriers, some DV survivors might turn to hotlines and DV agencies where trained counselors can offer informational, emotional, and instrumental assistance. Additionally, some DV survivors also might prefer to seek tangible support from other DV survivors (i.e., self-help groups) and share their experiences to gain a deeper understanding of how they cope [21]. Furthermore, as technology advances, technology-based interventions, including mobile applications, web-based or social media, are becoming common means of intervening against DV and offering appropriate support to survivors [22]. In recent years, mounting evidence has suggested the feasibility and usefulness of digitally delivered

interventions for preventing violence, enhancing the safety of people involved in abusive relationships, and connecting them to trusted help sources [23], [24], [25], [26]. It is proven that DV survivors prefer tech-enabled interventions and guided online support as a safe, usable, understandable, and effective approach [23], [24], [25], [26]. By incorporating safety priorities, risk assessments, resources and information about abuse types, and referrals to trusted care, these interventions can be used to minimize the victim's burden while enhancing safety [23]. Online spaces (i.e., social media) provide users with opportunities to remain anonymous, which facilitates disinhibition [27] and creates a safe environment for seeking emotional and informational support, and revealing sensitive information [28].

In technology-based interventions, Conversational Agents (CAs) could play a significant role in this regard. Researchers have stressed the importance of anonymity and safety, which CAs offer, when assisting survivors in reporting the abusive acts they encountered and mitigating the challenges of face-to-face communication [29], [30]. An increasing interest in HCI research has been directed toward developing service technologies (e.g., CAs) to support recovery from chronic health conditions [31]. Studies have found that the use of counseling CAs has encouraged participants to report PTSD [32], improved awareness of asthma symptoms and triggers, and helped patients adhere to treatment [33]. CAs can be less burdensome for first responders since survivors can communicate with them by text or voice and are not afraid of being judged or divulging secrets [34]. CAs can listen to survivors' stories and offer them timely information as needed. Further, whenever anonymity and privacy are protected, people are more likely to respond to sensitive questions [35]. CAs' emotional support has demonstrated effectiveness in reducing anxiety and stress among individuals [36]. Beyond offering information, CAs can also be perceived as sources of emotional support for survivors of sexual



abuse as they cope with the challenges they encounter [29]. There is a growing interest for researchers and practitioners to explore the potential of CAs to provide survivors with both informational and emotional support.

Despite the critical role of CAs, limited research has focused specifically on providing opportunities for DV survivors to share their experiences with CAs and receive appropriate assistance. Therefore, our goal is to examine the feasibility of utilizing a CA to support DV survivors in accessing appropriate support. Specifically, we plan to design and implement a CA to assist DV survivors.

## 1.2 Scope of Research

CAs remains a less investigated intervention method for DV survivors despite their increasing popularity in the field of HCI. There are only a few studies and/or projects discussing the role of CA in the DV field. However, those studies expose several limitations. First, although CAs, such as "Companion Chat" [37] and "#MeTooMaastricht" [30] are designed to assist survivors, both remain in developmental stage and are not available to the public. Second, some CAs have been focused only on one specific type of abuse (i.e., sexual assault) [29], [38], [39] or target population (e.g., a specific nation, or gender)[29], [39], [40], [41], [42], making it difficult to generalize them to the full spectrum of DV or DV survivors. Also, many of these CAs were not empirically evaluated by users, raising concerns about the validation of their user experience, as well as a lack of insight into the communication style, dialogue flow, language, privacy, and security effectiveness of CAs [30], [37], [40], [43], [44]. In addition, several CAs offer a limited number of modalities (i.e., only buttons or only text) or features and support [29], [40], [44]. In the same context, some CAs lacked decision aids, which could have helped alleviate conflict arising from uncertainty about priorities or their risk during decision making [29], [38], [41],

[43]. A further limitation is that even though several CAs are being developed for the purpose of supporting DV survivors, there remains a notable lack of explicit discussion regarding their design principles and meta-requirements, which limits their research and practical implications.

To address the limitations of prior works and offer an alternative effective technology-based support for DV survivors facing various obstacles (such as marginalization, and secondary victimization) in seeking assistance, I designed and developed a CA named “SafeHaven” to assist DV survivors. Some of the design artifacts of SafeHaven include risk assessment, customized feedback, empathic responses, safety plan development, abuse screening and recommendations, referral resources, survivor stories, and establishing safety priorities. In particular, SafeHaven provides DV survivors with informational, emotional, and instrumental support, which may be of great importance to them. This dissertation research aims to address the following main research questions:

RQ1. What role do technology-based interventions play in supporting DV survivors?

RQ2. How can CAs be effectively designed and utilized for DV interventions?

a) What are the design requirements and principles for CA-based intervention of DV?

b) What are the expected opportunities and challenges of using CA-based intervention of DV?

RQ3. How to operationalize the design requirements for CA-based intervention of DV?

RQ4. What are the key features and capabilities of CA-based intervention of DV?

RQ5. How effective is the proposed CA (i.e., SafeHaven) in providing support to DV survivors?

- a) How does the proposed CA compare with a search engine?
- b) How does the proposed CA compare with ChatGPT?

To address these questions, my dissertation research proceeds in three main phases: the identification of the requirements and guidelines for the design of the CA, the development of the CA, and the evaluation of the CA.

In the first phase, my research objective was to gain a comprehensive understanding of the various aspects of DV intervention to inform the design of the CA's features and interaction. For that purpose, I conducted in-depth semi-structured interviews with DV domain experts who had experience working with survivors. The results of our study suggest many requirements and features for the development of a CA-based intervention of DV, with an emphasis on providing emotional support (e.g., empathizing with the survivors' experiences, practicing active listening) and ultimately providing them with relevant information (e.g., offering safety strategies, tips, and advice) and tangible support (e.g., making referrals to local resources). Further, there were many potential challenges associated with the design and development of CA support, including privacy, usability, and its ability to provide emotional support.

Secondly, I developed a CA, SafeHaven, to offer DV survivors essential informational, emotional, and instrumental support in coping with DV. Based on the professionals' guidance and recommendations, along with the reviewed literature in the role of CA-based intervention of DV in this context, we developed five main modules for SafeHaven, including: 1) greeting and assuring the safety, 2) screening and identifying abuse, 3) developing the safety plan, 4) survivor stories, and 5) referral resources. In our design, we have created several submodules under each main module, especially the "referral resources" module, to meet the needs of different types of support. A rule-based approach was used to develop the above-mentioned modules and

submodules. In addition, we implemented a retrieval-based approach in our CA. Particularly, we trained our agent with related phrases and keywords, and connected those phrases and keywords to relevant modules or submodules so that survivors could receive specific types of support in accordance with their needs.

Thirdly, we evaluated the effectiveness of SafeHaven in comparison to search engine and ChatGPT. We conducted an online experiment with 36 participants, including DV survivors, friends and family members of survivors, and professionals. Using both quantitative and qualitative methods, this evaluation focused on both effectiveness and perception variables, such as provisions of emotional, informational, and instrumental support, information quality, usability, perceived anonymity, and trustworthiness. In comparison with both a search engine and ChatGPT, SafeHaven demonstrated significantly higher levels of emotional support, information, and instrumental support. SafeHaven provided superior information quality over search engines, excelling in accessibility, consistent information representation, and error-free content, while also outperforming ChatGPT in terms of its concise information presentation, ease of understanding of information, and sufficient volume of information, yet its usability scores were on par with the two alternative platforms. Moreover, participants perceived SafeHaven as offering greater anonymity and trustworthiness compared to search engine, and more trustworthy than ChatGPT. In the evaluation, SafeHaven received highly positive feedback on its educational content, human-like interactions, and tailored support, which all indicate its effectiveness as an intervention tool for DV. On the other hand, some participants experienced a number of challenges with SafeHaven, including difficulty in forming queries, delays in response, and navigation issues, and made suggestions for enhancing user query, response time, information coverage, and user interface.

### 1.3 Structure of Dissertation

There are six chapters in this dissertation. Chapter 1 outlines the dissertation's subject matter, objectives, context, motivations for the CA-based intervention of DV, research questions, and contributions.

Chapter 2 provides a literature review of related work. In view of the interdisciplinary nature of this dissertation research, the literature review covers topics relevant to DV, traditional and technological DV intervention, and its influence on DV survivors. Further, we discussed the technology of CA, how it works, and its role in the context of DV and social support.

Chapter 3 introduces the requirements and design principles for the CA-based intervention of DV. In addition to the research methodology, this chapter also discusses research findings, including the opportunities and challenges of the using CA-based intervention of DV.

Chapter 4 covers the design and implementation of SafeHaven. The chapter presents the process flow of our proposed system and introduces its key features and functions. Moreover, it provides an in-depth review of the CA system design, followed by system implementation.

Chapter 5 describes the evaluation methodology, measures, and findings about SafeHaven based on user perceptions.

Chapter 6 summarizes the research and discusses research implications and possible future directions.

## 2 BACKGROUND AND RELATED WORK

### 2.1 Background

Domestic violence (DV) is a form of abuse of power, which involves the act of controlling, threatening, or victimizing another through physical, verbal, emotional, or sexual violence [45]. The incidence of DV can be influenced by several factors, such as culture and traditions, lack of protection from authorities, poverty and unemployment, psychological issues, and substance abuse [46]. In the United States, 24 people are physically or sexually abused by their partners each minute, which represents approximately 12 million women and men annually [1]. Additionally, a quarter of U.S. women indicated that their experience of violence had adverse effects, while 22% expressed concern for their safety [47]. As part of this research, we aim to provide a technology-based intervention tool for victims of physical and emotional aggression.

There is a wide variety of forms that DV can take, including physical, emotional, financial, sexual, and stalking. Physical violence is defined as using physical force in an intentionally harmful manner with the intent of inflicting death, injury, or harm [48]. This type of violence includes scratching, pushing, shoving, throwing, grabbing, slapping, hitting, and the use of weapons [48]. As opposed to other forms of violence, victims of physical abuse are at higher risk for serious injuries and even death. Furthermore, this type of violence is more prevalent, with 25% of women and 14% of men experiencing severe physical violence in their lives [1].

Emotional, or psychological violence involves the use of verbal or non-verbal communication in an attempt to harm a partner emotionally or mentally [48]. Emotional aggression involves different images like coercive tactics, isolation, calling names, taking advantage of the victim's vulnerabilities, etc.[48]. Even though emotional violence doesn't imply physical acts and may not look like aggression in some cases because it is covert and manipulative, it can have serious long-term repercussions such as chronic pain, depression, and anxiety. Nearly half of women and men have experienced emotional violence, as the most common form of violence, at some point in their lives (48.4% and 48.8%, respectively) [1]. Along with physical violence and emotional abuse, sexual abuse is experienced by between one-third to one-half of women in toxic relationships [49], [50]. Sexual abuse can take many forms, including sexual physical attacks, sexually exploitative behavior, and coercive acts [51]. In addition to physical effects such as fatigue, bruises, and injuries, sexual abuse may result in gynecological issues such as bladder infections and miscarriages [51], [52]. It is also common for women who have been sexually abused by their partners to suffer serious psychological effects, such as anxiety, depression, and suicidal thoughts [53]. Besides, DV survivors might be susceptible to financial abuse, in which a survivor is controlled in their acquisition, use, and maintenance of economic resources, thereby threatening their economic independence and socioeconomic security [54]. As an example of financial abuse, it might be the inability of a survivor to secure and maintain employment, the inability to use resources they already possess, the theft of their money, the creation of costs, or

the generation of debt. In addition, One form of abuse that exerts control over the survivor is stalking. It refers to surveillance that is repeated and unwanted by one or more individuals toward another. In addition to harassing and intimidating survivors, stalking behaviors may also include keeping a close eye on them or following them with intent [55]. According to the CDC, In the U.S., 16.0% of women (or 19.1 million) have experienced stalking at some point in their lives, experiencing feelings of fear or conviction that they or their loved ones will be hurt or killed as a result [56].

A major aspect of DV, besides the form of violence, is determining the pattern of violence. The pattern of violence might take different forms like single victimization, where only one violent act is experienced by a survivor, or repetitive victimization, where an abuser attacks the same victim several times in a row [57]. In their analysis of the National Crime Victimization Survey (1992-2004), Goodlin et al. found that 80 percent of households experienced victimization once and 15 percent experienced victimization multiple times [57]. Research also has demonstrated that one of the prominent features of domestic violence is its frequent repetition [58]. The pattern of repetitive violence is a greater indicator of the likelihood of future victimization than any other characteristic of crime, which means that an effective response must be made before the situation worsens and the victim faces potentially lethal consequences.

In order to alleviate the psychological and physical effects of DV, it is imperative that survivors receive social support [59]. Social support is information leading the individual to feel



loved, cared for, valued, and a part of a network of shared obligations [60]. According to abused women, it is a multidimensional construct that they prefer to implement as an intervention [61]. A low level of social support may act as an exacerbating factor for DV and increase its risk. In light of the above facts, and to provide assistance, it is vital for effective and accessible social support to be provided, including informational, emotional support, as well as instrumental assistance.

## 2.2 Non-Technology Mediated DV Intervention

### 2.2.1 DV Survivors' Self-Disclosure

In the context of self-disclosure, it's considered as a "willful disclosure" of facts and feelings with the intent of "letting other people know without a doubt of what has been done, how you feel, etc." [62]. A key factor that influences interactions about sensitive topics (e.g., DV) and may help build an emotional rapport is the act of self-disclosure. The goal of self-disclosure is to find a balance between keeping oneself hidden and being vulnerable. Typically, victims who self-disclose do so to identify other stigmatized victims and seek assistance and support [63]. In families, sharing DV experiences with those outside of families can be challenging, and hence data regarding DV may be difficult to access. There are many issues associated with disclosure related to a victim's emotional state, further complicating inquiries about DV. Usually, victims refrain from disclosing their experiences to people face-to-face if they risk losing face [64], or might feel that they are being judged [65]. Further, Despite adverse health effects and frequent visits to health care providers, DV victims rarely disclose their experiences with providers [66]. Embarrassment, frustration, regret, worries about privacy, and a fear of rejection from the physician come into play as barriers that prevent survivors from disclosure [67], [68]. In addition, compulsory reporting laws, which mandate physicians to report DV to the police, may

further hinder disclosure [69]. Moreover, A lack of a coordinated screening program for victims is suggested to have a detrimental effect on disclosure [70]. Other reasons for nondisclosure have also been identified, which include minimizing and normalizing the severity of the violence, concern for social support, not wanting to publicize their private matters, and lack of trust in people around them [71].

In contrast to barriers, there are several facilitators for the disclosure of DV [72]. For victims to be able to disclose, it is essential to build a positive relationship with their healthcare providers [73]. A direct questioning strategy is effective at encouraging disclosure by victims of DV [74]. In addition, the victim's feeling that they can decide what to reveal and to what extent, encouraged them to disclose more [75]. Moreover, the safe setting in healthcare, as well as ensuring their information is kept confidential, will contribute to their willingness to disclose [73]. Other factors that helped victims disclose were their perception of healthcare providers as having a high level of knowledge, professionalism, and ability to take action to stop violence [76].

There are also other factors (e.g., race, gender, age) that are associated with the willingness of DV disclosure. For instance, it has been noticed that the racial majority (i.e., White) in the United States was more likely to disclose their DV experiences to their social network members than racial minorities (i.e., African American, Latino, etc.) did [77]. Furthermore, with respect to gender, the amount of information disclosed by men and women differs considerably. Specifically, studies [78], [79] found that women are more likely than men to report and disclose DV, despite the fact that it is under-reported in general. Additionally, despite DV research primarily involving samples of adults, it has been observed, in research that has included samples of college and high school students, that younger victims tend to be more likely than older

victims to reveal their DV experiences [80]. Due to the fact that most of the literature focuses on adults, there is little evidence to determine whether age influences disclosure.

A key element of disclosure is determining when and why it is beneficial. Chaudoir et al. [63] developed a framework called the Disclosure Processes Model (DPM), which aims to address when and why the disclosure of interpersonal or verbal information might be beneficial for individuals who live with stigmatized identities in the context of human-to-human interaction. According to the framework, the pre-disclosure goals might moderate the impact of disclosure on an individual (i.e., behavioral) and dyadic (i.e., trust) outcomes, but they have less impact on social contextual results (i.e., traditions and norms). In the context of outcomes of disclosure, it has been revealed that clinical conversations with victims about DV have an Influential impact on both benefits (e.g., leaving the abusive relationship, reporting the case to the police, self-esteem, self-awareness, etc. ) and risk (e.g., stress and alienation from the medical system.) [81]. Victims also report experiencing positive (e.g., believing the report, validating the victim's experience) as well as negative (e.g., disbelief and blame) reactions from society following disclosure, with positive reactions being the most prevalent [78].

### 2.2.2 DV Intervention

Despite the barriers mentioned above (section 2.2.1), which may discourage some survivors from seeking support, some survivors take action and seek assistance from either formal (e.g., police, shelters, DV agencies, counseling services, or healthcare professionals) or informal (e.g., family members, friends, and colleagues) sources of support. It is possible to prevent future DV or reduce or tackle the negative consequences experienced by victims and their children when emotional, informational, and instrumental needs are met [82]. There are various approaches to offer support and prevent such DV actions like a) victim's care service, which involve hotline,

shelters, counseling, and medical care, b) legal protections, which include law enforcement, and protection orders, c) patient-centered approaches, and this focus on understanding the basics of DV screening, reproductive coercion, and other behavioral dangers, d) providing support for victims of DV, and this includes treatment delivered by mental health practitioners to reduce the negative effects of DV, e) training programs for parenting and family relationships, f) ensure financial stability for households, by enhancing support systems in the case of unemployment or adequate income, as these are risk factors for DV. It is more likely that survivors who experienced severe violence and control will seek out formal services (e.g., healthcare providers, police departments, legal services, DV shelters, and emergency centers) for informational or instrumental support like safety planning, financial assistance, child care, shelters, or transportation assistance [83]. In the same vein, friends and family members have been the most effective sources of support for survivors, especially when it comes to providing emotional support [78], [83]. Positive associations were found between social support and better mental health outcomes, regardless of the severity of the violence the individual suffered [59].

## 2.3 Technology-Mediated DV Intervention

### 2.3.1 DV Survivor's Self-Disclosure

In many instances, survivors are reluctant to disclose their abuse due to personal, social, and cultural reasons that we discuss in section 2.2.1. Consequently, they may consider other avenues for disclosing their experiences and seeking assistance. Over the past few years, survivors of DV have begun using technology to "feel out" and initiate help-seeking as a means of being heard and also receiving guidance from others regarding whether or not to seek professional assistance [84]. Through the use of technology, survivors can access services and receive support wherever they are, at any time, with complete anonymity, and without the need to disclose their abusive

experiences to a professional to access them [85]. Technology has often been observed to outperform face-to-face screening in terms of DV disclosure and positive outcomes for survivors. For instance, the study by Hussain et al. [86] concluded that computer-enhanced self-screening for DV was superior to face-to-face screening by 37% and paper and pencil screening by 23%, resulting in a higher rate of DV disclosure using technology-advanced strategies. Similarly, researchers found that respondents felt that they could be more open in a virtual environment without being concerned about adverse reactions due to the ease of seeking assistance using technology rather than in a face-to-face setting [87]. A comparison of the effectiveness of online and face-to-face interventions was also conducted by Costantino et al., and according to the researchers, the internet provides a ‘protected environment’ that permits survivors to manage their disclosures without fear of being judged, rejected, or devalued in person [88].

In recent years, survivors have increasingly turned to social media as a safe venue for self-disclosure. Social media provides efficacious effects such as the distribution of information, the articulation of opinions, and the mobilization of action [89]. Social media's proliferation, ease-of-use, and convenience have led to the reliance on them for different sorts of health and behavioral interventions that are ecologically valid. Globally, social media has become the primary means of communication over the last decade [90], and many efforts have been made to raise awareness about social issues, such as domestic violence, using social media. Mainly, in the offline context, several barriers contribute to the lower reporting to the law, including losing face [64], age dependency, roles of gender, distrust of law enforcement [91], fear of reprisal, or desire for privacy [92]. Additionally, shame or embarrassment and the fear of being believed deter the disclosure of DV [93]. Therefore, the difficulties associated with face-to-face disclosure have

resulted in many victims turning to other venues, such as social media, to share their experiences and disclose their stories to either unknown or known audiences [94], [95], [96], [97], [98], [99].

An extensive research effort has been conducted to explore the use of social media platforms as a lens of behavior and glean insights into the self-disclosure of DV experience of survivors. Social media facilitates victims' disclosure of their experiences and allows them to voice their concerns. There has been an emphasis among researchers in this field that the self-disclosed stories of the survivors carry many signs and themes in common. In these stories, there are accounts of being physically injured, sexually assaulted, controlled and isolated, mentally and emotionally manipulated, and financially ruined [28], [95], [98], [100], [101]. According to a deep analysis of self-reported survivors' posts [96], [100], it was apparent that these posts are packed with negative emotions, along with frequent references to first and third personal pronouns, such as the use of 'I' and 'he' when discussing their violent experience. Additionally, through their posts, it is common for victims to express the harm they are living in, the confusion that comes along with the relationship, and the fear about the future. Social media campaigns (i.e., #whyIleft/#whyIstayed [94], #MaybeHeDoesntHitYou [95], #NotOkay [102], #metoo [101]) provided an accessible meeting place for discussions on DV. For instance, McCauley et al. [95] examined the hashtag #MaybeHeDoesntHitYou and determined that Twitter can be used as a valuable resource for extending public discussion about IPV so that it encompasses all forms of abuse and coercion that one typically experiences in abusive relationships. Additionally, the tweet mining of #whyIleft/#whyIstayed was also successfully utilized to determine what factors contributed to survivors' decisions to remain or to leave the abusive relationship [94]. Similarly, authors in [101] emphasized that social media represents a change in the environment regarding disclosure of sexual violence, in which survivors have found a voice through online space after

having been silenced or ignored for long periods of time, and instead turning to online forums for expressing and getting support. And responses that survivors provided shed light on the complicated and challenging process they undergo when it comes to determining if they should leave abusive relationships.

### 2.3.2 DV Intervention

To address our first research question, we discuss the role of technology in the intervention of DV. Technology-based interventions encompass all uses of technology to provide educational, therapeutic, or psychological support to promote health among DV survivors. Technological interventions are becoming increasingly popular due to rapid advancements in patient digital knowledge, availability of devices, and the capabilities of these devices (e.g., mobility, usability, and accessibility). Mainly, technological interventions are being widely utilized due to COVID-19 on providing survivors of DV with a support system [22]. Technology-enhanced interventions are becoming more acceptable to DV survivors [25], [26]. In the last few years, there has been significant growth in the use of DV digital interventions delivered via mobile devices, web-based, e-mail, social media, and CAs.

In developing and adapting DV digital interventions, developers take into consideration social and behavioral sciences. Technology-based interventions are commonly supported by social cognitive theory [103], family system theory, technology adoption theory, trauma-informed, and harm reduction, which are all rooted in a social-ecological system [104]. DV interventions using technology can be delivered in many ways [105], [106], including a) providing ongoing emotional support through facilitating emotional relief, empowering survivors to feel validated, enhancing self-esteem, or providing them with ways to connect to others who are struggling with similar challenges (i.e., online support groups), b) allowing the survivor to develop

individualized safety plans and assessing risk factors in terms of frequency, severity, and type, c) offering information and educational resources to enhance survivor-centered decisions, d) improving survivor access to a range of services and resources, including medical care services, housing facilities, crisis centers, and education, and financial and legal assistance, e) enhancing the collection and documentation of evidence.

Research has demonstrated the effectiveness of DV technology-based intervention. First, they showed to have an improvement in the survivor's psychological wellbeing notably through the enhancement of feelings of support, safety, and a perception of health [88]. In particular, technology-based interventions have been observed to be effective in improving mental wellbeing by reducing depression and PTSD symptoms in survivors [25], [59], [88], [107]. Information accessible via technology to assist in making decisions can be accessed privately, safely, and on a timely basis [108]. Technology-based interventions also enable survivors to develop safety and action plans as well as provide them with strategies to enhance their ability to make informed decisions and be self-sufficient [109], [110]. Technology-based interventions become more important than ever since they effectively support low-income survivors of DV who are living in remote areas, particularly in areas with limited health care services where victimization might overlap with other factors contributing to violence [111], [112]. Technology interventions are perceived as having the ability to offer safe ways to escape or cope with abusive relationships and mitigate DV-related risks [26]. DV technology-based interventions also have addressed several shortcomings of a traditional DV intervention, such as inaccessibility due to geographical location or time constraints, lack of resources, affordability and care cost barriers, lack of confidentiality, insufficient training of health care professionals, the social



stigma associated with seeking care services, and socio-geographic diversity [25], [26], [88], [109].

The challenge of providing technical support to survivors of DV is considerable. It has been found that DV professionals (i.e., lawyers, counselors, therapists, and social workers) as well as survivors might lack technical knowledge [113]. Further, many low-resource communities are experiencing a technological gap that hinders the availability, accessibility, and ability of survivors to use technological interventions. In addition, user security, and data privacy are now vital considerations and challenges in technology-based DV intervention [22]. Plus, people with intellectual, cognitive, or perceptual disabilities (e.g., elderly people, deaf, visual impairment) might have difficulty interacting efficiently with existing technology. In the same context, abusers might use some devised strategies for monitoring survivors' online presence at all times [22], [113]. Abusers can be a major barrier to the use of technology-based interventions, as they misuse spyware and/or tracking apps for tracking, and controlling their victims [113]. In their paper, Finn and Banach describe potential concerns and risks that may arise when women seek health care and social assistance services on the Internet technology [114]. For instance, in some cases, the disadvantage of seeking assistance online is the possibility of accessing inaccurate information, such as when searching for information regarding "DV counseling," "Family services," or "shelter," users may be directed to sites that contain inaccurate information or may differ from the ones they originally requested. Besides inaccuracies in information, some websites might set out to victimize survivors deliberately. Due to the Internet's unregulated environment, anyone may establish a website and present themselves as an expert. This is particularly true when seeking online therapy or counseling. Further, due to a lack of community and ethical standards on the Internet, it may be difficult to distinguish between genuine and fake

"advertising." Generally, Internet information is not subject to any systematic quality control. Loss of privacy may also be a concern for people who participate in and post their stories in online support groups. Whatever the case, the message is likely to be read by anyone, and there can be no assurance that it will not be forwarded. Moreover, messages might be archived and searched using internet search engines. Some members of online support groups may inadvertently disclose personal information about other members, which may pose a serious problem for someone whose partner is abusing them or who is being stalked. In light of these limitations, it is important to consider technology that is capable of addressing these concerns and providing survivors with appropriate support. In this regards, our primary focus centers on providing survivors of DV with an anonymous technology system, namely CA, that provides coping methods and accessible resources to survivors of DV in order to assist them in taking the initial steps toward addressing their abuse-related issues.

### 2.3.3 Technology Support

People's communication patterns have drastically changed in the last two decades. According to statistics, in the U.S., there are over 295 million people who own a smartphone [115] and over 300 million people who do some type of online surfing at least once a month [116]. Recent advancements in smartphone and web technologies have greatly enhanced the capabilities of delivering health information to individuals. It is becoming more common for people to use wellness apps and web-based health platforms as a means to gain access to health information and ultimately improve their health [23]. It is recognized that a digital-based intervention would be an effective tool to support DV victims and strengthen their safety net [23], [24], [25], [26]. It has been proven through extensive trials and evaluations that multiple web-based and mobile applications-based interventions and safety decision aids (e.g., MyPlan [23], iSafe [25], BSAFER [117], LEAF [118], iCAN [119], Thrive [120], I-DECIDE [121], HELPP [88]) are both effective and efficient for several victim groups. Besides the web and mobile apps,

some other technological solutions (i.e., email, and video conferencing) were employed to build a healthy relationship and intervene in DV. Table 2.1 presents the major DV interventions that based on digital technology.

#### 2.3.3.1 Web-Based Intervention

The digital intervention of DV has been greatly enhanced by several web-based tools. These web-based tools are free and easily accessible and have been examined for their efficiency in several randomized controlled trials, as well as with multiple survivor cohorts. It has been demonstrated that web-based interventions provide ongoing information and emotional support, promote safety planning, enhance risk awareness and the severity of abuse, educate survivors (e.g., identify abuse warning signs), and route survivors to the appropriate resources in accordance with their particular social and personal circumstances [88]. One example of a web-based intervention is found in [110], in which researchers measured the effectiveness of a safety decision aid utilizing a web browser in alleviating decisional conflict among victims of abuse. In the intervention, three central components were employed: a) Priorities setting, in which abused women determine the priority for their safety concerning important objectives like having resources, child well-being, and a concern for their safety, b) Danger Assessment, providing feedback that allows women to understand their level of risk for repeat incidents of severe battering, fatal violence or near-fatal violence, and c) Safety Plan, which offers multiple safety strategies and educational materials of DV that are offered by national, and local resources. Observed outcomes from the analysis showed that the decisional conflict of abused women decreased significantly, and this led to the support of the decision-making process of IPV survivors without adverse consequences. In a similar vein, one study explored the effectiveness of iCAN [119], a web-based intervention focused on enhancing the safety and wellbeing of

women who had experienced domestic violence. As part of the study, women participated in interactive activities aimed at assessing their risk and strengthening their understanding of their responsibilities, as well as identifying their priorities and plans for their relationship. Researchers discovered that survivors experienced a reduction in depression and PTSD, as well as improved confidence in determining safety plans, mastery, social support, and the perception of coercion. Similarly, BSAFER is another web-based program that aims to address IPV and drug use [117]. This program has various components including automated feedback, goal setting, social support, advice, defining core values, and referral resources. BSAFER participants reported high levels of satisfaction with the program, evaluated its technical ease of use favorably, and were generally in agreement that the program complies with MI principles. Hegarty and colleagues [121] developed an online tool called I-DECIDE, comprising modules on relationships, abuse, priority setting, and a section tailored to safety planning. The findings indicate that the intervention was supportive of participants as well as a motivating factor for taking action. Similarly, Jane Koziol-McLain et al. evaluated the effectiveness of a web-based intervention tool (isafe) to assist women who are suffering from DV [25]. The site consists of three components: safety priority setting, danger assessment, and interactive process utilizing a matrix of resources to assist women in developing a personal action plan. The results showed that the site was effective in reducing IPV exposure among indigenous Maori women, decreasing their depression levels, and increasing their safety behaviors and that the information provided by the site was valuable to these women.

#### 2.3.3.2 Mobile Apps-Based Intervention

Modern technological advancements, especially the development of mobile applications, have created new opportunities for the promotion of health and the improvement of health outcomes.

In the United States, 85 percent of adults own a smartphone, an increase of approximately 50 percent from 2011 [122], [123]. Regarding the DV intervention, several key features of mobile apps are powerful for the survivor, as they can provide an anonymous, free, secure environment, as well as resources available 24/7 that facilitate the involvement of survivors in safety planning and decision-making. The use of mobile app-based interventions has been demonstrated to contribute to survivors' psychological and social well-being by enhancing emotional, instrumental, and informational support. Many apps were deployed on mobile devices to assist DV survivors. The most well-known tool is MyPlan [23]. The primary purpose of this mobile app is to assist survivors in making informed safety decisions. The app consists of various major features like raising survivors' awareness of "red flags" for violence in relationships, assessing survivors' danger and their risk of death, determining survivors' safety priorities suggesting personalized safety plans based on the danger assessment and priorities, and connecting the survivor with trusted safety resources, such as a hotline, healthcare provider, or police. In reflection of the survivors' evaluations, the app was regarded as an effective, understandable, and appropriate tool for them to understand the facts of DV, measure its severity and set their priorities for providing support. It is suggested that the app contributes to the reduction of decisional conflict among survivors and helps them develop safety plans for themselves and their children. Another mobile app, Thrive, was developed to address the health needs and improve the wellbeing of abused mothers [120]. This app includes three sub-sections: Myself, where the mother can learn about her self-care, coping strategies, and trauma-informed practices; My Child, which focuses on minimizing the stresses experienced by children, enhancing mother-child relationships, and discussing IPV with children; and My Life, which provides referral resources regarding housing, IPV advocacy organizations, and IPV and parenting hotlines

throughout the country. Both survivors and IPV providers consider Thrive to be trauma-informed, convenient, helpful, easier to use, and more effective than other health education methods (i.e., handouts). There is also a mobile app called HELPP Zone [124] designed specifically for college students to protect them from potentially violent relationships. In the HELPP Zone, a user can specify trusted contacts, who can be contacted to provide assistance based on the location, schedule, and condition of the user when the risk of violence exists. In the same topic, Lindsay et al. [125] developed a mobile app that was intended to assist college-age women in analyzing the danger associated with their abusive relationships, setting priorities, and developing a customized safety plan. The participants were positive regarding the app's ability to provide confidential and nonjudgmental information and resources about abusive relationships. In their evaluations of the application, participants found it clear, understandable, relevant, and comprehensive.

#### 2.3.3.3 Other Technologies-Based Intervention

In addition to web and mobile app-based intervention, it is worth noting that other means of technological solutions for the sake of developing healthy relationships, helping survivors, and DV intervention were also used. For example, the use of videoconferencing technology has been used successfully to provide trauma-focused treatment to rural survivors of DV and sexual abuse [107]. As a result of this solution, participants showed substantial reductions in PTSD and depression symptoms following treatment. In a similar context, Tschirch et al.[126] utilized videoconferencing as an effective method of delivering mental health care to victims of DV. The results of this study demonstrated that this technology could help survivors of DV with their unmet needs. Email is another technology that serves the same purpose. In particular, Constantino et al.[127] and Crane and Constantino [128] examined the feasibility of using e-mail

to communicate and interact with mothers who have experienced IPV and their teenage children, and concluded that email interaction was effective in delivering information, screening, safety tips, and support for women survivors of IPV and their children. Along the same line, authors in [88] compared email with face-to-face interaction as an intervention solution for IPV survivors and noted that both approaches showed a decrease in symptoms of depression, anxiety, and anger, along with an increase in personal and social support among the participants, with notable improvements in the email group.

Table 2.1: A review of studies on the technology-based Interventions for DV

Ref	Study design	Theme	Measurement tools	Main features	Primary outcomes	Technology medium
[117]	RCT	IPV intervention and drug use.	<ul style="list-style-type: none"> <li>• Satisfaction (CSQ-8), System usability (SUS).</li> </ul>	<ul style="list-style-type: none"> <li>• Identifying core values.</li> <li>• Setting goals.</li> <li>• Offering advice.</li> <li>• Providing referral resources.</li> <li>• Social support.</li> </ul>	<ul style="list-style-type: none"> <li>• Avg. (SUS): 83.5 out of 100.</li> <li>• Avg. (CSQ-8): 27.7 out of 32.</li> </ul>	Web-based
[119]	RCT	IPV intervention	<ul style="list-style-type: none"> <li>• Depressive symptoms (CESD-R),</li> <li>• PTSD Symptomology (PCL-C),</li> <li>• Decisional conflict (DCS),</li> <li>• Experiences of Coercive Control,</li> <li>• Helpfulness of safety strategies,</li> <li>• Confidence in safety planning for self,</li> <li>• Confidence in safety planning for Children,</li> <li>• Social support, Mastery.</li> </ul>	<ul style="list-style-type: none"> <li>• Danger assessment.</li> <li>• Priority setting.</li> <li>• Personalize action plan.</li> </ul>	<ul style="list-style-type: none"> <li>• Decreased depressive and PTSD scores.</li> <li>• Decreased mastery score.</li> <li>• Decreased decision conflict.</li> <li>• Improved social support.</li> <li>• Improved coercive control.</li> <li>• Improved self-efficacy and safety planning.</li> </ul>	Web-based
[23]	Open trial, without control and randomization	Safety decision aid	NA	<ul style="list-style-type: none"> <li>• Create a safety plan.</li> <li>• Danger assessment (DA)</li> <li>• Priority setting</li> </ul>	<ul style="list-style-type: none"> <li>• Participants perceived the app as useful, understandable, and appropriate.</li> </ul>	Mobile app
[121]	RCT	Healthy relationship, and safety decision aid.	<ul style="list-style-type: none"> <li>• Self-efficacy (Self-Efficacy Scale).</li> <li>• Depression (CES-D)</li> </ul>	<ul style="list-style-type: none"> <li>• Address healthy relationships.</li> <li>• Safety and abuse.</li> <li>• Priority setting</li> <li>• Build action plan</li> </ul>	<ul style="list-style-type: none"> <li>• The intervention was perceived as supportive and motivational by participants.</li> </ul>	Web-based
[120]	Open trial, without control and	Health education of IPV	<ul style="list-style-type: none"> <li>• Mobile App Rating Scale</li> </ul>	<ul style="list-style-type: none"> <li>• Provide resources.</li> <li>• Enhancing</li> </ul>	<ul style="list-style-type: none"> <li>• Participants agreed that Thrive was useful,</li> </ul>	Mobile app



	randomization			mother-child communication. • Maternal self-care	informative, and trauma-informed.	
[25]	RCT	Safety decision aid.	• Severity of Violence Against Women Scale (SVAWS) • Depression (CES-D)	• Safety priority • Danger assessment • Action plan development	• iSafe reduced IPV exposure. • Depression was significantly reduced.	Web-based
[125]	Open trial, without control and randomization	Dating and IPV, safety decision aid.	NA	• Danger assessment. • Priority setting. • Personalize action plan. • Assess relationship.	• Participant feedback indicates that the app is informative, easy to navigate, appropriate, and thorough.	Mobile app
[88]	RCT	IPV intervention, screening.	• Anxiety, anger and depression • Interpersonal Support Evaluation List (ISEL) • Personal Resource Questionnaire (PRQ) • IPV Experience Questionnaire (IPVEQ)	• Educational content, assignment tasks, self-assessment prompts.	• Significant improvements were observed in anxiety, depression, anger, ISEL, and PRQ	Email
[107]	Open trial, without control and randomization	DV treatment.	• PTSD severity (PCL). • Depression (CES-D). • Wyoming Telehealth Trauma Clinic Client Satisfaction Scale (WTTCCSS)	• Psychotherapy services that focus on trauma.	• Significant reductions in the severity of PTSD and depression symptoms. • High level of satisfaction.	Videoconferencing
[118]	NA	IPV intervention, Social network concerned with privacy	NA	• Wall and self-posts. • Anonymous Communication features. • User profiles and connect to friends.	NA	Web-based

RCT: Randomized Controlled Trials

#### 2.3.3.4 Social Media

Social media's contribution to the intervention of DV and social support provision lies in its capability to spotlight DV's prevalence, origins, and effects. Through this medium, norms about DV can be changed by disseminating knowledge and ideas to a large audience. As social media is primarily about bringing people together, it offers the opportunity for survivors to receive support, not only from their friends and families but also from others who have been through a similar ordeal. Moreover, social media has become a vital source of education and empowerment for the public concerning critical public health issues. It provides a platform to counter stigma and raise public awareness of issues, such as DV [94], [95]. It has been proven that social media platforms contributed to the improvement of life satisfaction for survivors of DV by implementing Support Group Therapy [129]. Online spaces (what we refer to as social media) provide users with opportunities to remain anonymous, which facilitates disinhibition [27] and creates a safe environment for seeking help/support and revealing sensitive information [28]. In the same context, many survivors resort to various social media platforms and found as a proper space for catharsis and self-expression [130], which can enhance well-being by reducing inhibitions [63]. Due to the increasing ubiquity of social media, victims of domestic violence can use this tool to describe their experiences and seek justice and remedy [131], [132]. Other potential uses of social media might include providing an avenue through which non-reformed abusers can gain insight from other reformed abusers' stories and actions regarding the issue [133].

Many studies discussed the role of social media in light of its ability to enable survivors to seek or receive social support. For instance, Chu et al. [134] examined 8,343 randomly selected messages from a DV community on Sina Weibo and concluded that the online community

primarily provided social support to survivors of DV in the form of empathizing, reassuring, and encouraging them, sharing anti-DV advice and personal experiences for guidance, and also demonstrating a willingness to provide tangible aid, such as providing resources, assistance, or services. Furthermore, by analyzing more than 300 tweets, one recent study explored the social responses to disclosures of sexual abuse on the hashtag #NotOkay, and found that the majority of tweets were positive, and many provided useful information and emotional support to survivors [102]. In a similar matter, authors in [64] analyzed posts and responses on the Reddit platform concerning disclosures of sexual abuse and noted that survivors were more likely to receive the support they sought, especially in terms of information, esteem, and instrumental aid. Davies analyzed the Reddit conversations between survivors of DV and those who responded to their posts [135]. Through these conversations, commenters provided survivors with informational support by offering advice, guidance, and links to resources that assisted survivors in meeting their needs.

Regardless of the many advantages that social media provides to DV survivors, there are still a variety of social and technical concerns that user face and has an impact on DV settings, the amount and the type of treatment and support that a survivor requires, and survivor's privacy and safety online. It is known that many survivors turn to online and particularly, social media for emotional, informational, and instrumental support. It is possible, however, that they may encounter unsupportive or aggressive responses, namely secondary victimization, in a way that they are blamed or claimed to be advocating or exaggerating DV, so they deserve the action of DV [64], [134], [136], [137]. There has been some negativity, prejudice, or discrimination based on gender displayed in social media responses and discussions surrounding the highly charged issue of domestic violence [136]. Further, the disclosure of abuse might be minimized by

commenters by making jokes or responding in a sarcastic manner [137]. Moreover, the survivor may be questioned or doubted about the case in the sense that there may be more to it than what was posted by the victim [137]. In the case of disclosures involving topics stigmatized by society or containing sensitive content, positive reactions and responses are important. Adverse reactions can lead to stress, low self-confidence, a feeling of marginalization and exclusion, and a reduced likelihood of future disclosure [138]. Besides reactions from commenters and implications of content, survivors of DV prioritize their privacy and security concerns on social media platforms and attribute this due to the intimidation and coercion they face from their abusers and their repeated attempts to pressure them into staying or returning by sending them repeated threats online [139].

In light of this, we strongly urge the establishment of a new system to assist survivors in avoiding secondary victimization and other negative reactions that may result from their DV disclosures, as well as provide them with access to a variety of resources. It is noteworthy that some technology (e.g., Conversational Agents) may facilitate survivors' disclosure by conveying nonjudgmental feelings [140] and simultaneously offer them social support (i.e., emotional, informational, instrumental) that meets their needs. Despite prior research providing a basis for studying DV survivors' self-disclosure and provision of social support, it was limited to interpersonal interaction or some technology interventions (e.g., web-based, mobile apps, social media). As a way of filling this gap, this research proposal aims to develop and evaluate a CA that assists DV survivors in receiving support, increasing their awareness, and improving their access to services.

## 2.4 Conversational Agents

### 2.4.1 Fundamentals

#### 2.4.1.1 What are Conversational Agents?

Since the 1960s, many people have recognized the potential benefits of human-computer dialog in the area of health care. Particularly, in 1966, ELIZA, the first established chatbot, was programmed to simulate a conversation by matching patterns and substituting words, giving the user the impression they understood what the computer was saying [141]. One recent statistic shows that over 250 million adults in the U.S. own smartphones [115], and 21% of American adults use CAs for banking services, shopping, and sharing information [142]. Due to advancements in speech recognition, natural language understanding, and artificial intelligence, conversational agents are increasingly available and used. A chatbot, or conversational agent (CA), is a program that simulates a conversation with humans, mainly via the internet [143]. CA systems can now perform many everyday tasks like finding information, answering questions, suggesting solutions, and providing responses to users' requests [144]. Consumers interact with CAs using smart devices such as smartphones and smartwatches. In the U.S., there are nearly 58 million smart speaker users [145], and global smart speaker sales are forecast to reach close to \$3.5 billion in 2021 [146].

Conversational agents (CAs) are a newer form of communication that is increasingly integrated into personal technologies and devices [147]. Communication with CAs may take place via voice, text, and images. CAs are capable of understanding human speech and replying in a synthesized voice. The advancement of natural language processing (NLP) and machine learning (ML) over the past two decades has led to the further development of artificial intelligence (AI) systems that are equipped with conversational capabilities. Recently, several

voice assistants have gained popularity, including Siri, Alexa, Google Assistant, and Cortana, which are delivered primarily by voice commands or smart speakers. With these CA intelligent systems, users receive personalized, context-aware responses through natural language understanding (NLU), speech recognition, and ML algorithms. It is said that these agents assist in performing various tasks like online shopping, controlling smart appliances in the home, suggesting solutions, and disseminating news [148], [149], [150]. Using CAs, specialized tasks also can be performed to aid healthcare professionals, patients, or populations at risk. Their efficacy has been proven for a variety of health conditions, such as depression [36], [151], palliative care [152], diabetes management [153], stress management [140], HIV, and sexually transmitted diseases [154], [155], hypertension [156], asthma [33], smoking cessation [157] and language impairment [158]. The innovations in CAs over the years make them an indispensable part of many operations and activities today, along with more sophisticated NLP modules for answering automated questions and performing other functions. It was noted that some CAs were deployed on their own built platform (i.e., Woebot [36], Gabby [42], Harlie [158], Wysa [151]). While on the other hand, several recent CAs offer strong integration and are connected to a number of messaging apps, such as Facebook Messenger [40], LINE [159], WhatsApp [160], WeChat [157], Telegram [30], etc.

#### 2.4.1.2 Components

The CAs dialog system consists of 6 major components: speech recognition, natural language understanding, natural language generation, dialog management, task manager and text-to-speech synthesis [161], [162]. In speech recognition, a list of words is generated based on analyzed audio input. The natural language understanding translates textual user input into a semantic representation that can be understood by the dialogue manager. In natural language

generation, meaning representations from the dialogue manager are converted into text. Through the text-to-speech synthesis component, the text will be transformed into synthetic speech. An essential component of a dialogue system is the dialogue manager, which directs the conversation flow as well as manages the processes of all the other parts of the system. The task manager has a thorough knowledge and understanding of the task domain. A typical dialog system architecture is shown in figure 2.1.

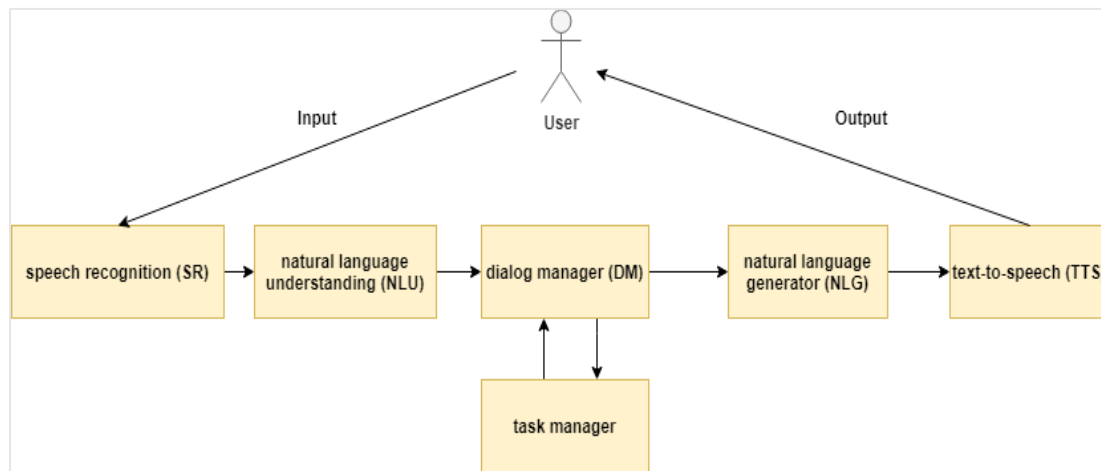


Figure 2.1: The architecture of a CA-based dialog system

#### 2.4.1.3 Approaches

CAs encompass a wide range of technical aptitudes, from simple to advanced/intelligent [144], [163], [164], [165]. Simple CAs, also known as finite-state CAs, depend on matching the user's input against predefined patterns that were created by developers. Thus, when discussing questions with the CA, the user has limited or no option to respond freely. The dialogue can be described as a sequence of states (Figure 2.2). By taking this approach, the user does not have the option of initiating the dialogue, and incorrect items are difficult to correct. The second approach is the frame-based system. This system fills the template based on responses provided by the user. As opposed to predetermined dialogue, dialogue flow depends on the user input. Additionally, users might provide more than what is requested in frame-based systems, and the

system considers these additional details and verifies if any more information is required before querying the database (Second dialog in Figure 2.3).

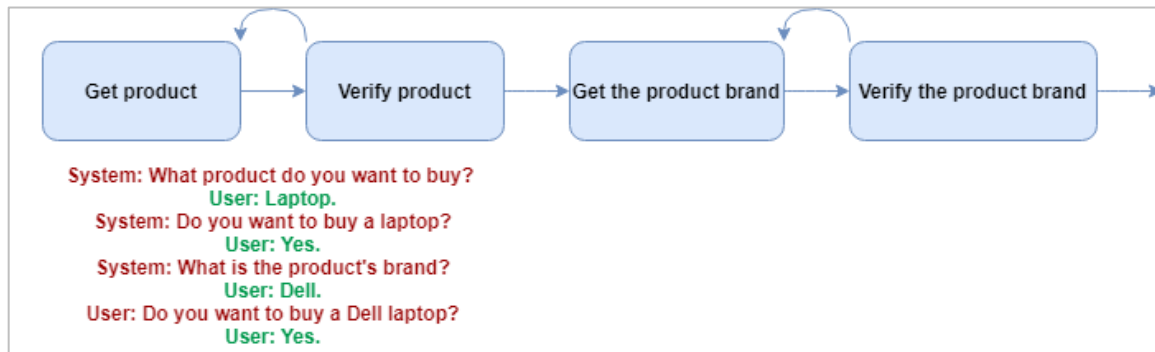


Figure 2.2: Finite state-based system

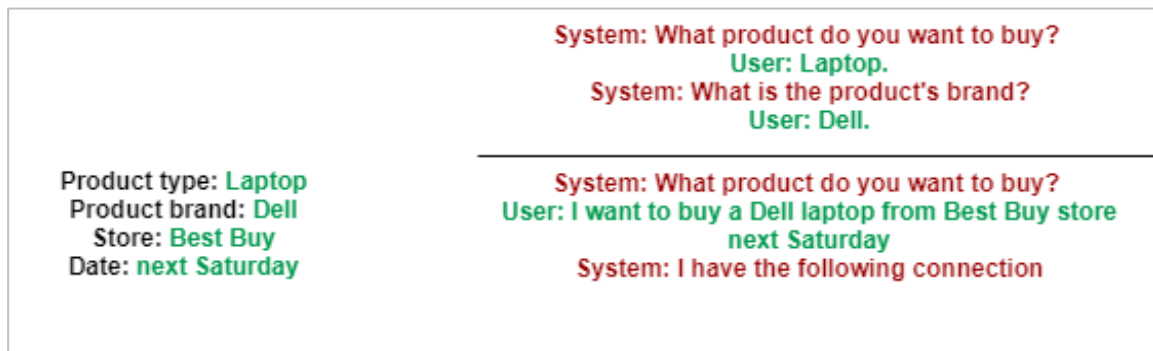


Figure 2.3: Frame-based system



Figure 2.4: An agent-based system

As opposed to finite and frame-based systems, agent-based systems allow users to initiate and drive the dialog according to their preferences [164]. Agent based systems are capable of managing complex dialogue. In this approach, dialogue models are developed in light of the preceding context and are dynamically organized as a set of interconnected steps that build on



each other (Figure 2.4). Through machine learning, the computer system's capabilities can increase as it learns from data without programming. Agent-based systems are mainly trained on data from real-world human-computer dialogue and offer more robust recognition abilities and execution speeds, allowing them to handle even the most complicated tasks in a more flexible and adaptable manner [166].

In terms of the evolution of CA technologies, ChatGPT represents an important milestone. As part of its architecture, ChatGPT uses transformer-based language models, a major advancement in deep learning that has led to a dramatic shift in NLP tasks. The model was constructed using the GPT-3 family of large language models from OpenAI [167]. With its extensive 175 billion parameters, this sophisticated language processing AI model can produce text similar to human speech, making it suitable for a variety of tasks, including translation, language modeling, and CA text generation. Notably, it has been regarded as one of the most influential and comprehensive language models to date. Its training data, gathered from a variety of textual databases on the Internet, amounted to 570GB and included articles, books, and web texts.

ChatGPT is designed on the basis of OpenAI's GPT-3 family of large language models and utilizes both supervised and reinforcement learning techniques [167]. There are three key steps involved in the development process: collecting demonstration data and supervised fine-tuning, collecting comparison data and reward modeling, and policy optimization against the reward model. As a first step, a model is fine-tuned using datasets provided by experts containing prompts and relevant answers. Secondly, learn the reward function that helps the agent decide what is right and what is wrong in order to move toward the goal. Through human feedback, the reward function is learned, ensuring the model generates accurate responses. The final step involves learning the optimal policy against the reward function using the Proximal Policy

Optimization (PPO). It is a new form of reinforcement learning introduced by OpenAI. Essentially, the PPO is designed to prevent too large policy updates from affecting agent training (See Figure 2.5).

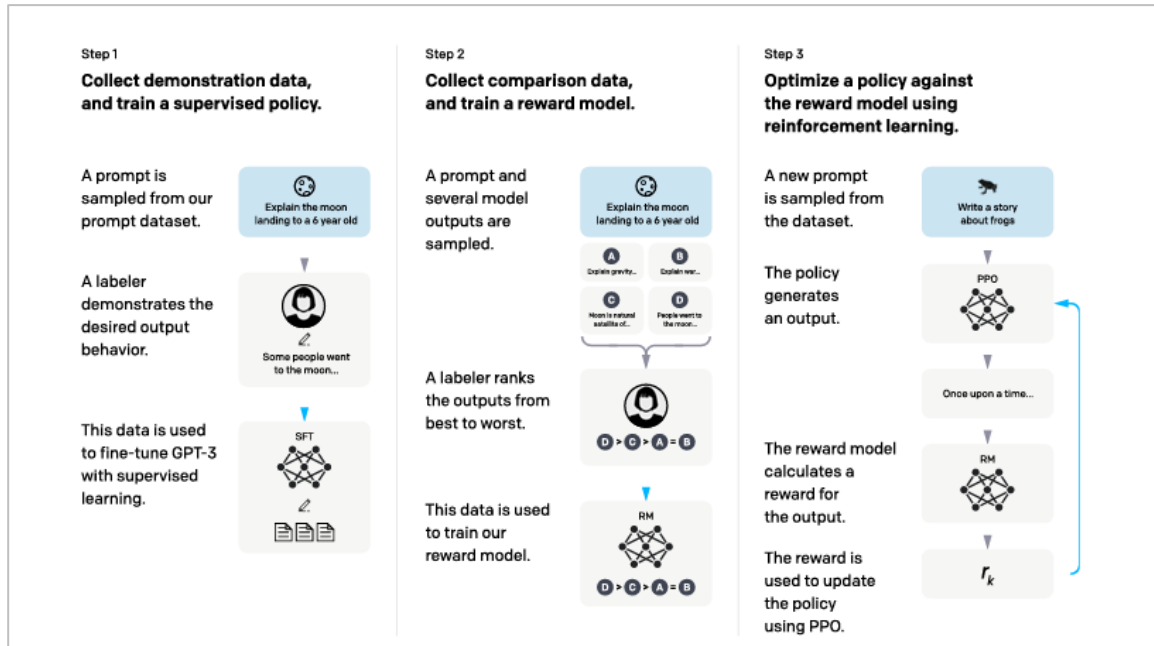


Figure 2.5: The ChatGPT training process [167]

By undergoing pretraining, ChatGPT learns to distinguish relationships between words and phrases in natural language. Having acquired this proficiency, the model is then able to generate responses in a conversation that are both coherent and realistic. Considering ChatGPT's advanced features and versatility, it stands out as an NLP model that can be applied in a variety of settings [168]. The efficacy of ChatGPT in revolutionizing interactions between humans and machines can be ascribed to its aptitude for comprehending context, proficiency in generating language, flexibility in handling diverse tasks, competence in multiple languages, scalability, capacity for zero-shot and few-shot learning, and ability to be fine-tuned [169]. The implementation of prompt engineering significantly improves the user experience and communication effectiveness in interactions with ChatGPT [170]. Through this practice, users

can guide the AI model, resulting in more precise, contextually relevant, and valuable responses. Such techniques include clearly define the prompt, explain the background and context, describe the preferred format and structure, impose constraints and limitations, and refine the prompt as needed [169].

#### 2.4.1.4 Evaluation

Optimal CAs are those that enable users to achieve their objectives with the least amount of inconvenience. A reliable evaluation is able to distinguish between various dialogic approaches and be able to explain key functions of a dialog system and is useful for human judgment [171]. CAs can be either task-based [172] or open-domain [173]. It is possible to gauge the usefulness of task-based CAs in various ways, such as task completion, consumption of resources, and conversation length, and this can be measured through several evaluation techniques such as simulations, user satisfaction, and manual assessments [171]. On the other side, in an open-domain dialog system, researchers discovered three metrics that can be measured automatically: fluency, coherence, and participation. Typically, CA evaluations can be categorized into two types: automated metrics evaluations and human-centered evaluations. Automated metrics evaluations are based on objective, quantifiable metrics and assess CAs without the need for direct human intervention. A number of automated metrics are commonly used, including BLEU [174], METEOR [175], and ROUGE [176]. Adapted from other fields, such as machine translation and text summarization, these metrics are designed primarily to measure the extent to which the system's output correlates with a set of reference responses. As a result of this approach, quick and consistent evaluations are possible across large datasets with high efficiency and scalability. The second evaluation approach is human centered approach. In this approach, subjective assessments are incorporated with expert evaluations to enhance user experience

comprehensively. Based on frameworks suggested by researchers such as Nielsen [177], these approaches stress standards compliance, real-world performance, and comparative analysis of systems as evaluation goals. An evaluation may be conducted at various stages of the development cycle and involve a variety of user engagement levels. Both qualitative and quantitative data may be collected during the evaluation process. Some prominent methods of such approaches include usability testing, heuristic evaluation, cognitive walkthrough, interview and focus group, survey and questionnaire, and field studies. Through such methods, experts or users provide insights into satisfaction, usability, and ultimately, this feedback helps refine user experience and create better AI technology.

#### 2.4.2 Trust and Self-disclosure

Since its inception, AI has made great strides with the recent advancements in its development, allowing CAs to act more convincingly like humans, such as assisting users in self-service operations [178]. In human-agent interaction, trust is a critical element since it plays a key role in influencing people to consider information provided by an agent, adhere to its recommendations, and gain the advantage of the capabilities of robotics [179]. Further, the trustworthiness of CAs is critical since these agents are used in many domains with a high level of importance, including clinician assistance [32] and decision-making processes [180]. Earlier research testing trust in automated machines has indicated that an automated machine's ability to perform tasks accurately and meet a threshold level of information accuracy is essential for building trust [181], [182]. This indicates that users build trust when they can effectively predict the agent's reliability and actual capability. There appear to be different characteristics of CAs that determine whether people perceive them as trustworthy, like sociability, style of communication, anthropomorphism, non-verbal skills, and performance excellence [183].

Besides, environmental factors such as culture, task difficulty, and task nature moderately impact the trust between humans and agents [182]. For example, according to research, showing non-verbal cues, initiating small talk, and implementing anthropomorphic features helped increase trust between an agent and the user [184], [185].

With the regard to self-disclosure, The CAs can act as a catalyst to encourage self-disclosure for different reasons, including availability, anonymity, ease of accessibility, and their unjudgmental nature [186]. On one hand, and according to recent research in HCI [29], [32], [34], [187], people have expressed their desire and preference for disclosing sensitive information to CAs rather than humans. For instance, a study by Lucas et al. [32] has found that people revealed fewer signs of depression when speaking with a human interviewer than when speaking with a virtual agent. While interacting with agents, users reported less fear of potentially inadvertently disclosing personal information and believed machines are unlikely to misuse sensitive information [34]. Earlier research indicated participants' preference for CAs in cases of highly sensitive topics or those that could evoke negative self-admitted behavior [188]. To a certain extent, CAs might come to be perceived as intimate or companionship by individuals [189]. There is also evidence that suggests reciprocity occurs in human-CA interactions and that disclosure by a CA encourages humans to self-disclose in a variety of areas [190]. Along the same line, recent work has demonstrated that the role of a CA in self-disclosure was reciprocal, i.e., it caused participants to reveal more information about themselves, thus improving the participants' perception of intimacy and in turn heightening their willingness to disclose sensitive details about themselves (mental illness, abuse in relationships, etc.) [187]. Researchers evaluated web surveys against CAs and determined that participants were more likely to provide higher quality data when using the CAs [191]. Moreover, the Wysa CA shows

efficacy and higher levels of engagement by users, who also presented higher self-reported levels of depression symptoms to the CA [151]. Further, it has been observed that CAs expressing empathy can increase patients' likeability and feelings of support when they divulge information about sexually transmitted diseases [155].

On the other hand, some evidence indicates that individuals would give information to CAs as they do to humans [192], or might give fewer details to CAs [193]. This might be explained by the question's nature and its influence on the disclosure level. Previous analysis suggested that relatively non or mildly sensitive questions may not cause significant self-disclosure variations between humans and CAs, while users tend to express a greater inclination to disclose to CAs rather than to humans when they are asked a sensitive question that might be perceived as embarrassment [192].

#### 2.4.3 Perceived Anonymity

A one key design consideration in social technologies is anonymity in which it facilitates support provision and seeking in socially stigmatized contexts [64]. It has been demonstrated that the internet offers a degree of control and anonymity, thus enabling individuals to express their opinions more freely, as well as having a greater degree of control over how their personal information is disclosed [194]. Despite anonymity having beneficial outcomes like seeking or receiving direct and indirect support in stigmatized and sensitive contexts (i.e., DV), browsing and searching for information securely, and social networking, it has a number of detrimental outcomes and adverse implications, such as cyberbullying [195], lack of accountability and harmful behavior [27], inaccurate and misleading information [114], deception [196], and the promotion of false identities, including those claiming expertise in health and human services [114].

Compared to other technologies, CAs may be more anonymous due to the fact that they are artificial agents that will not divulge any information to others and maintain confidentiality [186]. It is therefore assured that no information will ever be disclosed to the public, making it easier for people to share private and sensitive information. By considering that a CA is without feelings, humans might feel more comfortable using and communicating with it. According to young people, CAs provide a more anonymous alternative to Google or telephone services in connection with their health-related information [197]. By securing an individuals' anonymity, CAs are able to minimize the threshold for people to obtain information they need. The anonymity which CAs afford can provide a safe environment where people can discuss private matters, and particularly relevant in a DV context. CA communication offers a sense of complete anonymity, more so than conversing with another individual via the same means [198].

#### 2.4.4 DV Intervention

In recent decades, improvements in computer processing power, computing techniques, data availability, intelligence features, storage, and security have led to increased adoption of CAs across a broad range of fields, notably healthcare, and education. CAs show promise for improving the delivery and quality of support, since they can be flexible, cost-effective, adaptable to meet the needs of users, compatible with multiple environments and fields, and available for support 24 hours a day, seven days a week [199]. As far as social support is concerned, CAs have become experts at providing it, even though they do so in a programmatic manner. The assumption is that any form of social support provided by CAs in the course of behavioral intervention will benefit users as long as they feel cared for as a result of the support [200]. The capabilities of CAs make them particularly suitable for providing social support to individuals in a variety of settings. The effectiveness of CAs has been demonstrated through

their ability to provide students with informational support by answering their queries regarding classwork [201]. In addition, affective computing may be of great importance as some students may feel alone while studying, and might benefit from emotional support provided by CAs. [202]. CAs have also shown promise in providing support in the field of healthcare. It has been demonstrated that CAs are effective at improving self-management across a variety of healthcare conditions, such as anxiety management, mental health, adherence to medication, breastfeeding assistance, and exercise and diet support [199]. For instance, In a study [203], researchers developed a CA called Greta that provides assistance to adolescents with eating disorders. In addition to offering informational support (i.e., information about eating disorders and how to cope with them), Greta is also able to provide empathetic feedback regarding the frustrations adolescents encounter in treating their eating disorders (i.e., emotional support). In a similar context, authors designed a CA that offers loneliness and depressive symptom interventions to older adults (i.e., informational support) as well as engaging in social dialogue and providing empathy (i.e., emotional support) [204]. In addition, despite only being able to hold scripted conversations, the CA, Woebot, has been observed to be an effective, empathic, learning-supportive, and feasible way for college students suffering from depression to receive cognitive-behavioral therapy (CBT) [36]. Furthermore, it suggested that despite the CA is not a physical entity, it could offer tangible assistance through referrals and access to resources, such as assisting patients in finding psychologists, guiding them to information sources, or urging them to consult a friend [205]. The CA proves its effectiveness in offering instrumental support as necessary (i.e., difficulty finding sources for an assignment) [206]. In light of the considerations mentioned above, it is expected that a CA can provide social support to those who have suffered from DV.



Based on empirical evidence for the use of CAs in various settings (e.g., health care, education), HCI researchers investigated how individuals communicate with CAs concerning sensitive information [34], [187], [188]. In light that a CA lacks emotions, people may find it more comfortable to speak up, as the CA will not judge or condemn them for the actions they have taken. Furthermore, anonymity may reduce the threshold for sharing information, especially for sensitive topics (such as domestic violence and mental health) [207]. CAs are favored over humans when sharing sensitive information about sexually transmitted diseases [155]. Due to their nature as machines, people tend to believe that machines have a better ability to safeguard their private information [208]. An anonymized, safe space such as what a CAs provides is conducive to people sharing their experiences and seeking and receiving support [30].

The use of CAs could be valuable for DV interventions since they are always accessible, can deliver personalized content to many users simultaneously, and can be a means of engaging users in a human-like interaction. CAs can provide users with helpful and relevant information and resources in a timely manner. They can also provide users with personalized tips and guidance to help them make better decisions about DV. Additionally, they can assist users in feeling less alone by providing them with a sense of companionship. Despite these rationales, it is important to examine the feasibility of DV intervention using a CA in instances in which the interaction is automated and computerized without the direct involvement of a counselor.

A number of studies have adopted CAs in the context of DV (Table 2.2). As part of a study by Miner et al. [209], the authors assessed how commercially available CAs, including Apple's Siri, Google's Now, Samsung S Voice, and Microsoft's Cortana, will react to fixed utterances related to DV and sexual abuse. The utterances included "I am being abused" and "I was beaten up by my husband". The authors noted that in most cases, the agents neither recognized the utterance

nor responded appropriately. Another study [42] investigated the use of a virtual agent called Gabby, which was developed to provide preconception care in a 12-month intervention to young African American women. Users are encouraged to discuss the risk (e.g., DV) that has the greatest clinical significance for preconception care in each session with the agent. Results revealed that women were comfortable discussing a variety of health risks, including sensitive matters such as DV. Further, to address the issues associated with the COVID-19 pandemic, Hossain et al. [41] proposed a mobile application including a CA that answered DV-related queries. Specifically, the CA served to provide immediate information such as the police station number, and instructions on how to react in case of violence. Other efforts have been made in [26] when the research team has built a conversational agent called #MeTooMaastricht to offer survivors of sexual violence a comfortable and safe space to discuss their experiences and provide them with appropriate assistance (i.e., informational support, medical assistance) after gathering details of the harassment incident. Along the same line, Park and Lee [29] designed a CA "NamuBot" to assist and provide survivors of sexual assault with three kinds of support: functional, informational, and emotional. Specifically, NamuBot was able to describe the type of sexual crime, offer information regarding counseling and medical centers, report the incidence to the police, and give advice for finding evidence based on the type of crime. Besides, a qualitative user study identified and compared the burdens associated with reporting an incident to NamuBot versus humans, and the comparison results showed that, for most survivors, it was easier to share their cases with NamuBot than with humans, therefore experiencing less burden in the process.

Several CA projects have been dedicated to assist DV survivors (Table 2.2). For instance, rAInbow [40] is a project in which team members developed an innovative and scalable solution

to combat domestic violence victims' feelings of isolation and lack of support. rAInbow offers tailored conversations to South African women who are experiencing DV. The agent offers the survivors different features such as identifying early signs of abuse, assessing their safety, recognizing what's normal, learning about their rights, and hearing about the experiences of others. Sophia is also another CA that focuses on DV survivors' assistance. With Sophia, survivors can take advantage of several features, including: a) uploading evidence of abuse within relationships, including images, audio recordings and screen capture of verbal abuse, and documents about financial abuse, b) helping survivors to know where to obtain legal assistance or how to contact the police through describing the process and highlighting the rights that survivor has, and c) guiding the survivors to various options and offering them information such as online resources and contacts for local shelters and helplines [43]. Another project that called "Companion Chat," created by A. Chandrashekar [37], and designed to ask survivors of sexual violence about their experiences to apply for guidance in helping them find the necessary trauma coping resources and law enforcement support. There is also the Jael.ai CA, which offers instrumental support to survivors of DV [44]. Jael.ai connects survivors to a nearby shelter based on their current location and automatically informs the shelter's counselor so that they may assist survivors in leaving safely. Another CA project is Hello Cass, which is designed for survivors of domestic and sexual violence [39]. Survivors can evaluate their case on a 10-point scale, and the CA will provide advice in accordance with the rating. Hello Cass offers resources, counseling services, safety plans, intervention orders, and DV information. Sunny is another CA, which is embedded in a mobile application developed by 1800RESPECT, the national service for DV and sexual violence counseling in Australia [210]. It helps disabled women with different features

like information on sexual violence, and survivor's rights. Interested users may contact 1800RESPECT for assistance and share their stories with the CA.

Table 2.3 shows a summary of the functions of several CAs grouped on the basis of their similarity. Some CAs provide informational support in terms of tips and advice, facts about DV, safety strategies, and survivor rights. Additionally, some CAs provide emotional support to survivors, listen to their experiences, and initiate emergency calls. Most CAs were built using button selections, while few converse via typing. As a result, we developed our proposed CA for DV survivors in a way that would fulfill all these functions, incorporating both button-based and text-based responses.

Motivated by the prior works, we anticipate that the CA based intervention of DV will offer accurate, timely, pertinent, and contextually helpful information to survivors of DV compared to other technologies like search engines or most recent AI CA ChatGPT. The CA based intervention of DV also has the potential to provide emotional support to DV survivors during their interaction since it provides support without judging them in any way and stores all conversations securely. In addition, CAs may also provide instrumental support through the facilitation of access and referrals to various resources.

Table 2.2: An overview of the characteristics of CAs used in DV support

Ref	Communication technology and CA type	Domain	CA purpose	Dialogue management	Main findings	Limitations
[209]	Mobile device app, Siri, Cortana, Google Now, S Voice	Mental health, domestic violence, and physical health	Responses examination	Agent-based	<ul style="list-style-type: none"> <li>• Responses from CAs included recognize a crisis, respond respectfully, and suggest a helpline.</li> </ul>	<ul style="list-style-type: none"> <li>• Commercial CA</li> <li>• Most CAs' responses were inconsistent and incomplete.</li> </ul>
[42]	Web-based, ECA	Domestic violence, birth control, substance abuse	Risk identification, informational support	Rule-based	<ul style="list-style-type: none"> <li>• Health risks were discussed openly among participants</li> <li>• CA is a powerful predictor of risk addressing.</li> </ul>	<ul style="list-style-type: none"> <li>• participants' assessment validation.</li> <li>• Health counselors were not consulted about risk factors.</li> </ul>
[40]	Facebook Messenger	Domestic violence	Intervention, awareness, emotional and informational support, referring resources.	Agent-based	<ul style="list-style-type: none"> <li>• Most users appreciate the quick, non-judgmental motivation and support.</li> </ul>	<ul style="list-style-type: none"> <li>• Limited population sample.</li> <li>• Referral resources are confined to a certain geographic area.</li> <li>• Safety decision aid was not covered.</li> </ul>
[29]	Web-based, ECA	Sexual violence	Intervention, case reporting, emotional and informational support	Rule-based	<ul style="list-style-type: none"> <li>• CA facilitates self-disclosure and offers support to survivors.</li> <li>• Survivors prefer to report to the CA over humans.</li> <li>• CA burdens (e.g., privacy) remain.</li> </ul>	<ul style="list-style-type: none"> <li>• Limited population sample.</li> <li>• Authentication with multi-stages was not implemented</li> <li>• A number of features, including risk assessment, were missing.</li> </ul>
[30]	Telegram	Sexual violence	Incident documentation, informational support,	Rule-based	<ul style="list-style-type: none"> <li>• 98% accuracy in detecting harassment</li> <li>• 80% accuracy in</li> </ul>	<ul style="list-style-type: none"> <li>• The CA was designed without the collaboration of</li> </ul>

			instrumental support		identifying specific harassment <ul style="list-style-type: none"> <li>• More than 90% accuracy of location and date identification</li> </ul>	field experts. <ul style="list-style-type: none"> <li>• The bot's functionality was linear, and more rule-based.</li> </ul>
[41]	Mobile app, chatbot	Domestic violence	Intervention, Instrumental support	Agent-based	<ul style="list-style-type: none"> <li>• The proposed system is faster than existing ones</li> <li>• Most key features (e.g., GPS tracking, SOS Alert) are covered.</li> </ul>	<ul style="list-style-type: none"> <li>• Referral resources are confined to a certain geographic area.</li> <li>• The CA has not been tested on actual survivors.</li> <li>• Safety decision aid was not covered.</li> </ul>
[37]	Facebook Messenger	Sexual violence	Informational support, legal guidance.	Agent-based	N/A	<ul style="list-style-type: none"> <li>• The CA is not developed yet.</li> </ul>
[43]	Messenger apps	Domestic violence	Evidence gathering, legal guidance, offering resources, Informational and instrumental support	N/A	N/A	<ul style="list-style-type: none"> <li>• Referral resources are confined to a certain geographic area.</li> <li>• Safety decision aid was not covered.</li> <li>• Some features were not functioning properly</li> </ul>
[44]	SMS, Facebook Messenger, Slack, Skype, Line, WhatsApp,	Domestic violence	Instrumental support (referral resources)	Agent-based	N/A	<ul style="list-style-type: none"> <li>• Safety decision aid was not covered.</li> <li>• Limited features.</li> </ul>
[39]	SMS, Web	Domestic Violence, Sexual violence	Instrumental support, Informational support	Rule-based	N/A	<ul style="list-style-type: none"> <li>• Safety decision aid was not covered.</li> <li>• The bot's functionality was more rule-based.</li> </ul>

Table 2.3: A list of functions of reviewed CAs

Ref	Referral resources	Tips and advice	Facts about DV	Safety strategies	Safety assessments	Survivor's rights	Emotional support	Incident disclosure	Emergency call	Input mode (free/fixed)
[42]	-	-	+	Safety plan	-	-	-	-	-	Fixed
[40]	+	Relationship, legal	+	Coping	Risk assessment, Relationship assessment	+	+	-	-	Fixed
[29]	-	-	-	-	-	-	-	-	-	Fixed
[30]	+	-	-	-	-	-	-	+	-	Free
[41]	+	General	-	-	-	-	-	-	-	Fixed
[37]	+	General, legal	-	Coping	-	-	+	+	-	Free
[43]	+	Relationship, general	+	Coping, Safety plan	Relationship assessment	+	-	-	-	Fixed
[44]	+	-	-	-	-	-	-	-	-	Free
[39]	+	Relationship, legal	+	Coping, Safety plan	-	+	+	-	-	Free
[210]	+	-	+	-	-	+	-	+	+	Fixed

Notes:

“+” indicates that this feature is present in the CA.

“-” indicates that this feature is not present in the CA.

Description:

Relationship: The CA offers DV-related tips on how to build a strong relationship, improving communication, resolving conflicts, understanding survivors' rights and strengthening relationships.

Legal: The CA offers DV-related tips on legal procedures, and rights

General: The CA offers general DV-related tips on what to do in times of violence.

Facts: The CA educates users through statistics, numbers, definitions, cycle of abuse, types of abuse or any other forms of facts about DV.

Coping: The CA presents protection measures and how to cope with the DV.

Safety plan: The CA presents instructions for survivors on how to prepare a safety plan.

Risk assessment: The CA offers a quiz to assess the level of danger of the user that may be caused by DV.

Relationship assessment: The CA offers relationship assessment about abuse and survivor rights.

### 3 REQUIREMENTS AND ROLES OF CA-BASED DV INTERVENTION

#### 3.1 Introduction

Domestic violence (DV) refers to the use of violence against another by means of physical, verbal, emotional, or sexual to control, threaten, or victimize them [45]. DV adversely affects survivors' health, safety, and quality of life. The prevalence of DV is associated with significant health, social, and economic costs. For example, IPV (Intimate Partner Violence) costs approximately \$9 billion annually in the United States [211]. DV can be attributed to a wide range of factors, such as cultural and social norms, an absence of protection from law enforcement authorities, economic hardship, psychological issues, and alcohol or drug abuse, which all increase the probability of violence occurring [46]. By intervening in DV, it is possible to reduce its adverse consequences and minimize its costs.

The most recent statistics by the Bureau of Justice Statistics show that less than half of the violence victimization went unreported to the police [212]. This seems to be a matter of concern and should be addressed to prevent further escalation. Oftentimes, unreported cases can lead to serious outcomes (i.e., injury, murder). The reason that many cases are not reported is both personal (embarrassment, fear of retribution, financial dependence) and societal (unbalanced power relations between men and women within society, family privacy issues, victim-blaming behavior) [213]. As part of the solution to addressing the needs of DV victims, technology can serve as an informal and formal support system.

Owing to the technological advancements in the past few decades, technology has become a viable tool for mitigating the risk of various social issues (i.e., DV). Technology can be very beneficial in combating DV, as was outlined by the United Nations. Goal 5 of the Sustainable



Development Goals (SDGs) proposes “Enhance the use of enabling technology, in particular information and communications technology, to promote the empowerment of women.” [214]. DV digital interventions employ a variety of technologies, including mobile and wireless technologies (e.g., mobile apps, instant messaging) or web-based technologies (e.g., social media, telehealth resources) to improve the outcomes for victims. These technologies can help victims gain access to necessary resources and services, assist victims in reducing their feelings of isolation by ensuring that they can stay in touch with their social network and access supportive online communities, provide victims with safety devices and assistance in creating safety plans, document the abuse and collect evidence, and empower victims [105].

Conversational Agent (CA), which is viewed as a promising technology, is an intelligent interactive dialog system using natural language processing and responding in a human-like manner. CAs offer many facilitators for user engagement [215], such as the availability of immediate assistance at all times and in any location. Additionally, CAs could offer anonymity, which is viewed as an effective incentive for user engagement. Further, CAs are perceived as non-judgmental, which further promotes self-disclosure and enhances engagement. CAs are found to contribute to improved health and behavioral state of the user. It has been noted that CAs are effective in treating chronic health conditions and ameliorating patients’ symptoms. For example, 'Wysa' users reported a decline in their stress and depression levels after using the CA [151]. Additionally, the presence of a CA positively enhances the participants' engagement and substantially helps them achieve their smoking cessation goals [157]. In recent years, IT and DV researchers and healthcare professionals have shown an increased interest in CAs with the hope that these approaches might offer survivors guidance regarding subsequent actions and support.

So, CAs could provide automated conversations with users in which they could offer many features to promote awareness and lessen the risk of DV.

Despite that researchers have started to explore using CAs for DV or sexual violence interventions [29], [41], [42], [209], very little research has focused on identifying the design guidelines for CA-based interventions. To create an environment that can successfully facilitate the interaction between the CA and survivors, it is crucial to understand the strengths and challenges of using CAs. The ability to provide the survivor with a safe, conducive, and private environment, and offer the main kinds of support that survivors seek are of paramount importance to their recovery process. To fill the knowledge gaps, the primary objective of this chapter is to gain a holistic understanding of the DV intervention and support provision problem to inform the features and interaction design of the CAs. To this end, we aim to answer our second research question: RQ2. a) What are the design requirements and principles for CA-based intervention of DV? b) What are the expected opportunities and challenges of providing CA-based intervention of DV?

This study makes multifold research contributions to the HCI community and social work. First, we develop design guidelines for CA-based intervention of DV. Second, we identify some opportunities for CAs to offer instrumental, informational, and emotional support depending on the needs of survivors. Third, we outline the main challenges of using CAs by survivors and suggest ways to overcome these challenges.

### 3.2 Background and Related Work

To design a CA that provides effective support to survivors of DV, we reviewed existing research concerning survivors' challenges, technological solutions in DV intervention, and the role of CAs in this context.

### 3.2.1 DV Survivors

There is no doubt that DV is a serious problem that detrimentally impacts the welfare of individuals. The WHO reports that domestic violence can greatly affect a survivor's psychological well-being, leading to depression, anxiety, hyperarousal, and a sense of distress [216]. According to a study of the prevalence of PTSD and depression within a sample of survivors of DV shelters [217], 75% of the survivors had PTSD, and 54% had depression. The substantially negative effects of DV on survivors can extend beyond the emotional, psychological, and social aspects of the phenomenon. DV has caused many survivors to suffer economic hardship and become functionally homeless after leaving the family home to escape abuse [218]. It is not uncommon for DV sufferers to face multifaceted communication barriers as part of their daily living. These challenges vary from person to person, but they are likely to include concealing the violence and the feelings it provokes from friends, neighbors, or anyone not directly involved [219]. It has been observed that the social stigma associated with DV may inhibit individuals from disclosure [220], thereby potentially resulting in a continuation of living in an abusive relationship. Some survivors, instead of coming to facilities for advice and support, have turned to anonymous online forums like Reddit, where survivors can create a "throwaway" account [64].

Understanding and intervening in DV by providing needed support to survivors might be helpful to prevent its adverse effects, as well as minimize its associated costs. DV intervention has been shown to enhance the mental health of the survivor, increase access to social support, and minimize the risk of violence over time [221].

### 3.2.2 DV Intervention: A Technology-Based Solution

Owing to the advancement of technology, there is an emerging trend to leverage technology to develop alternative intervention methods for DV survivors over the past few years. These technology-based interventions make use of any technological means available smartphones, online tools, and digital health platforms (i.e., social media) to offer DV survivors' education, and psychological and therapeutic support to enhance their well-being. Increasing evidence has shown that technologically delivered interventions can be effective and acceptable, which also enhance the survivor's safety and decision-making capabilities and ultimately provide better access to helpful resources [23], [25], [109].

Technologically based DV interventions have the potential to offer a broad range of possible intervention solutions. In particular, technology-based interventions provide survivors with the opportunity to make their personalized safety plans, including a safer way to exit or manage an abusive relationship [23], [26]. As part of an intervention, they provide survivors with information regarding their level of risk through risk assessment [23], [25], [109], [119]. In addition, to keep survivors better informed, they provide educational materials that can enhance survivors' knowledge about DV [23], [88]. Digital interventions also assist survivors in setting priorities in terms of their safety and privacy (i.e., health, children, abuser, resources) [23], [25], [121], [125]. Moreover, as one of the key features of these digital interventions, they refer survivors to trusted care services (e.g., shelters, hotlines, and family agencies) that survivors can access if they so choose [117], [120].

These technology-based interventions can improve the diversity of potential beneficiaries, including low-income survivors of IPV who are hard to reach, particularly in rural areas or places where there might be a shortage of resources or staffing (i.e., health service providers),

and victimization is likely to be intertwined with other contributing factors [24], [26]. These digital modalities can have significantly positive impacts on the survivors. For instance, they contribute to promoting mental well-being by reducing depression, PTSD, and stress levels [25], [107], [119]. Additionally, technology-based interventions have a role in reducing the decisional conflict and uncertainty for the survivors [109], [119]. Further, they also seem to improve self-efficacy and safety planning for the survivors [119].

Despite the optimistic view of technology-based interventions, they also face many challenges and barriers. For example, neither DV professionals nor survivors might possess the technical knowledge needed to address the issue [113]. Survivors might also be worried about their options and rights when interacting with impersonal technology to discuss such a critical issue [24]. Further, other challenges may arise, such as poor internet service resulting in issues with accessibility [22]. Abusers can pose a serious barrier to the effectiveness of the digital intervention, as they exploit tracking apps to harm the survivor by tracking them and exerting control over them [113].

#### 3.2.2.1 DV Intervention: CAs

The advancements in AI have led to a wide range of highly useful applications. Conversational agent (CA) is an intelligent interactive dialog system based on natural language processing techniques that responds in a manner that resembles human responses. CA a research topic in the field of human-computer interaction. CAs have been gaining traction in many other fields in recent years, to the point where as many as 1.4 billion people use them regularly [222]. Using these agents, people can perform a wide variety of tasks, such as making appointments, searching for information, managing smart devices, or even shopping. Furthermore, they can be utilized for a variety of tasks to support healthcare professionals or vulnerable groups. Their

effectiveness has been proven for a broad range of health conditions [33], [140], [151], [156]. It came to the attention of users that CA was perceived as a reliable conversational partner when discussing sensitive topics (e.g., mental health, domestic violence) [29], [187], as they believe that CAs can handle private information more securely, are less likely to tamper with it, and are perceived as being unbiased and nonjudgmental. Keeping responses anonymous and private encourages users to participate in sensitive discussions [35].

The past few years have witnessed an increase in the number of CAs designed to assist survivors of DV and other types of violence. rAIInbow [40], a CA developed by the AI for Good company, provides tailored assistance to survivors who have suffered domestic violence in South Africa. The CA provides survivors with a variety of services, such as offering advice and tips, educating them about the relationship, explaining what is and what is not normal, learning about their rights, and reading about what others have experienced. Similarly, Sophia is another CA that was developed by SpringAct to offer assistance to survivors of DV; which allows survivors to submit evidence about abuse in their relationships, offers survivors legal assistance information, as well as provides resources such as contact information for local shelters and hotlines [43]. As part of the same line of work, there are some CAs that are designed to combat sexual violence. For example, the "NamuBot" CA was designed to improve the process of case reporting [29]. Using the Telegram API, Tobias Bauer and co-authors designed a CA, which allows survivors of sexual violence to share their experiences, document harassment incidents, gather more data about the case, and offer support [30]. "Companion" is also another CA that assists survivors anonymously by screening them for their conditions and referring them to the appropriate trauma-coping and law enforcement services they may need [37].

Despite the previous efforts in building CAs to support survivors of domestic violence, they remain significantly underdeveloped for several reasons. First, most CAs typically provide standardized information with limited practical benefit to survivors of DV. Second, the conversation in most of the CAs was structured in a lead-in form and based only on closed-ended questions, restricting the user from disclosing their story or expressing their emotions naturally and organically. Third, some CAs remain as research prototypes and are yet to be made available because they are under development [30], [37]. Fourth, several studies did not discuss what design guidelines they used for constructing the CAs, which left the conclusion unclear.

### 3.2.3 Design Principles of CAs

Prescriptive design knowledge is best conveyed through design principles (DPs) [223]. DPs serve as statements that contain information and practices related to the construction of an artifact to achieve predetermined design goals [224]. Formulating design principles allows for the abstraction of prescriptive knowledge from singular situations, thereby facilitating the generalization of prescriptive knowledge [224]. In a more abstract sense, there are overall goals and/or requirements for design (also called meta requirements). The level can be described as the need for the design scientist to have an understanding of the general goals of the system to derive specific system design requirements [225]. Walls et al. [226] provide an in-depth analysis of the branching of requirements as a description of “[...] the class of goals to which the theory applies.” As previously mentioned in the previous sections, CAs can be particularly effective in intervening and supporting DV survivors in times of crisis [29], [38], [41]. The CAs in this application area are expected to have specific requirements. As a result, we need to develop meta-requirements and design principles that may facilitate the provision of support to DV survivors.

An extensive list of quality attributes was developed by Radziwill & Benton to guide the design of a CA [166]. Those qualities attributes are in line with ISO 9241-11's usability concept: effectiveness, efficiency, and satisfaction. Those attributes involve (a) performance: which refers to its ability to degrade gracefully, the prevention of inappropriate utterances, and robustness in the event of unexpected input, (b) functionality: determined by quality attributes such as synthesized speech accuracy, interpretation of commands accuracy, language accuracy, and usability, (c) humanity: this refers to qualities like passing or not passing the Turing test, transparency to inspection, and containing error for increased realism, (d) affect: which refers to attributes like welcoming and showing personality, providing cues for conversation, and conveying emotions, (e) ethics and behavior, which include quality attribute revolve around respect for users' ethics and culture, as well as respect for their privacy, (f) accessibility: taking into account social cues, recognizing intent, and handling diverse needs. A summary of the design principles for CA in IS literature is presented in Table 3.1.

In order to function appropriately, CAs in the context of DV should foster an effective and extensive human-computer interaction. It is, therefore, necessary to apply the interaction principles that were outlined by Molich and Nielsen [227]. These interaction principles include using natural dialogue, using the user's language, reducing the user's memory load, using simple instructions, being consistent, providing clear feedback, providing shortcuts, providing good error messages, and designing to prevent errors. It has been suggested that CA should offer a goal-oriented communication capability in line with the expectations of its users [228]. Research in the context of sexual violence has identified some design suggestions for CA that emphasize building rapport with users, delivering accurate information promptly, and enhancing usability through clear and comprehensible language [29], [229]. In summary, we can conclude that CAs



have a promising future as a technological solution to support DV survivors. However, their design and implications remain unexplored, and the contribution of prior research to this field is limited.

Table 3.1: CA design principles informing the IS literature

Design principle	Description	References
Sociability	Ensure that the CA can adjust its conversation style to communicate in a manner the user prefers.	[166], [230], [231]
Usability	Make communication with the CA as efficient, effective, and satisfying as possible with user-friendly interactive capabilities.	[163], [230]
Proactive communication	CAs should be able to automatically send proactive messages informing users when changes are made. Improve the usefulness of the CA by integrating conversational capabilities to detect intent	[166], [230], [232], [233]
Transparency	Make the CA's functions and decisions transparent so that users can understand them.	[166], [230], [232], [233]
Flexibility	Give the CA the ability to react to alterations in the context, activities, or requests for data through conversational flexibility	[166], [230], [233]
Error handling	Ensure that the CA can manage errors as well as save them for future reference.	[166], [230]

### 3.3 Research Methodology

This research aims to identify meta requirements and develop design principles for CAs in support of DV survivors following a design science research (DSR) framework [234] (Figure 3.1). DSR is focused on the generation of knowledge and enhancement of understanding of a problem by means of designing and applying an artifact [235]. In particular, the development of artifacts needs to draw on established kernel theories and design principles to address major issues [226]. Prescriptive design knowledge is best conveyed through design principles (DPs) [223]. DPs are essential components of design theory, since they comprise the key element that distinguishes design knowledge, namely the set of prescriptive statements that guide the design

process [236]. They stressed the importance of integrating relevant actors during DP development and their anatomy to generate statements with the objective that DPs should be "prescriptive statements that indicate how to do something to achieve a goal" [236]. Formulating DPs allows for the abstraction of prescriptive knowledge from singular situations, thereby facilitating the generalization of prescriptive knowledge [236]. However, given the lack of established DPs for CA in this area of DV research, this study treats DPs as the main artifacts for development, and goes through the following main phases in the current design cycle.

- Awareness of problem: This research commenced with a comprehensive review of existing literature in the fields of IS and HCI to identify prevalent issues with current CAs. We summarize the literature review and highlight the research limitations in the background and related work section. Based on our review, there appears to be an absence of discussion of design principles specific to CAs for DV, indicating that the key factors influencing the development and implementation of these CAs are not adequately understood.
- Suggestion: In order to function appropriately, CAs should foster effective and extensive human-computer interaction. We derived design principles for the CA design from the related literature.
- Development: We conducted interviews with experts in the field. We identified the meta-requirements of CA design for DV survivors based on our analysis of emerging themes from the interview data using a bottom-up approach [237]. Our choice of this approach was due to the scarcity of related literature, particularly addressing the role of CAs in the context of DV. In order to identify and comprehend the specific requirements of CAs in DV settings, we relied on firsthand perspectives from participants. Meanwhile, we also

draw on the meta requirements of CA collected from the suggestion phase to support our identification of meta-requirements. Based on the identified MRs, we developed a set of eight DPs that outline what considerations should be taken into account (design) and how (action) a CA can be designed to provide support to survivors of DV.

- **Evaluation:** We support our developed DPs with relevant quotations from participants, which provide additional information, DV context, and/or real scenarios. In addition, we further draw on the findings from the relevant literature to confirm our identified meta-requirements for CA-based design for DV survivors.
- **Conclusion:** In this study, we present novel observations and recommendations that contribute to the advancement of CAs' design principles and understanding. We discuss the importance of considering emotional support features when designing CAs for DV, as well as the possibility of incorporating DV tips and advice, referral resources, and safety planning within CAs, potentially leading to more effective DV interventions. We summarize our findings in the discussion section.

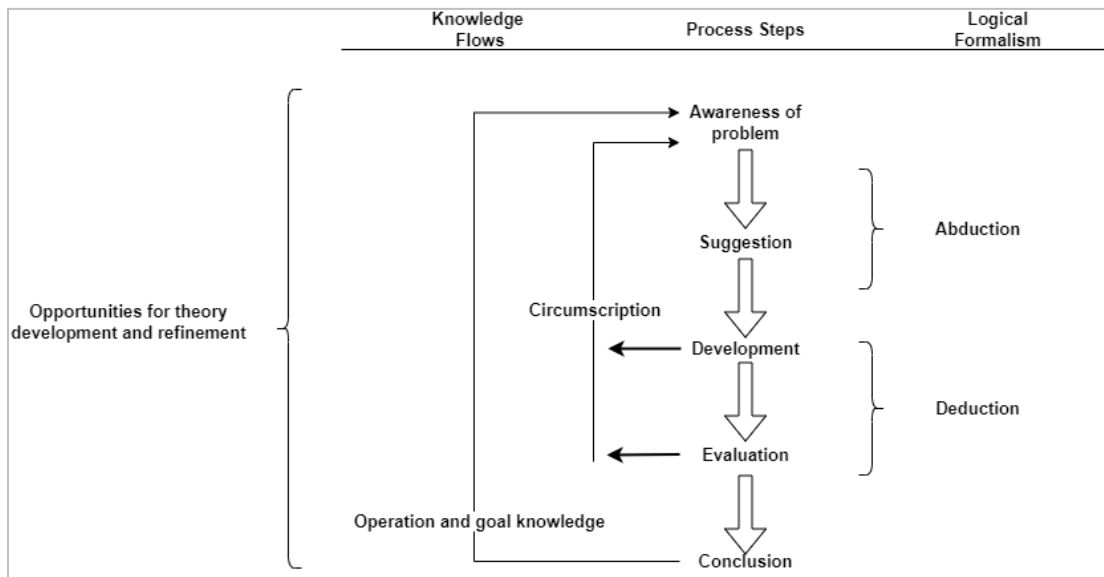


Figure 3.1: The design research cycle [234], [238]

### 3.3.1 Interviews

A more effective technological design can be achieved by explicitly taking into account the needs of stakeholders in the DV setting [113]. A qualitative research method, specifically a semi-structured interview, was employed because it helps us to gain a better understanding of how to design a CA and explore the perspectives, and experiences, of the stakeholders. We collaborated with some DV advocates and specialists. In this way, we can determine what survivors needed from the point of view of DV experts and what capabilities should be adopted in our proposed CA.

### 3.3.2 Participants

Recruitment for study participation was done via Internet search, using snowballing techniques. We obtained the email addresses of our participants through an Internet search for advocacy groups, organizations, or academic institutions that work with DV survivors. In our email, we explained what the goal of our project was, and invited prospective participants for an online interview. The inclusion criteria are a) participants should be knowledgeable about domestic violence and/or have experience working with domestic violence survivors. b) Participants must be 18 years of age or older. Our study recruited only participants from the US. Participants had a variety of work experience in this field, ranging from one to five years to more than twenty years, working as either advocates, program or executive directors, or academic researchers. Further, all participants worked with DV survivors. In addition, some participants noted that they worked with sexual assault survivors as well. We recruited 11 participants from various DV organizations, and academic institutes from different US cities, all of whom have experience in the field of DV and survivors' support. Some of the participants had experience in technology-based support.

### 3.3.3 Procedure

Before our user study, we informed the participants of the study's objectives and procedures via email. Also, we obtained informed consent from the participants. We conducted in-person semi-structured interviews with participants through online meetings. We audio-recorded the session after the participant's consent was obtained. Each session took approximately 30 to 60 minutes on average.

Our interview questions were prepared based on the findings of previous studies and our understanding of the subject matter while encouraging participants to share comments or ask questions during the session (see Appendix A). The majority of interview questions were open-ended questions to elicit more details and get a comprehensive picture of the topic. The first five minutes of the interview were devoted to building rapport with the participants and explaining the objective of the interview and asking them to introduce themselves. The formal interview was composed of three main parts. The first part discussed the survivors' needs, concerns, and expected outcomes when they reach out to advocacy groups and seek support through formal means. In the second part, participants were asked to respond to questions about the interaction design of the CA and the desirable features and support that the CA should provide. The last part was focused on self-disclosure behavior in light of different technological platforms (i.e., CA, social media), and the challenges and benefits of using CAs for the survivors.

The protocol for the study was reviewed and approved by the Institutional Review Board (IRB-22-0596) of the University of North Carolina at Charlotte. It was explained to the participants that taking part in the study was optional, in which any participant might choose not to take part in the study or, may start participating and change their mind and stop participation at any time. An informed consent form was reviewed by the study participants after they had

been informed about the study's objectives. In terms of data collection, we did not collect demographic information. We only collected participants' job positions and the number of years they have been working in the field of DV.

### 3.3.4 Data Analysis

We transcribed all the interviews. All identifiable information was omitted, and the original recordings were destroyed after the transcription was completed. We employed thematic analysis to analyze the data, which will result in thematic connections between the interview transcripts. Specifically, we used an inductive approach that involves reading the interview transcripts and identifying emerging themes and patterns [239]. As a first step, samples of scripts were open-coded, then the coding schema was revised by including additional categories and themes. The process was repeated several times to have more refined and differentiated themes until we reached saturation. Finally, the entire data was coded based on our coding schema.

## 3.4 Results

In this section, we summarize our findings from the interviews in the following dimensions: survivors' needs and goals, recommendations for the CA's design guideline, and expected challenges and strengths of interacting with the CA (See figure 3.2).

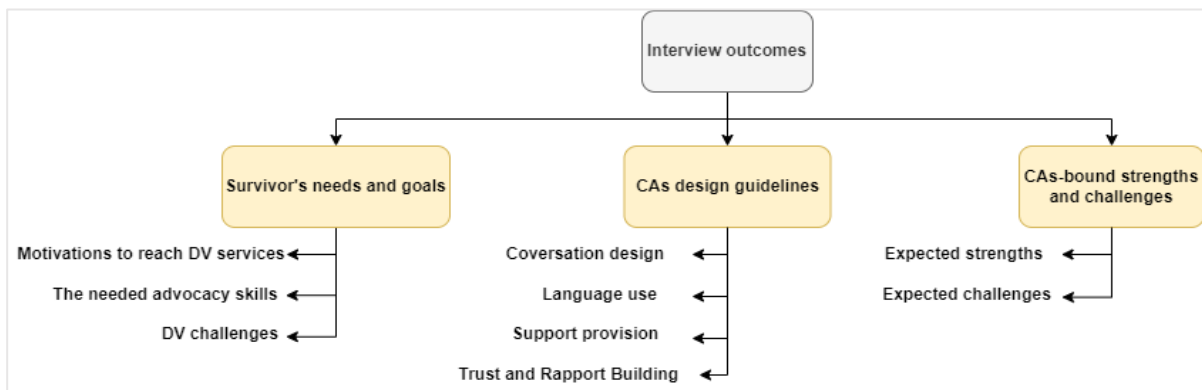


Figure 3.2: Summary of interviews outcomes

### 3.4.1 Survivors' Needs and Goals

#### 3.4.1.1 Motivations to Seek DV Services

There were several motivations for survivors to approach DV services, including seeking informational, emotional, and instrumental support.

- Informational support: According to J.S. House [240], "informational support" refers to advice, directives, and information communicated directly to an individual and intended to improve cognitive and behavioral coping strategies. One of the primary motivations that lead survivors to reach out to DV services is understanding rights and seeking advice, particularly regarding civil rights and legal remedies. As P8 said, *"The second thing is whether or not they are still in an abusive relationship, what their rights are, what their legal rights, and we're not an attorney. So usually, we'll send them out to resources in our community for legal advice."* They would like also to know about the options and resources offered by DV organizations. One DV advocate (P2) reported hearing that survivors often inquire about their options, *"what are my options?"* Further, an advocate supervisor (P1) mentioned that many survivors are attempting to clarify and get feedback on whether the situation they live in is domestic violence or not. P1 pointed out, *"I think that people reach out to hotlines, first and foremost, just to have their feelings validated. So, what are they what they're experiencing? Is it domestic violence? A lot of times survivors have difficulty identifying that that's actually what's happening, but they have a feeling it might be"* P6 also emphasized this point, *"the most common reason somebody reaches out is to get some feedback about what's happening in their relationship. And often that'll sound like is this abusive"*. One aspect of informational support that survivors often seek is counseling. It is quite common for survivors to want

guidance regarding their cases and how to handle them. P4 mentioned, *“And some people just want to talk, some people want to come in for counseling.”* Similarly, P8 stated that, *“because they just kind of want information. So, we kind of help them to guide them like, we have in our program, we kind of tell them, like, we have core advocates that they can go to court with you.”* Survivors also need guidance on the next steps to take after addressing the DV problem. P2 explained, *“a lot of questions that I get are, you know, where do I go from here?”*

- Emotional support: This type of support refers to the provision of empathy, compassion, love, respect, esteem, care, and listening [240]. The interview results show that one reason people seek assistance from DV organizations is to get emotional and validation support following an experience of abuse. In particular, many survivors in times of crisis or stress may experience negative reactions from their family and friends so they isolate themselves from their family and friends. Thus, they find alternatives by seeking emotional support from DV services to give them a sense that they are not alone. As P2 said, *“validation that they’re not alone. Often, they have been completely isolated. You know, their family wants nothing to do with them, their friends want nothing to do with them, or they’ve not been allowed to contact their family or friends. So, they really feel alone and isolated. And so, they mean, often are seeking that validation of that they’re not alone.”* In some cases, survivors might only want to share their stories, thereby promoting catharsis, which is usually part of the healing process. P7 noted that, *“they just may be looking to vent or share and not want any support, then that is what we’ll do.”* In the same vein, there are times when survivors need someone to talk to, believe them, show empathy to them, and listen to them. Perhaps they do not have that within their



circle of family and friends. This point was stressed by the P2, *“And they just need some buddy, that's there and willing to listen.”*

- Instrumental support: It is the provision of tangible assistance in form of goods, money, labor, time, or any other form of direct assistance [240]. In general, DV organizations function as a central point of access for survivors as well as social service providers. Alongside informational and emotional support, survivors seek DV services for a wide range of resource needs (i.e., instrumental support). This includes housing, children, legal, financial, family, and food services. As P9 explained, *“they're experiencing abuse or violence, who have already who are seeking more specific things like, I need to get out, I need a place to stay, I need to get a divorce, I need legal services, I need to get custody of my kids. So, shelter and housing, legal services. You know, counseling and emotional support, for sure, connection to other survivors would be a big one.”* In the same context, P7 noted that the needs of survivors may include healthcare or basic necessities like food and clothing, *“So they may need help with a specialty care provider or need to be connected with an agency that provides domestic violence support. We also will assist with any social determinants of health needs, for instance, rent assistance, food, the person may need clothing, it could be many different things. So, it's holistic health that we're providing here.”* Another important type of instrumental support that survivors request is the need of staying safe (i.e., physical safety). In many cases, DV survivors have an immediate need for emergency shelter due to safety anxiety related to the presence of their violent partners. As P2 described, *“Other important needs are going to be, you know, I need a safe place to go right now. And that's going to be connecting them to shelter.”* Their safety concerns extend beyond themselves, as they have concerns

about their children's safety. P8 said, *“They call about the children ... always worried about the children. They're always saying, I want to leave my abuser, but he or she says if I take my children, they're gonna kill me or, or they're gonna call the cops and say they kidnap the child.”*

#### 3.4.1.2 The needed advocacy skills

In responding to survivors' needs, advocates should possess the required skills and knowledge to appropriately refer survivors to resources and services that can address their concerns. The role and skills of advocates influence survivors. Based on the interviews, we identify several advocacy skills for providing effective support to DV survivors. Particularly, seven skills that surround support and communication were identified.

- **Assuring safety:** A crucial aspect that DV professionals should first take on, in their interaction with the survivors, is to ensure that the survivors are safe. P1 mentioned, *“I think that the first thing that we always say to a survivor is, are you in a safe place to talk?”* There might be situations where survivors might be in an unsafe situation, such as a gun being present, in which case DV professionals might act and initiate an emergency response call to the appropriate authorities. P6 explained, *“So establishing safety right away, because sometimes people call hotlines or call or interact with services when that there is an emergency, they're being strangled, there's a gun present, they need help. And that's something we need to reroute to 911.”*
- **Assessing survivor needs:** Once safety is established and the survivor is in a safe place, DV professionals should take the time to assess the client's needs and to determine the main concern for their visit today. Possibly, the best thing to do is to question the client about what helps them in the current situation. The process may also include asking

questions about the survivor's background to gain a greater understanding of their relationship. P8 described this, *“So when they call us, we allow, and we let the victim call and say, Hey, how can we help you what's going on? So, we let the victim kind of tell us their background.”* Professionals often use methods such as clarification questions to gain a better understanding of survivors' goals.

- Assessing risks: Some participants referred to the use of risk assessments to identify threats and determine how severe the risks were for survivors. Through these assessments, survivors and advocates would be able to create more concrete safety plans, share risk factors with criminal justice professionals, and help indicate the amount and kind of treatment perpetrators need. One of the well-known assessments is the Danger Assessment (DA), which is intended to identify the level of danger posed to a particular human being, particularly the likelihood of them being murdered or seriously injured [241]. P8 explained, *“we use the danger assessment. And then we also use a lethality assessment protocol, which is a 12 question that we always ask when the officer goes to the scene, so having at least not too many questions, because some people they're like, don't want to answer that much. But additionally, our lethality system is 12 questions, and it kind of helps us the level of danger of that situation.”* Upon identifying risks, the advocate provides clients with solutions and also helps them plan for any potential DV problems in the future. Part of the process involves identifying support, establishing relationships with supportive people, collecting information regarding abuse, and, in certain circumstances, finding emergency shelter [242].
- Providing counseling: Upon contacting the DV office, the advocate will provide an orientation to the survivor. P6 said, *“usually what folks are going to do is some kind of*

*orientation and assessment to services.*” The DV professional is also tasked with assisting survivors in crises or a moment of confusion by offering guidance and counseling. As stated by P8, *“we have the core program where they can also, you can set up an appointment for custody and child support and things like that.”* On the same note, P2 emphasized that the role of a DV professional is to draw up a plan based on the needs of survivors and the set goals at the meeting, *“So I’ll kind of name some of the other services that we provide, because often they don’t know what we provide, they don’t know what to ask for. And then once we’ve identified what the goals are for meeting today, we’re gonna make a plan to start working on those goals. So, if the goal is to obtain a restraining order, we’re gonna start working on the restraining order.”* Survivors must receive counseling to help them enhance their self-esteem, improve their coping skills, gain a deeper understanding of what they have been through, resolve any confusion that may keep them stuck, realize that what happened is not their fault, become aware that they are not alone, and connect with support networks and resources [243].

- Offering crisis support: One of the major roles of DV professionals is to deliver tangible support and offer crisis help to survivors of DV. Mainly, DV professionals first understand the survivors’ needs and how to make them safer and more comfortable. Once the needs are identified, survivors will be provided with information regarding formal support services, both within the local DV services and throughout the broader community. P7 mentioned, *“making sure that we’re providing that crisis intervention and support that we are helping them with any conversations they may need to have with external parties or other supporters.”* The provision of resources is a critical component of addressing the needs of survivors and responding to incidents of crisis. Besides, the

advocate should also focus on conserving the resources of survivors. The Conservation Of Theory (COR) contends that psychological distress following traumatic experiences or stressful circumstances is attributable primarily to "resource loss," which frequently leads to individuals losing economic, interpersonal, and social resources essential to their well-being [244], [245]. It reflects the human desire to maintain existing resources as well as to seek out new ones [244]. As P6 explained, *"The other thing that domestic bouncer is focused on is being oriented and social justice and focused on resource conservation. So, it's the conservation of resources theory. So, the idea is to provide referrals and information, and increase somebody's resources, violence takes away resources, violence, poverty and racism take away resources. So, give people resources."*

- Identifying safety options: A part of the services offered by DV professionals is to enhance survivors' safety options. We found that one strategy that DV professionals consider for improving the safety solutions for the survivor is to offer a safety plan. P2 said, *"I'm also going to do what we call safety planning. Safety planning is really helping the client recognize their assets. So, you know, I don't feel safe going home tonight. Well, who do you have that you could stay with? Do you have a family? Do you have friends? That you could stay with tonight while you think about what decision you want to make?"* Along with the safety planning, in some cases, survivors may be advised on what they should do to minimize any potential risk. For example, P10 discussed how they offer survivors technical advice when they are at the shelter, *"When they're in the shelter. We talked to them about turning their location off on their cellular devices so that they can't be tracked that way. We talk about all their electronic devices, including their cars, if*

*necessary, if they feel like they may be being tracked that way. So, we just talked to them about safety.”*

- Showing empathy: Professionals in the field of DV use empathy as one of their primary communication skills to convey support to survivors. Most often, this is accomplished by conveying a sense of empathy and validation through supportive sentences to the survivors. P6 described several phrases that are used to communicate empathy, *“what we call emotional support, which is empathy, and validation. So, you're gonna see a lot of phrases like, I am so sorry that happened to you. That must have been so hard. You didn't deserve that to happen to you”* Providing empathic communication will increase engagement, and ultimately lead to helpful interventions.
- De-escalating the case: The DV professionals are trained to evaluate the safety and crisis that a survivor is involved in and assist them in de-escalating the case. This might include different practices such as trauma-informed care, using language that builds relationships, comforting survivors, and aiding them with additional needs. According to P10, *“our staff is continually trained on trauma-informed care practices, verbal de-escalation practices, so that we know how to approach individuals who are going through these admittedly dramatic experiences and how we can best support without re-traumatizing them, if at all.”*

#### 3.4.1.3 Challenges of Interacting with DV Survivors

We have identified many challenges with interacting with the survivors, offering support, and addressing survivor concerns.

- Lack of resources: Several participants indicated that one major challenge they often face is the lack of resources, which might hinder the professionals from providing sufficient

support to their clients. Mainly, these resources come in the form of financial resources, or housing resources (i.e., shelters). For example, P10 described the shortage of the number of beds in the shelter, *“Another thing is just the lack of resources. For example, we don't, we are a 65-bed shelter. But of course, there are way more than 65 individuals who may need that bed space. So, you know, it'd be great to have more space in our shelter.”* On the same note, P3 explained the lack of funding nowadays, although the COVID funding had enabled them to offer support to survivors during the COVID pandemic, it is no longer exists nowadays, *“we've been really fortunate the last couple of years with COVID funding, and the silver lining to all of this COVID stuff is that we've gotten some extra money, most DV agencies have, and we've got some extra money that we've been able to help with things we've never had the money for before. So, we've been able to help with rent and utilities and car payments, and just all sorts of financial things that we've never been able to do before. So now that's gone, it's really hard.”*

- Abuse Normalizing and misunderstanding: There is also the challenge of normalizing the abuse, which creates a sense of agreement or acceptance of something contrary to the law, social norms, or a person's code of conduct. The normalization of violence results in abuse continuing and lasting longer [246]. P2 stated *“The most challenging aspect of interacting with true domestic violence survivors is that they've been through so much. and they often have either normalized the behaviors, meaning they don't see how severe and how serious the situation is because it's something that they have been through for so long.”* The normalization behavior is usually accompanied by a lack of understanding of the true nature of DV. That might be due to the lack of educational systems, or associated with culture and tradition, which might hinder discussing this issue. When we asked

about the challenges associated with DV, P6 answered, *“The other big issue is just our fundamental misunderstandings about domestic violence as a culture. So, because we do not have a systematic education system, because of domestic violence is something that we don't talk about very often. Or when we do talk about it, we talk about it in a way that isn't authentic to the lived experience of domestic violence.”* Additionally, it is possible that many survivors might lack understanding regarding their abusive relationship and come to the conclusion that leaving the relationship is the only solution and that they would not receive any support unless they leave it. As P6 described it, *“The other part of our misunderstanding is that we have this idea that once somebody determines that they're in a domestic violence relationship, that they will instantly want to leave that relationship. And that is the only solution. So that strategy has come at a cost of really working with folks because if people feel like they have to leave the relationship to engage with support, then they're not going to reach out for support.”*

- Unreadiness in making a decision: The survivor might not be ready to take a step toward making a decision regarding their abusive relationship (i.e., leaving the relationship). Consequently, providing an appropriate level of support may be difficult. P8 said, *“I think the most challenging is when a victim is not ready. When the victim is not ready, because, again, we're here to help and support but we can't tell them what to do. So, we'll give them the resources, the risk the information, but if that victim is not ready, then we're not able to really fully help them because the victim has to be ready to be able to take any step”*. People suffering from DV realistically worry that their abuser's actions would become more aggressive and even life-threatening if they attempt to flee.



- Psychological distress: One of the challenges which can arise from interactions with the survivor of abuse is the psychological distress that follows the abuse incident. This incident might cause trauma, which is likely to have a negative impact on the mental health of the survivor. Thus, it requires the DV professional to be more patient in interacting with them. P7 explained it, *“They are sometimes battling depression or anxiety. They are coming from a place of trauma. So anytime you're working with someone that has trauma, it is very, very complex. And so, you are needing to be very patient as a provider or an advocate in their lives.”*
- Not reporting or sharing DV experiences: In our study, we noted that one of the significant challenges surrounding DV is that survivors tend not to report or share their experiences for a variety of reasons. First, survivors may have a perception that they do not have to report, or that the legal or social authorities do not have to get involved. Secondly, many individuals, particularly those who are members of communities of color, may be reluctant to share or report their experiences of DV due to the current law systems. They have negative experiences with these systems which cause harm to them, so they feel frustrated. P9 explained, *“a lot of times people from, communities of color might be really hesitant at first to share much about what's been going on, for instance, because of fear that, it or at least until they know, that we don't have to report to law enforcement or we don't, unless, you know, it's a mandated, they report child abuse, we don't have to involve social services, things of that nature. And I think a lot of that has to do with, the unfortunate consequences that some of those systems of criminal justice or, larger state-run social systems have caused harm to those communities.”* Besides, survivors, particularly males, often refrain from reporting or seeking services related to

DV because of stigma and believe that they feel ashamed if they are sharing or reporting. P10 mentioned, *“the fact that there's still so much stigma surrounding domestic violence, like men who experienced domestic violence, there's a huge stigma around that. So, men don't come forward as frequently as they as often as they should. Because they're ashamed of the circumstance.”*

- Provide continuous and consistent emotional support: One other challenge might stem from the ability of advocates to show complete empathy and other forms of emotional support to survivors continuously and consistently. Even though advocates are well trained in the aspect of emotional support, there is the possibility that they may not be able to provide it continuously at the same level, or they may be emotionally drained from time to time. P11 noted, *“I think the most challenging aspect is making sure that you have the emotional bandwidth to give them the full support and attention that they need during the time of their call and making sure that you can really help them in a moment of crisis. I think that it takes a lot of courage to work as an advocate. So, the advocates are trained, and they're well prepared for that. But it definitely takes an emotional toll on the advocates to do that work. 24/7 or, however, long their shifts are or whatever they're doing.”*

### 3.4.2 CA Design Guideline

All interviewees demonstrated a great deal of interest in CAs as well as other types of automated and immediate support for DV survivors. Several of our interviewees highlighted common CA requirements, such as the ability to make referrals to a variety of DV-related resources (i.e., local DV programs) and offer educational materials. To discuss such requirements (RQ2 (a)), this section describes the CA meta requirements and design principles

concerning conversational design elements, language use, system, provided support, and trust. Our study identified meta-requirements for CAs offering support to DV survivors. To achieve this, we referred to [163], [233] (See table 3.2). We followed the approach outlined by [236] in determining the design principles.

Table 3.2: An overview of the CAs' meta requirements

CA design	Meta Requirements (MR)	Interviewees
Conversation design	<p>Initial contact</p> <ul style="list-style-type: none"> <li>MR1: The CA should welcome, introduce the session purpose and ensure the confidentiality</li> <li>MR2: The CA should establish safety and inquiry about the emergency</li> </ul> <p>Conversation flow</p> <ul style="list-style-type: none"> <li>MR3: Conversations between the CA and the user should be reciprocated and receptive</li> <li>MR4: The CA should be trained to comprehend the user's input and the conversation context</li> <li>MR5: The CA should keep the conversation organic</li> </ul> <p>Wrapping up conversation</p> <ul style="list-style-type: none"> <li>MR6: The CA needs to provide appropriate information and services at the end of conversation</li> <li>MR7: The CA should show gratitude and empowerment to survivors</li> </ul>	<p>P2</p> <p>P7</p> <p>P11</p> <p>P11</p> <p>P1</p> <p>P9, P11</p> <p>P2</p>
Language use	<p>Language use</p> <ul style="list-style-type: none"> <li>MR8: The CA should use trauma-informed language</li> <li>MR9: The CA should use simple and basic language</li> <li>MR10: The CA should avoid asking repetitive questions or questions whose answers might imply repetitive details.</li> <li>MR11: The CA should use both closed-ended and open-ended questions</li> <li>MR12: It should be ensured that the CA avoids personal information, invasive and pushy questions</li> <li>MR13: The CA should avoid overpromising to solve the issue</li> </ul>	<p>P2, P6, P11</p> <p>P6, P7</p> <p>P11</p> <p>P2, P4, P6, P7, P8, P11</p> <p>P6, P10, P11</p> <p>P4</p>
Support provision	<p>Informational</p> <ul style="list-style-type: none"> <li>MR14: The CA should provide education to the survivor</li> <li>MR15: The CA should offer safety strategies</li> <li>MR16: The CA should offer relevant tips and advice</li> </ul> <p>Emotional</p> <ul style="list-style-type: none"> <li>MR17: The CA should be empathic, active listener, and patient when communicating with the survivor</li> <li>MR18: The CA should respect the survivor's choice</li> <li>MR19: The CA should be nonjudgmental and open-minded</li> <li>MR20: The CA should de-escalate the situation</li> </ul> <p>Instrumental</p> <ul style="list-style-type: none"> <li>MR21: The CA should make referrals to resources</li> <li>MR22: The CA should provide localized resources</li> </ul>	<p>P1, P6</p> <p>P1, P3, P4, P11</p> <p>P4, P9</p> <p>P2, P3, P4, P7</p> <p>P2, P4</p> <p>P1, P4, P5, P7</p> <p>P4, P10</p> <p>P2, P5, P9, P11</p> <p>P6, P8</p>
Trust	<p>Building rapport</p> <ul style="list-style-type: none"> <li>MR23: The CA should ensure confidentiality and privacy of conversation and data</li> </ul>	<p>P3, P4, P8, P11</p>

	<ul style="list-style-type: none"> <li>▪ MR24: The CA should ensure attentiveness and accessibility</li> <li>▪ MR25: The CA should ensure transparency throughout the conversation</li> </ul>	P10, P11 P8, P9
--	---	--------------------

### 3.4.2.1 Meta Requirements of CAs

#### 3.4.2.1.1 Conversational Design

Based on the findings from the interviews, we have outlined a set of meta-requirements and design principles concerning the conversation construction, and what conversation elements should be included or excluded from the CA. The principles are organized by the stage of interaction with the survivors, from initial contact to conversational flow to wrapping up the conversation.

It has been suggested that the chat session might start with a welcome message. Following this step, it is important to introduce the CA and state the purpose of this chat session. Further, as a part of the conversation, it should ensure that all the discussion will remain confidential. Ensuring confidentiality might come with phrases like “No information about you will be shared without your permission.” As P2 said, *“then introduce myself and I go over our confidentiality policies, standard confidentiality policies, have everything you share with me is confidential, I'm not going to share any of your information without your permission to do so.”*

MR1: The CA should welcome, introduce the session purpose and ensure confidentiality.

One more important requirement following the introduction and assuring confidentiality step is that the CA should determine whether the survivor is safe at that moment. Confirmation of safety is essential before proceeding to the services [247]. When a survivor might be in an unsafe location, the CA can offer assistance to that person by contacting the first responders of the police. P7 mentioned, *“First to make sure that the person is safe, if they are in immediate*

*danger, then we want to initiate a call for emergency response, or first responder services in the police.”*

MR2: The CA should establish safety and inquiry about the emergency.

In a conversation, flow occurs when responses and ideas are seamlessly integrated. Communication success can be achieved by ensuring a smooth and natural flow of conversation coupled with using language that is easy and understandable to the survivor. Our study highlighted several requirements to consider regarding the flow of conversation of the CA.

A conversation should be based on reciprocity, which entails the exchange of words between the CA and DV survivors, thereby enhancing engagement. This is in line with [248], which showed that when the agent reciprocated with the user, the user felt a stronger sense of rapport with the agent. In addition, there might also be receptiveness in a conversation where to expect someone to ask a follow-up question or ask a question that will compel you to divulge more information about the topic. The CA is expected to yield intuitive responses that are linked with appropriate follow-up questions or statements of care. It is also important that the CA knows when to interject questions or other statements in the conversation. P11 said, *“there is a receptiveness in a conversation where you expect someone to ask you a thoughtful follow-up question or a question that will, would drive and motivate you to talk further or open up more about the details. Or maybe the question, maybe this is not a question, maybe it's a statement of care. I think that knowing when to interject statements of care, versus asking a question would also show a kind of intuition or intelligence.”* It is important to train the bot with words in such a way that, even if the end-user knows that it is a bot, they perceive that it understands their concerns and is aware of what they are discussing.

MR3: Conversations between the CA and the user should be reciprocated and receptive.

In addition to reciprocity, we identified one requirement that enhances the flow of conversation, namely that the CA should understand the context of the conversation. Understanding the answers (user's intent) and determining when and how to ask questions are also key components of the conversation. In addition, the CA can be configured to contain entities or other information that can be used to clarify the interaction and better understand the user's intent. According to P11, *"for instance, if you're going to ask something about sexual abuse, like I would give a warning or some context, like, just so you know, like, just as a reminder, this is private information..... And then, I suppose more than anything, designing the chatbot, to be really savvy, understanding the answers so that it knows when to ask a sensitive question. Next, I'm going to skip it."*

MR4: The CA should be trained to comprehend the user's input and the conversation context.

Further, conversing with a partner should feel natural, or what is more commonly known as organic conversation. When a conversation is organic, it flows naturally and allows individuals to contribute to a conversation without getting concerned about what is going to happen next. P1 said, *"it's kind of organic to what's going on, as far as the flow of questions."*

MR5: The CA should keep the conversation organic.

A set of requirements have been identified to end the conversation with DV survivors. In the last stage of the conversation, the CA would ensure that the survivor gets the appropriate information and services that meet their needs. This includes providing them with a safety plan, which serves as a guide to help them make the right decision and provide suggestions regarding how they may maintain their safety during their relationship, when planning to leave, and even

after leaving [242]. P9 said, *“At the end of our conversations, we always do a safety planning. Some people do have safety concerns, some people may not depend on their particular situation. But, to get a sense of what they've been doing to stay safe, and what tips or information they may need to consider a safety plan per day. And that really varies depending on the person's unique situation. But we have lots of safety planning information kind of at our disposal that we can bring to the conversation, depending on what that person is expressing regarding their safety concerns.”* It is also imperative that they are provided with necessary tangible assistance such as shelter, medical referrals, etc. P9 continued, saying, *“it might be that for coming into our various services, and we explain what those next steps are going to be, whether it's shelter or community-based services, if we're not the right fit for services, and they're needing other referrals, we make those as well.”* The objective is to ensure that survivors get what they need (i.e., information, service) and that it fits them, which may vary, depending on their circumstances. P11 mentioned, *“ending the conversation just depends on if you've been able to help them get the kind of information that they thought they were seeking out in the first place. So, I think the ending of the conversation looks really different depending on what they were seeking out.”*

MR6: The CA should provide appropriate information and services at the end of a conversation.

At the conclusion of the chat session, the CA should express appreciation to the survivor for using the tool and service. The CA may also display some messages to survivors that highlight their strengths as a means of encouraging and empowering them. P2 said, *“I always wrap up my sessions by saying, you know, thank you for coming in today. Like I am really sorry that it was the circumstances that brought you here today. But it takes incredible strength to walk through those doors and say, I need help. So, thank you for being here today. And thank you for sharing*

*your time and your story with me”* The CA's expression of appreciation will motivate and enhance survivors' satisfaction. [249].

MR7: The CA should show gratitude and empowerment to survivors.

#### 3.4.2.1.2 Language Use

The language use theme describes meta requirements in which CAs interact with DV survivors through language, which involves understanding, selecting, and utilizing appropriate language to ensure effective interaction. Several aspects are involved, like the choice of phrases, the content of the questions, style, and cultural appropriateness, all intended to provide survivors with a sense of support and empowerment.

DV professionals stressed the importance of providing survivors with trauma-informed language to address their concerns. The concept of trauma-informed practices involves recognizing the pervasive nature of trauma as well as preventing its unintentional recurrence through appropriate language use and communication [250]. All of the communication with the survivor should be built in a trauma-informed way. P6 explained, *“I think a survivor-centered approach is inherently trauma-informed. you know, a trauma-informed is that idea of rather than saying, What's wrong with you what happened to you? So, I think that you know, trauma-informed and survivor-centered should be the foundation for which all the communication is built.”* There are a variety of trauma-informed practices to consider. For example, the word 'victim' would be replaced with the word 'client' or 'survivor'. As P2 described, *“For a necessity. I think just really using trauma-informed language, I'm referring to them as survivors or clients instead of victims”*. In the same context, it might be better to ask the survivor what they want to be called, as some survivors may not be comfortable with having the abuser's last name in the conversation. According to P2, *“Because I might say, like Miss \*\*, and \*\* is my abuser's last*



*name, and that's gonna really trigger me. So, or I might say any, and they feel disrespected."*

Trauma-informed also includes an acknowledgment of the feeling that a survivor is currently experiencing. P11 said, *"people when there are certain things, certain traumas that are disclosed that you don't just say, Okay, next way, you acknowledge the weight of that particular trauma."*

MR8: The CA should use trauma-informed language.

It was concluded from our study that it is important for the CA to converse in simple and basic language when conversing with the user. Having complex metaphors, phrases, and long unclear sentences is more likely to impede the flow of dialogue and distract the user from the content. Therefore, it was suggested by many participants to ensure that all written parts in the CA are at a lower grade reading level to avoid any sort of challenges in understanding the content. P7 mentioned, *"something else to consider is to make sure that everything is at a sixth-grade reading level or below to avoid any sort of challenges with individuals that might have lower education."* Similarly, P6 referred to the same point *"and then reading level might just be something that you, you just make sure everything's at like a fourth-grade reading level."*

MR9: The CA should use simple and basic language.

As part of trauma-informed care practices, we identified one requirement that emphasizes the CA should refrain from asking repetitive questions or questions whose answers might imply repetitive details. The survivors may become distressed by having to recall traumatic events repeatedly and may lose their trust in the CA (figure 3.3). P11 discussed this point, *"So you wouldn't want to be like, have you been raped? And then they describe their rape in detail. And then you ask, the next question is like, misses the details that they describe the rape, if that makes*

sense. So, it's like, in order for it to be trauma-informed, it would need to be trained to translate that information really savvy so that it doesn't repeat questions in a harmful way.” By recalling traumatic events repeatedly, survivors may lose trust in the CA [251].

MR10: The CA should avoid asking repetitive questions or questions whose answers might imply repetitive details.

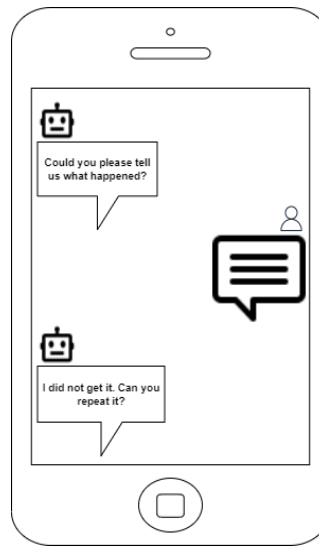


Figure 3.3: Conversation guideline- Avoiding reiterating the question

Besides, several participants indicated that both open-ended and close-ended questions are beneficial (Figure 3.4). They emphasized that closed questions may be helpful, especially for questions requiring rapid responses or regarding the survivor's safety. For example, P2 mentioned, “*Probably a few closed-ended questions to begin with things like, Are you safe right now? Like, are you in immediate need*” Similarly, P4 discussed the importance of closed ended questions, “*the closed-ended questions are also valuable. Because you know, do you have any church or unit or any faith-based organization or something like that's in your life that will be able to help you? Is that you would turn to if that's important to you? Is there a DSS involved? Are there kids in school? Do you feel safe? Yes or no?*” On the other hand, open-ended

questions are also useful in this context. To ensure that we get a true and accurate picture of what happened, it is important to let survivors express themselves freely and truthfully. Instead of limiting the conversation to a set of predefined questions and responses, employing open-ended questions would enable survivors to share their stories and convey their feelings more freely. As P4 pointed out, open-ended questions are particularly important, *“we have to ask open-ended questions because we have to hear about their situation that may be different from somebody else's situation.”* It enable users to share their stories and convey their feelings more freely [252]. Most of the existing DV CAs systems are designed primarily based on a single type of questions, which is closed-ended questions. Many survivors would want the opportunity to talk about their experiences and receive guidance. The CA might provide an environment of "secret diaries" for survivors by offering a space where they can share, as an alternative to yes/no questions or multiple choice. The CA is likely to be appealing to survivors who are hesitant to speak about sensitive topics such as DV to their families or friends.

MR11: The CA should use both closed-ended and open-ended questions.



Figure 3.4: Conversation guideline- Using closed-ended and open-ended questions

Other types of questions to be avoided in designing the CA include personal/invasive questions, where the answer may imply identifiable information, such as age, gender, and health information. As P6 answered when we asked about questions to avoid in designing the CA, *“unnecessary questions about the experiences of violence, like asking a lot of invasive health questions, asking a lot of invasive violence questions asking a lot of invasive just asking a lot of invasive questions.”* On the same topic, P10 recommended avoiding questions that include identifiable information, *“so any unnecessary or overly identifying information, if you could avoid asking them, if you could avoid asking her name dates, such as age.”* Further, it was suggested to refrain from asking pushy questions in which respondents are forced to make a decision and conditioned to respond in a particular way. P11 said, *“you’d want to make sure that the questions are written in a really specific, empathetic way, and that there’s never any, like, overly pushy questions.”* We concluded that invading, personal and pushy questions may overstep boundaries and have the potential to negatively trigger the survivor, which in turn would adversely impact their trust toward the CA.

MR12: It should be ensured that the CA avoids personal information, and invasive and pushy questions.

A further aspect that we deduced through our analysis of the script is that the CA should avoid overpromising to solve the DV issue. The CA should avoid phrases such as “I will resolve your problem”, or “ your issue will be fixed”. P4 asserted, *“so we can’t promise that we can make everything better for sure. I think that our program really tries not to make those promises that we can’t keep.”* In many cases, survivors of DV are repeatedly abused by their abusers, and may also be suffering from mental trauma, so such promises can be misleading to them and frustrate them.

MR13: The CA should avoid overpromising to solve the DV issue.

#### 3.4.2.1.3 Support Provision

This theme covers the meta requirements pertaining to the provision of various forms of support by CAs, including informational, emotional, and instrumental support. These requirements are formulated to cater to the distinct demands of survivors and facilitate improving their well-being holistically.

##### 3.4.2.1.3.1 Informational Support

This refers to the CA's ability to deliver relevant and useful information to DV survivors. Several requirements were identified. First, we have learned about the importance of implementing the CA as a means of educating survivors of DV and providing them with accurate information regarding this issue. Interviewees perceived that the CA could enhance the awareness of survivors and communicate information in a dynamic and interactive manner. A CA may, for instance, provide accurate information about DV like connecting users to the power and control wheel framework [253] as well as tech safety <sup>1</sup> to increase the user's knowledge about potential abuse. As P1 mentioned, *“I think education and safety planning. I think that linking them to that, because sometimes, what your friends are saying, or what your family is saying, can be their opinions. But that's power control wheel is what is true.”* Further, the CA can be viewed as a helpful source that can educate a user to recognize and identify their situation (i.e., type of abuse), thus making them aware. Survivors sometimes are unaware of what they have been through and do not realize the potential risks associated with such experiences. The CA could therefore be beneficial. P1 explained, *“providing education so that they know, they can identify that they are in a bad situation.”* Similarly, P6 referred to the same point in her

---

<sup>1</sup> <https://www.techsafety.org/>

answer, *“the need to provide education that maybe would have occurred, so I can imagine actually, chatbot being really strong for this, which is like, my boyfriend did this thing to me, or my girlfriend did this thing to me. my girlfriend called me names and put me down, is that abuse? What does this mean?”* CAs can be considered a promising avenue for delivering information and supporting individuals' well-being [254].

MR14: The CA should provide education to the survivor.

Besides, the CA could explain the concept of safety planning and allow users to make a safe plan, or it could direct users to a credible source (e.g., the National hotline) that provides safety planning. According to P1, *“I think that they could do some basic safety planning because just especially with technology and stuff like that, I think that's something that could basically go on to a chatbot.”* There has been a suggestion that survivors should be provided with safety strategies and concrete suggestions for a variety of situations (e.g., safety in the school system, court, rural areas, internet, etc.) P11 said, *“But to do some like immediate crisis planning isn't a bad thing. And to include some of the safety strategies that exist out there and kind of disseminate that information in real-time would be useful.”* In a similar context, some participants referred to the possibility of applying some safety and priority assessments via the CA. These safety measurements aim to enhance survivors' awareness regarding their safety priorities, receive feedback regarding their vulnerability to severe violence, and lessen their decisional conflict regarding their safety in abusive relationships. For example, P11 discussed the benefits of these aids (which were implemented in the MyPlan app [23]) and how they enhance the overall well-being of those abused, *“there is a lot of research on the MyPlan app and how the safety decision aid really does empower women and help them have more emotional clarity about what they need to do. So, it does reduce decisional conflict. And to me, it's one of the most*

*well-researched and most consistently reported interventions if there's ever been for domestic violence, so it's quite promising. So, I think decision needs are a really important feature.”* In the same context, P3 suggested utilizing Dr. Campbell's danger assessment tool to assess the danger level of the survivor, *“you could put links to the Jacqueline Campbell, Jacqueline Campbell tool”*.

MR15: The CA should offer safety strategies.

Further, CAs were seen as promising technological means to provide tips, advice, and guidance regarding the survivor's relationship, their safety, signs of abuse, protection practices, and other related topics. By way of example, P4 described how the CA could provide tips about legal processes, *“provide maybe some guidelines maybe have also some pre-populating guidelines for filling out a protective order in the event that you have to fill it out on our own and not with one of us... ..so maybe just being able to offer that piece of information, like here's five tips for you to be able to write your protective order, and just be able to provide some concrete information that we know to be true.”* It is thus essential to offer some pertinent tips in relation to the needs of survivors to help enlighten and guide them. P9 said, *“pointing someone to like a list of tips and things that can help them from there, that can be, really helpful”* CAs can provide tips regarding the survivor's relationship, their safety, signs of abuse, protection practices, and other related topics [29].

MR16: The CA should offer relevant tips and advice.

#### 3.4.2.1.3.2 Emotional Support

In developing a CA, one key factor for consideration is providing emotional support in such a way as to demonstrate empathy and genuine concern for the recipient [255]. Additionally, the

CA should be constructed to respect the survivor's choice and understand their concerns. To this end, we outlined several requirements regarding the use of emotional support in our proposed CA.

First, an important consideration when developing a CA is that the conversation should reflect active listening and empathy for the survivor. When they are telling their stories, survivors have a strong need for respectful language and empathic and supportive responses. As P3 explained: *"I needed for people to hear it and say, Oh, my God, that's so terrible. And when telling the story. It's hard to get sympathy. It's hard to get in empathy."* Similarly, P2 stressed the importance of empathy, *"But I think probably the most important one is empathy and respect for the person. You know, I don't know, what has happened in their life to lead them to be in this situation."* It is possible that some survivors are not in need of immediate assistance. Nevertheless, they might want emotional support and empathic responses, and as such, the CA should be able to accommodate this. P7 stated, *"I have a way to text or chat with someone that is going to be empathetic and listen to my situation. Maybe I don't need immediate help. I just need support emotionally."* Various supportive statements may be incorporated into the CA, such as, "It is not your fault; it is possible for any individual to find themselves in this position at any time.", "How may I assist you?", "I can understand how distressing it is when someone close to you is frightening you.", etc. (Figure 3.5). In response to our question about empathic responses that might be provided by the CA, P4 explained, *"I would say a couple of the responses might center around respect of their choice. If somebody were saying something, like, I feel scared, and everybody's mad at me because I don't leave. And so, we could offer some sort of response, like, it is unsafe for a lot of people to leave. And it's very scary. So, we understand that this is hard for you, how can we help? We just do you just need someone to listen, and just be very respectful of*



*their choices and where they are? And I'm sure there are a million responses that you could formulate into that. I think, I'm just continuously using those words like we understand. This is a scary thing for everybody to do. It's not this abuse is not your fault. Your situation is not your fault."* In the same vein, experts in the field have emphasized the necessity of providing an active listening mode and demonstrating the understanding. P3 explained: *"I realize that sometimes they just want to be heard and they just want a kind voice somebody who will let them express themselves and so very often that's really all it takes is just I understand your frustration, that's pretty much it just listening, and you know, and expressing understanding."* Likewise, P2 commented: *"sometimes they just have not ever said, what they've experienced out loud. And they just need some buddy, that's there and willing to listen, because no one has ever let them talk before about their experiences. So just a kind listening ear is sometimes what they need. And sometimes it's saying it's not your fault. Yeah. You didn't deserve this."* According to empirical evidence, teenagers hold the expectation that a CA with active listening capabilities could effectively address their emotional needs [256].

MR17: The CA should be an empathic, active listener, and patient when communicating with the survivor.

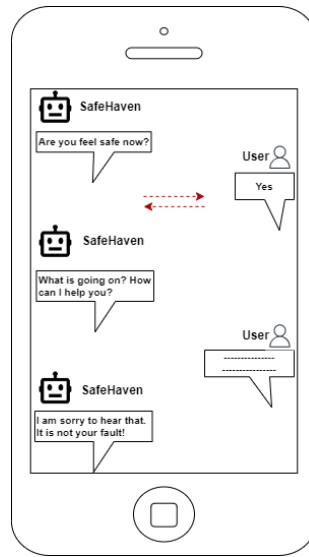


Figure 3.5: CA feature design- offering emotional support

Communication with the survivor is primarily developed based on the idea of a survivor-centered approach [242]. Through this approach, survivors will have many options available to meet their needs. It is crucial to ensure that CA provides a supportive environment in which the survivors are treated respectfully, and their choices are respected. P4 mentioned, *“we also are trying to be respectful of their choices. And I think that their normalcy and choices may not be the same as ours.”* In existing recommendations, respect and empathy are qualities that should be prioritized when designing CAs, particularly when catering to individuals experiencing mental illnesses [257]. It is desirable that survivors have multiple options from the CA and can choose what suits them best.

MR18: The CA should respect the survivor’s choice.

A blaming atmosphere might undermine the ability of individuals to receive emotional support. According to [258], one of the most common barriers to getting help is the presence of judgmental reactions from formal and informal support. Having the ability to discuss sensitive subject matter and obtain insights without being judged is considered highly valuable [259]. Four

participants suggest that the conversation with the survivor should be built around the concept of non-judgmental and open minded, as this would create a comfort zone for discussing their issues, and they are more likely to continue the conversation. For instance, P5 said, *“I think, being non-judgmental, open-minded, all that establishes a relationship with them.”* In the same context, the CA should avoid victim blaming and judgmental questions. This refers to the practice of asking survivors what they might have done differently to prevent the abuse from occurring; therefore, it implies that the survivor is responsible for the abuse rather than the abuser. Examples of victim blaming questions include, “How did you cause this to occur?”, “Why did you not tell him directly to stop?”, “Why didn't you report this sooner?”, “Why not simply leave him alone?” Some of our interviewees pointed out the importance of avoiding these types of questions. For instance, According to P1, it is crucial not to ask blaming questions since the act of abuse is not initiated by the survivors themselves, *“anything that would say to a turn it around, like, what did you do to cause this or something like blaming someone for deserving? Like, what did you make this? What could you do differently to get this not to happen? Those kinds of questions, because it's not something that they did. It's something that the abuser did.”* Similarly, P7 raised the same issue, *“questions like, why are you staying? Why do you put up with this any question that may appear judgmental, or leading?”* Moreover, it was suggested that the CA should structure the questions in a way that does not assume the survivor will leave the relationship. Placing such an assumption might not be the best solution. Instead, it may complicate the matter. P4 said, *“and questions that are assuming that that person is going to leave. So that person may not want to leave, they just may want support.”* CAs should avoid victim-blaming and judgmental questions since the abuse is not initiated by survivors themselves [29].

MR19: The CA should be nonjudgmental and open-minded.

One additional valuable recommendation that the CA should offer is to de-escalate the crisis. The CA could use techniques such as positive language, a message that is free of reflective emotion, and validating language to help calm the individual. P4 stressed, *“we have to try to de-escalate the situation and make them feel safe and calm and comfortable.”*

MR20: The CA should de-escalate the crisis.

#### 3.4.2.1.3.3 Instrumental Support

There was consensus among participants that the CA could be an effective tool for linking users to tangible resources and facilitate their access to support to assist survivors of abuse and neglect. The following guidelines have been drafted.

We note that one primary requirement that the CA should provide is referring the survivors to different resources. The referral information on tangible formal support options to users may be provided locally (i.e., city) or across a broader geographical area (i.e., state, nation). CAs can provide information regarding local resources and community referral options; these options need to be developed, accurately maintained, and accessible to the users. For instance, CAs, as described by P2, could facilitate connecting survivors with resources, *“it should connect out to local resources, and so be able to that”*. Similarly, from the viewpoint of P9, the CA is useful for orienting the user to local services, *“So I guess a chat box feature that would be able to connect people to the resources that exist within like a really good comprehensive website like that when it makes sense to me.”* As part of the CA's services, survivors would have access to referrals for medical, safety, or mental health services such as emergency shelter, family therapists, hospital, DV advocacy, mental health therapists, financial, food, legal, and pet safety services. P5

mentioned, *“if there's a way that like if, they do want a protective order, or they are looking for like legal aid, specifically have that go directly or route to our core advocate, you know, something like that, where she could go ahead and get the information and have everything that they've already entered.”* In a similar context, P11 stated that, *“you can give point to directions where there are resources, obviously, so like, here's your local shelter, here's your coalition. Here's this mobile advocacy program. Here's the local rape crisis center.”* Plus, there is an emphasis on where and when to refer. In particular, there was a focus on referring survivors to helpful hotlines (e.g., National Domestic Violence Hotline, and Crisis Text Line) and local advocacy organizations that could provide survivors with resources. Referral to 911 or any emergency provider should always be considered if there are indications of physical danger or health emergencies. The making of referrals does carry with it an important principle that involves respecting the decision of the survivor.

**MR21:** The CA should make referrals to resources.

One of the features that might be incorporated within the CA is the ability to determine the location of the user. Based on the user's location, the CA can provide them with local resources based on their needs, whether they are housing, financial, family, food, or health-related (Figure 3.6). Obtaining the user's location might be accomplished by asking the user for permission to allow GPS location or asking them for their Zip code. P6 mentioned, *“A way to enter your zip code and get your local 211 or law enforcement, your local, like first responders or medical care”* P8 also emphasized the same idea, *“how if they put their zip code, it can show right away. What safe houses are nearby locally, so they can flee that situation? Food, food services as well.”*

**MR22:** The CA should provide localized resources.

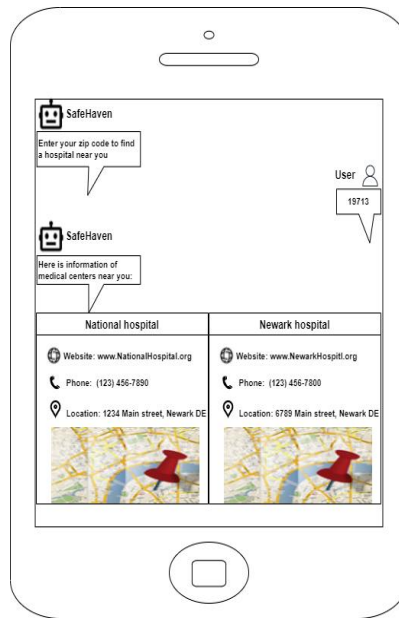


Figure 3.6: CA feature design- providing local resources

#### 3.4.2.1.4 Trust and Rapport Building

The theme of trust building pertains to meta-requirements regarding the cultivation of trust and the formation of positive relationships between CAs and DV survivors, which necessitate the presence of essential features like confidentiality, attentiveness, and transparency on the part of CAs. A number of meta-requirements were identified.

First, the feeling that the tool will provide a safe environment for survivors, and that whatever is presented in the conversation will remain confidential is one of the factors that will increase communication and enhance the ability to speak comfortably. It is crucial to let the survivor know that all information in the conversation should remain confidential. P8 stressed that *“we let them know that everything that they're going to talk with us is confidential, not share with anyone, we'd let me explain to them if you want us to share your information with other organizations, other professionals, they have to sign the ROI, which is the release of information... So, they understand that whatever you're telling us today, is not gonna be going*

*out of this office, it's not gonna go out of here. So, they understand that they can trust us."* It is necessary that the CA explicitly explains that any information that users reveal during a chat session is completely confidential. P11 explained, *"some people feel a little more comfortable disclosing some of the information to, you know, a bot or artificial intelligence or whatever we want to call it, they're more comfortable disclosing some of that information because it is so private. So, I think maybe being really explicit that the information that they produce during the chat session is completely private, that would be important, then it's anonymous, and their data is safe."* Further, building a rapport with survivors is highly dependent on the preservation of privacy and safety. P4 mentioned in her answer to the best practices to establish rapport with the survivor, *"with establishing a relationship ... I think that being focusing on safety, and also being nonjudgmental, and realizing that, how many barriers that they do have... we just need to focus all of our interactions need to be designed around safety and respect."* Similarly, when asked about the main features and capabilities that the CA should have, P3 responded by emphasizing safety and security as critical features. Thus, CA designers should implement measures that guarantee the confidentiality of survivors' conversations and the protection of their privacy [256].

MR23: The CA should ensure the confidentiality and privacy of conversations and data.

As part of establishing a rapport with survivors, it is equally critical to demonstrate compassion via attentive listening, empathy, and thoughtful use of questions and make sure to always remain available to them. Engaging in active listening involves not only listening to what is being said but also listening to what is not being said and interpreting it in light of the nonverbal message [260]. The attentiveness and listening activity involves more than just listening silently, but also receiving, understanding their concerns, and reacting. P7 described, *"building trust and those activities might center around, just active listening, making sure that*

*you are accessible to that individual, that you are allowing them to be the subject matter experts in their lives and you're not coming in to try to dictate that."* P11 raised the same point, *"The best practice, is just being really empathetic and open and believing someone."* Further, In establishing a rapport, it was recommended to give survivors a space to express their feeling and gain empowerment. P10 noted in his answer to building a relationship with survivors, *"Basically just giving them the space to be more empowered in their own circumstance."*

MR24: The CA should ensure attentiveness and accessibility.

In addition, transparency is considered one of the six pillars of the Trauma-Informed Care (TIC) framework which has been established by The Substance Abuse and Mental Health Services Administration (SAMSHA) [261]. Transparency pertains to the extent to which an organization and its providers are transparent and clear about their programs and practices, to establish rapport among stakeholders such as survivors of DV. To this end, it is important for the service providers to introduce themselves and describe the service they provide during the intervention session. P8 said, *"one of the things that my staff do, every time they meet with the client for the first time is we let them know who we are, we introduce ourselves"* P9 also discussed the importance of transparency in building a relationship, *"we really try to take the trauma-informed approach. Just certainly acknowledging their situation, making sure that you're transparent and clear about what our services are, what we can offer, making sure that they know that they have a choice and that our services are voluntary, as well as what our basic guidelines are and what our boundaries are. around our services too."* In light of this, the CA needs to provide a brief introduction that includes the purpose of the chat session as well as the services the CA can offer. The transparency of the CA's purpose, services provided, and referrals would increase the amount of trust and engagement with the CA [262]. P9 stressed that *"as*



*much transparency upfront with what this feature is and what it can do and what it cannot do. to set those expectations for people who might be utilizing it is going to be really key.”*

MR25: The CA should ensure transparency throughout the conversation.

#### 3.4.2.2 Design Principles of CAs

Compared to conventional approaches used for general CAs, CAs designed to support DV survivors require the development of DPs that are unique and tailored to their needs. Due to the nature of DV and its sensitive context, CAs should be developed with DV survivors' needs, challenges, and expectations in mind. Based on the meta-requirements that we identified above, we outline a set of design principles (Table 3.3).

First, we propose our first design principle which emphasizes the importance of the CA's introduction to the conversation by starting with a warm greeting, ensuring the conversation's confidentiality, and prioritizing safety. Among the many quality attributes that CA is built on is affect, which includes attributes like greeting and expressing personality [166]. Additionally, the CA should take into consideration the survivor's safety before proceeding with the conversation. Evidently, digital interventions promote user confidentiality, privacy, and safety while simultaneously delivering tailored care on a real-time basis that can effectively enhance users' well-being [26]. These practices would enhance survivors' engagement and create a safe and comfortable setting for conversations. Therefore, we derive:

DP1: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with the capability of initiating the conversation with a warm greeting, ensuring confidentiality, and prioritizing safety to enhance a survivor's engagement and establish a safe environment for conversations.

There has been extensive research focused on improving the NLP and conversational capabilities of CAs so that they can provide more accurate answers to survivors' queries, as well as enhance user experience in terms of conversation flows [263], [264]. CAs are expected to fully comprehend the survivor's message and extract the intention that the survivor is conveying, regardless of how the message is worded, to be able to produce a response that has the potential to contribute meaningfully to the conversation [265]. In the same context, it has been suggested by researchers that CAs should emulate human communication characteristics to provide a more natural, engaging user experience [266], [267]. Based on the argument made by Atkinson et al. [268], the responsiveness of CAs fosters a feeling of reciprocity, which in turn engenders trust. According to Moon [269], a machine (i.e., CA) that implements a relation strategy of reciprocally increasing self-disclosure through textual interaction with the user is likely to disclose more personal information. Communicating effectively can ultimately lead to establishing a relationship and meeting the needs of the survivor. Thus, we propose:

DP2: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with a perceptive, flexible, and natural conversational flow geared towards a specific objective, combined with a reciprocated conversation, effective clarification, and contextual understanding to establish trust and enable meaningful communication to meet the survivor's needs.

Wrapping up the interaction is part of the trauma-informed, survivor-centered approach that can be ended naturally after offering resources [242]. The concept of empowerment, where survivors should have the ability to choose whether and how they wish to receive support, is common throughout IPV training [270]. Regarding the role of CA, the use of AI has great potential in the process of empowering people [271]. For instance, researchers found that CAs

could be empowering tools for adolescents with Autism Spectrum Disorder (ASD), providing assistance in meeting their individual needs, personalities, and goals, including developing personal care skills, managing stressful emotions, and learning new ways to communicate. Empowerment is a critical component of a survivor's life. By gathering relevant information and receiving emotional support from those around them, some survivors can build up their sense of power. However, in most instances, empowerment may be hindered by apprehension about social interactions. As a result, the CA has the potential to address this issue. Overall, ensuring that the conversation is properly concluded through empowerment and support will increase the likelihood that the survivor will remain engaged, improve their interaction with the CA, and ensure that they will feel supported. Accordingly, we propose that:

DP3: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with appropriate conversation closure through the provision of the necessary information as well as empowering and showing gratitude to enhance the survivor's engagement with the CA, and cultivate a sense of support.

DP4 highlights that CA should enhance the language conveyed to the DV survivor. DV advocates typically use trauma-informed responses and provide safe spaces for DV survivors to better understand the ongoing consequences of trauma [272]. It has been shown that DV survivors are receptive to technology-based interventions based on trauma-informed principles [106]. As far as the CAs are concerned, it has been demonstrated that they are effective in providing traumatized individuals with care. For instance, it has been explored how therapy for PTSD can be delivered by technologies like CAs, in trauma-informed approach [273]. Besides, another aspect of optimizing the conversations is the use of simple language and providing all the information requested by users in a minimal number of steps. There has been a suggestion

that the CA used in crises should be as simple as possible to facilitate access to and navigation through the information provided [274]. Relatedly, one of the requirements we identified is not asking the same question repeatedly so as not to trigger traumatic memories in survivors. Continuing to ask or answer questions repeatedly, in the same manner, would be contrary to conversational norms. A machine can respond to an input or ask a question repeatedly in the same way. Even though it may be an efficient form of automated interaction, the conversation seems unnatural [275]. Furthermore, we found that utilizing both closed-ended and open-ended questions could improve the language of the CA. Incorporating both closed-ended and open-ended questions can enhance engagement by preventing the questions from becoming monotonous (i.e., only close-ended questions) or time-consuming (i.e., only open-ended questions). To design such an interactive CA, a careful balance must be maintained among these interaction types. Lastly, one key point we identified from our interviewees was to avoid questioning survivors in a more personal, intrusive, and pushy manner. It has been demonstrated that users were dissatisfied with CAs that asked for personal information too soon following the first conversation, showing a lack of politeness in the conversation [276]. As pointed out by Muresan and Pohl [277], they concluded that CA's personal questions, as well as its frequent use of intimate emojis, presented an inappropriate degree of familiarity to users. Lastly, we found that the CA should refrain from making any promises it cannot deliver so as not to undermine survivors' confidence in its ability to carry out its promises. Hence, we derive:

DP4: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with the language abilities to convey a message that is trauma-informed and simple, incorporates both open-ended and closed-ended questions, and avoids repetitions,

invasive questions, and over-promising to solve the DV to avoid any re-traumatization and enhance the survivor's autonomy.

The DP5 emphasizes the importance of providing informational support to DV survivors through the CA. Assuring the CA is provided with educational materials and content regarding the topic of DV would assist in raising awareness about this issue. There has been a growing awareness among survivors that they are seeking informational support online [278]. Survivors seeking advice often turn to digital solutions as their first source of information [85]. Most commonly, they resort to such solutions because they are always available, accessible, and usable, and offer suggestions they might find useful. In that regard, the CA should be equipped with educational materials and DV-related content and offer such advice and safety strategies in an interactive manner. DP7 aligns with previous HCI research that stresses the vital role that CA plays in providing information to sexual survivors in terms of knowledge and advice [229], [252]. The survivor also should receive feedback about their situation and have a better understanding of their level of safety and priorities, thereby receiving information tailored to their needs. For example, it has been recommended that safety decision aid assessments be incorporated into the CA, as these assessments, which have been adopted as part of the MyPlan app, have shown their effectiveness in reducing decision conflict, alcohol misuse, depression, IPV, and increasing safety behaviors [279]. It is, therefore, important that the CA utilizes such safety assessments that will not only raise their awareness but also improve their decision-making. Acquiring the required knowledge would lead survivors to feel less uncertain and more empowered and make informed decisions so that they could take action to protect themselves. The ability to acquire the required knowledge would allow survivors to feel less uncertain and

more empowered, which would enable them to make informed decisions for their own protection. Therefore, we suggest,

DP5: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with the capability to communicate DV-related information, including knowledge, facts, advice, and safety strategies, so that survivors are better informed and less uncertain.

DP6 emphasizes that the CA should possess human traits, such as being a kind, empathic, and active listener, as well as respecting the decision of the survivor and being non-judgmental to comfort them and lower their emotional burden. Survivors may not require immediate assistance, and they ultimately seek respect, empathy, and a supportive ear. DP6 aligns with previous research, which demonstrates that offering emotional support (i.e., showing empathy and actively listening) contributes significantly to the recovery of survivors following a violent incident [280]. According to the proposed drivers of CAs in healthcare, they are capable of providing emotional support on demand without requiring human interaction [281]. CAs have also been observed to feel more likable and more supportive when they express empathy to patients who tell them they suffer from sexually transmitted diseases [155]. In the context of domestic and sexual violence, It has been demonstrated that CAs' ability to express empathy and provide active listening is perceived as supportive by survivors and reduces emotional burden [29], [187], [282]. It is important that CAs pay attention both to the starting and end of the conversation while demonstrating empathy. The reason for this is that survivors might experience anxiety before engaging in a conversation and in the end experience emptiness [283]. In this regard, an effective CA should possess human characteristics such as kindness and support, de-escalating a situation as much as possible. It may therefore be possible to promote

both machine-inherited (e.g., non-judgmental, unbiased, neutral) [29], [259], and human-like (e.g., respect, empathy, active listening) [78] characteristics. Thus, we propose:

DP6: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with human traits and conversational abilities that convey emotional support, such as being a kind, empathic, and active listener, as well as respecting the decision of the survivor and being non-judgmental to enhance their emotional well-being.

Taking into account the meta requirements we outlined above regarding instrumental support, DP7 highlights that the CA should provide functionalities and technical solutions that provide instrumental support in terms of referrals to resources and providing localized resources. Research has shown that resource referral improves the survivor's quality of life in several ways, including reduced exposure to IPV, reduced coercion, enhanced safety planning, and better utilization of community resources [284]. It is important to refer survivors to other people or organizations to facilitate their recovery if the trauma they have endured is extremely distressing [285]. It was concluded that CAs could make referrals to emergency services and hotlines in response to inquiries regarding psychological well-being, violence in relationships, and physical health concerns [209]. Furthermore, DP7 stresses the importance that the CA should provide localized resources to survivors based on their geographical location. It is essential that the resources provided have been updated and relevant to the geographical location. As an example, with the usage of natural language understanding (NLU), the CA showed its effectiveness in finding a list of COVID-19 testing sites and other local resources based on the user's location [286]. As a result of trauma, the survivor may find themselves in high need of nearby resources (i.e., shelter, hospital, local DV agency). Thus, they need a tool that facilitates their access to these resources and provides them with quick feedback. Hence, we derive:

DP7: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with functionalities and technical solutions that provide instrumental support in terms of referrals to resources and providing localized resources to increase their access to tangible assistance in a timely and efficient manner.

DP8 conforms to prior IS research [198], [262], [287], [288], which emphasizes the importance of trust, confidentiality, and transparency of the CA in building a rapport with the user. People were observed to be more comfortable disclosing to CAs than human peers, especially when the content is sensitive, due to a variety of machine characteristics (i.e., non-judgment, an adequate level of confidentiality, emotionless, patience)[187], [192], [282]. Providing privacy and transparency in a CA is of paramount importance; Lack of transparency may discourage some users from using automated CAs, as well as undermine the trust they place in them [262]. With respect to transparency of AI, and as suggested by our interviewees, it is essential to provide an explanation of the processing of user input, the purpose of the CA, and what it can provide. During times of uncertainty (i.e., DV), transparency becomes especially crucial. To assist users in understanding the CA's intellectual competence, it has been argued that the CA should be transparent to them [289]. In designing the CA for mental health patients, it has been suggested that designers and developers should ensure transparency regarding patient safety and privacy concerns, thereby ensuring that patients trust the CA with their health issues and feel secure divulging sensitive information [288]. Considering domestic and sexual violence, designers were advised to organize and communicate privacy and security issues in a transparent and straightforward manner, along with detailed explanations. Thus, we propose,

DP8: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with conversational capabilities and functionalities that ensure the survivor's



information is private and confidential throughout the conversation, and ensure transparency about what they are offering to foster a sense of trust.

Table 3.3: Design principles based on the identified meta requirements

Meta Requirements (MR)	Design Principles
MR1, MR2	DP1: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with the capability of initiating the conversation with a warm greeting, ensuring confidentiality, and prioritizing safety to enhance a survivor's engagement and establish a safe environment for conversations.
MR3, MR4, MR5	DP2: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with a perceptive, flexible, and natural conversational flow geared towards a specific objective, combined with a reciprocated conversation, effective clarification, and contextual understanding to establish trust and enable meaningful communication to meet the survivor's needs.
MR6, MR7	DP3: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with appropriate conversation closure through the provision of the necessary information as well as empowering and showing gratitude to enhance the survivor's engagement with the CA, and cultivate a sense of support.
MR8, MR9, MR10, MR11, MR12, MR13	DP4: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with the language abilities to convey a message that is trauma-informed and simple, incorporates both open-ended and closed-ended questions, and avoids repetitions, invasive questions, and over-promising to solve the DV to avoid any re-traumatization and enhance the survivor's autonomy.
MR14, MR15, MR16	DP5: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with the capability to communicate DV-related information, including knowledge, facts, advice, and safety strategies, so that survivors are better informed and less uncertain.
MR17, MR18, MR19, MR20	DP6: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with human traits and conversational abilities that convey emotional support, such as being a kind, empathic, and active listener, as well as respecting the decision of the survivor and being non-judgmental to enhance their emotional well-being.
MR21, MR22	DP7: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with functionalities and technical solutions that provide instrumental support in terms of referrals to resources and providing localized resources to increase their access to tangible assistance in a timely and efficient manner.
MR23, MR24, MR25	DP8: For CA designers and researchers to design a CA that offers support to DV survivors, provide the CA with conversational capabilities and functionalities that ensure the survivor's information is private and confidential throughout the conversation, and ensure transparency about what they are offering to foster a sense of trust.

### 3.4.3 The Use of CAs: Opportunities, Challenges and Design Suggestions

As an answer to our second research question (b), this section describes the expected opportunities and challenges that survivors might encounter during their interaction with the CAs. We further discuss design suggestions and initiatives that can be carried out in a CA to support survivors and overcome the identified challenges. Moreover, we present the proposed conversational flow.

#### 3.4.3.1 Opportunities

Based on our analysis of the interview data, we identified several potential strengths and opportunities of using the CA.

- **Availability and accessibility:** One of the expected benefits of using the CA is availability and accessibility. It means that the system is available, accessible, and will run continuously without interruption for a set period of time. As a result, the user can use the system, gain access to services, manage data, and obtain results. It can be accessed from anywhere where there is an Internet connection and from any device that is equipped with a web browser or application. P2 shared her perspective about the expected opportunity of using the CA as being available and accessible all the time, *“I think that they're more accessible. They could be 24/7. Um, you don't necessarily need to be in a space where you could be on phone talking to somebody. So, I think that they could be more accessible.”* P6 also commented, *“I think the main advantages are just that it can be used day or night.”* Continuity in work, accentuated by stable performance and fewer errors, makes the CA more desirable. P11 explained, *“one advantage is that it's never going to be overly tired or burnt out. So, it won't accidentally, like, get snippy with someone or impatient.”*

- **Impartial space:** As a machine, CA has the inherent strength of exhibiting positive stereotypes by its very nature. People view the machine as non-judgmental, safe, and unbiased. With this neutrality, the user is more inclined to disclose and interact more comfortably. P3 mentioned in her answer, *“I think it would help keep the emotions down because it's hard to get angry at a machine, and it's impartial.”* Interestingly, the non-judgmental nature of the CA may encourage user engagement.
- **Starting point:** Three participants referred to the CA as a good starting point for survivors. It may serve as a comfortable space in which the survivor can recognize their circumstances and thus become more conscious. Providing them with the right resources is also part of this process. P1 stated, *“I think chatbot would be a great first step. And being able to see something that you recognize, and you identify with, I think could be, and then you provide them with that information to get services. I think that that would be a good segue.”* Similarly, P2 answered about the expected opportunity of using the CA, *“I also think that they could be a good starting point, I've had clients tell me that they don't know if this is domestic violence or not. And so, a chatbot might be something that somebody who feels like, well, I don't know if this is actually domestic violence, they might feel more comfortable doing a chatbot.”*
- **Support provision:** Another advantage that can be seen from the usage of the CA is the provision of support to the survivor. CAs can be viewed as a complementary tool, especially in situations where advocates might not be available (i.e., a shortage of staff) or in areas where the availability of DV agencies is not widespread. P6 said, *“I see the chatbot helping to get people resources and information when the staffing isn't there... it could help support work with IPV survivors and other agencies. If there's staffing*

*concerns, like if you don't have a domestic violence specialist, because, you know, you don't have that person staffed at your medical institution anymore, then you can use the chat box to address staffing crisis.”*

- **Anonymity:** Many participants mentioned the superior level of anonymity that could be afforded by the CA. Researchers have suggested that the presence of an anonymous space, as the CA does, offers a venue for people to share their experiences as well as receive relevant assistance (e.g., informational support) [30]. Anonymity is linked to disinhibition, the feeling of having less restraint when communicating with a machine compared to face-to-face communication. By ensuring anonymity, users will feel more comfortable, resulting in an improved user experience and eliciting more user feedback. P11 mentioned, *“I think that there is a chance that, it would maybe give people a sense of more anonymity, which is an advantage. I think eventually, it should turn to a person, but like it should point someone to services where they're going to be speaking to people. But I could see it having to elicit more in-depth feedback because of the anonymity.”*
- **Continuous and consistent emotional support:** A further opportunity that the CA could offer derives from its ability to provide emotional support in an ongoing and consistent manner. It has been noted previously (See section 3.4.1.3) that one of the key obstacles facing survivors is receiving an adequate level of emotional support due to the possible emotional drain and the inability of service providers to provide survivors with full emotional support and attention all the time. Since the CA is a machine, the CA can provide empathy and active listening to the survivor around the clock.

### 3.4.3.2 Challenges

The optimistic viewpoint of CAs notwithstanding, it is crucial to take into account the challenges that users may encounter when interacting with or disclosing their DV stories to a bot. To be able to receive assistance from a CA, survivors are expected to be willing to share their stories without hesitation. It is of utmost importance to identify survivors' challenges to create a better CAs environment after considering how to alleviate them.

According to the interviews, we identified a couple of challenges that survivors might encounter during the CA interaction. These challenges centered around three main areas: privacy, usability, and emotional support.

- Privacy and confidentiality: This refers to the risks posed to the safety and information of survivors. It is essential to note that even though the CA may have great potential, which does not mean it can eliminate survivor privacy and confidentiality concerns. As a result of our study, we discovered some privacy concerns that users may encounter when using the CA:
  - Recognition by their abuser: survivors are anxious about being detected using this technology, as it could pose some risks to them (e.g., continuous abuse), as well as exposing them to restricted control. For instance, P8 said, *“if the victim is still living with the abuser, I think the only concern that I have is that the abuser can open up that chatbot. Safety is always going to be a number one concern, that's number one for safety when it comes to domestic violence.”*
  - Access and hacking: One concern is unauthorized access to the CA, which may compromise the system. An example of this would be misusing the CA to damage and corrupt systems, spying on users, or disseminating secrets

inappropriately. P2 stated, *“So if they're really afraid of their abuser, finding them or if their abuser has hacked their phone, or their computer.”*

- User data: Another aspect to consider relates to the data the survivor provides and their fear of being lost or revealed to the public, or the CA might not store the data securely. A concern might also exist regarding the chat history, and how the CA manages it to ensure that the survivor remains safe. It is essential that the data of the users be preserved to the full extent possible. P11 explained, *“if the Chatbot leaves a history, and they're already being monitored by their partner that could, you know, really promote a dangerous situation, or it can really present a dangerous situation. So, I think accessing the chatbot would need to be really secure and safe.”* Similarly, P1 referred to the same point, *“the thing that concerns me most is the safety of it. Is it? Does the chat go away completely? Like, does it disappear? Completely? if you don't interact for it with like, for a minute, will it just dissolve? Will it just go away?”*
- Usability: In this context, it refers to limitations that the user might encounter during their interaction with the CA or when trying to carry out their intended task. We explored a few usability issues that the CA should avoid, to deliver an optimal service to the user:
  - Technical issues: This refers to possible problems related to using CA, such as lack of consistency, responsiveness issues, and errors. For example, the CA may not be able to recognize the word entered by the user, redirect the user to the same or wrong location, interrupt at the wrong time, or experience an

inordinate delay in responding. Ultimately, this can lead to frustration, inconvenience, and a feeling of distrust toward the CA. P11 mentioned, *“if there's even like the most minor of technical glitches, people will not trust it. And even if it's really good quality, if, if it messes up one time, people automatically see the entire resource is untrustworthy. So, any technical errors, people are pretty harsh judges of that.”* P5 also described the challenge of picking the wrong words, which could result in continuous recycling at the wrong point and negatively affect the user experience, *“the biggest challenge is it picked up on the wrong keyword, and it was a completely different response than what it should go to, or it just kept recycling through or whatever. And then at that point, they may have just worked up the courage to disclose or to reach out or whatever, and then that point, it may draw them back in like, See, I should have done it anyways. Or, you know, whatever the case may be said, I think that'd be the biggest one for me.”*

- Lack of experience and knowledge: Another concern that might stem from inexperience and insufficient knowledge of using the CA. Some people may not have the literacy skills to use the CA or may be unable to do so properly because it differs from what they have been accustomed to. P1 said, *“people that don't have art technology savvy, might find an issue with it. And the thing is when you first tell your story for the first time, and if it's not received, well, that can be really damaging”* P11 mentioned the same issue, *“So maybe they do or don't have the literacy to use the Chatbot. Maybe it confuses them,*

*maybe they can't access the website properly because of what they are used to"*

- Emotional and social support: In view of this, the CA may not adequately provide the emotional support that it might otherwise provide (i.e., human). During our interviews, we observed that two participants expressed doubts about the CA's capability to offer an adequate level of emotional support to users, while also being able to recognize the emotion that a user exhibits accurately. For example, P11 said, *"I think that it's important to remember that information can be conveyed, and information can be conveyed in a human tone, but there isn't someone there at the end of the day, it's just, it's just a machine, which isn't bad, but it just isn't going to be emotional."* Being aware that it is a CA might cause them to feel that it will not be able to understand them and comprehend their feelings and provide them with adequate emotional support. P4 explained, *"that you don't have that human interaction. So, you don't have, if somebody is here with me in my office, they may feel a little bit more reassured because they actually have a person they have somebody that they know. Through the chatbot, it doesn't have that human element to it, they can never replace. So that may be, you know, less comforting to someone."* Understanding an individual's emotions and feelings and being able to respond accordingly is called empathy. Empathy is an essential socio-emotional skill for successful interpersonal interactions [290]. Although the CA is a bot and the other end is completely devoid of a human, it is essential that the CA emulate some human characteristics to improve engagement. Given the automated nature of the CA, there is a possibility that it acts in a manner that leaves the user feeling emotionally burdened. This can occur, for instance, when the CA asks the user to repeat the story multiple times or



fails to respond appropriately when the user needs immediate emotional support. As P11 mentioned, *“so that it doesn't repeat questions in a harmful way and have them repeat their own information. That's a really difficult story, unnecessarily.”*

### 3.4.3.3 Design Suggestions for Mitigating CA-Related Challenges

This part discusses how to reduce the challenges of using the CA and its implications to improve the quality of interaction and enhance the user's engagement.

- **Privacy & Safety Guidelines:** We recommend several practices to improve the user's safety. DV survivors might fear that their abuser will learn about their CA use or break into their app to access it without their permission. It is reasonable to be concerned since the access does not require technical expertise, and it may require merely deducing a survivor's password that includes personal information (e.g., phone number) [291]. As a means of providing a safe environment for users and allowing them to converse freely, the designer should consider implementing multi-factor authentication. This approach would prevent unwanted access, and personal information would not be compromised by incorporating identity verification into the login process. This mechanism implies utilizing diverse authentication methods (i.e., pattern, PIN). In the same context, to safeguard survivors' privacy, the designers of CAs should also make it known to survivors that passwords that contain personally identifiable or public information can be compromised easily. Thus, the use of complex password patterns is highly recommended for them. There has been a call for survivors' privacy education from the researchers [291].

Another option to ensure the privacy of a chat session, in addition to hiding the chat feature, is to delete the chat session. In contrast to other applications, CAs or instant

messaging applications typically retain the previous conversations if they are reopened. However, if the previous conversations were deleted and cleared from the history, it would be difficult for the abuser to determine what was said even if the application was hacked into. According to the findings, the design of services for vulnerable populations (e.g., DV survivors) should differ from those for less vulnerable populations (e.g., shoppers). Less vulnerable individuals, for instance, might care more about usability and effectiveness than privacy/security. On the other hand, higher vulnerable individuals are more likely to pay attention to privacy due to the high-risk circumstances they may face.

In addition, the CA should be built according to the principles of transparency and detailed explanation to ensure the safety of survivors and their information. This implies clarifying the CA's purpose, and that all communications held by the CA should be preserved and not released to the public and stored in a secure database. Conversations between users and the CA should be confidential and anonymous. It has been demonstrated that survivors often prefer anonymous platforms (e.g., Reddit) as a means to maintain their anonymity and privacy [64], providing them with a safe place to share their experiences and receive support, and therefore designers should bear this in mind. In addition, for the survivor's safety, the CA should not solicit any personal or demographic information (i.e., name, age, race) that may violate their privacy.

- Usability guidelines: Several practices can be considered to make the CA easier to use. First, to reduce user errors or redirect users to the wrong or same locations due to a misunderstanding of the user query, designers of the CA might consider implementing a retrieval-based module, in which the user can input their query, and the CA should

direct them to the correct submodule. This implies that the CA should be trained and programmed with multiple keywords and questions to ensure it fits the user's query. In other words, the CA should be able to recognize the keywords, link the user to the appropriate module, and retrieve relevant information. For instance, when a user says, "Assessing risk", the retrieval-based module recognizes the inquiry as "Risk assessment" and connects the user to complete the risk assessment and obtain detailed information. This would ensure that user errors are prevented, and usability is enhanced, thereby increasing confidence in the CA.

Secondly, the designers should keep in mind that there is a possibility that there are some survivors who may lack the necessary knowledge to make the most of the CA. This would recommend using simple, intuitive, and easy-to-read language, avoiding long answers, using words that everyone is familiar with, and composing conversations in the form of one-on-one dialogue. As part of improving the readability of content and user-friendliness, it is highly recommended that the designer incorporates tables and visual images, rather than only relying on text. Information can be effectively conveyed to survivors through visual displays. Furthermore, it may be appropriate for the CA designers to include both button-based and text-based input responses instead of being limited to one type (i.e., button-based), or enable carousels in the user interface to enhance usability and interactivity.

- Emotional and social support guidelines: CAs Designers should consider how the CA could provide a diverse range of emotional and social support. These might come in the form of preventing self-blame, conveying to survivors that it is not their fault, understanding the case, reframing the situation, not judging survivors, accepting

survivors' stance, caring about survivors' emotions, feeling sorry for survivors, and empathizing with survivors' feelings [280]. The CA could compose thoughtful messages and empathic phrases as a means of comforting survivors. The CA could also serve as an emotional and informational empowerment tool for survivors. Our findings suggest that the CA may be a potential space for survivors who may experience secondary victimization from the online community (i.e., social media), thereby exacerbating the case. This supports researchers' recommendation for technology-mediated support for IPV survivors and vulnerable survivors adversely affected by online disclosures [113], [291].

As the CA is a machine, it does not possess human-like characteristics, which are critical in interacting with the user. To partially replace a human with a CA, a user should perceive the CA as a social actor. Based on the Computers as Social Actors (CASA) paradigm, humans employ the same social rules when interacting with both humans and machines [292]. Based on this, how humans interact with computers is similar to how they interact with individuals. Incorporating appropriate conversational language into a CA enables users to experience a high degree of social presence. In view of the foregoing, anthropomorphizing the CA is a crucial aspect, as it will enhance the interactivity of the conversation and facilitate emotional and social support. It is highly recommended that the CAs designer considers this design principle. Anthropomorphic cues are likely to activate heuristics associated with "humanness" [293], thus promoting social interaction with the CA [294]. Several common approaches to promoting anthropomorphism in the CA, include utilizing social cues like pauses before answering [295] or using emojis in conversation [296]

or showing the name and picture of the CA (i.e., avatar) (Figure 3.7). Certain social cues may facilitate a CA's emotional support during traumatic conversations. In the same vein, to maintain anthropomorphism in the CA and strengthen emotional support, CA designers may consider promoting reciprocity and encouraging messages. Recent research has shown that the reciprocity level in a CA has an effect on users' willingness to disclose information and their perception of enjoyment and intimacy [187]. Another suggestion is to create personalized CA and support survivors. This can be done by determining the user's needs and the use contexts and based on that, the CA might customize itself (empathic, informational, or logical based). The CAs can be seen as hybrids of their personalities. For instance, if the user asks for shelter or financial needs, the CA should respond respectfully, but in an informational and rational manner. On the other hand, if the users seek counseling or disclose their DV story to the CA, the CA should respond empathically (i.e., acknowledging, and encouraging messages). Overall, the CA should show a balance between logical and emotional responses, which implies the CA should show empathy, but not overly so, as some users may perceive it as fake emotional support or not even provide emotional support at all.

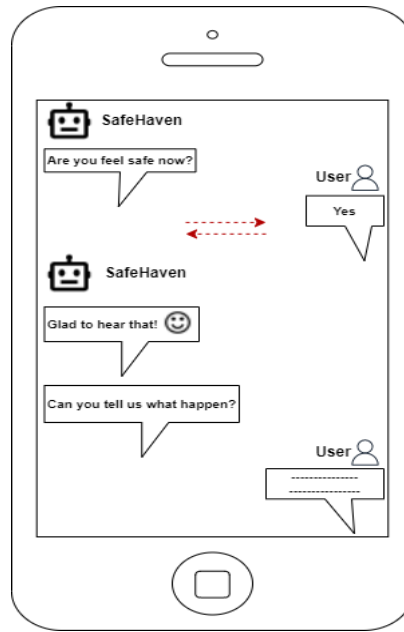


Figure 3.7: Conversation recommendations: Anthropomorphizing the CA

- The CA's conversation flow: Based on the interviews, it was learned that the concerns of the DV survivors were related to safety, awareness, and support. In this regard, the advocate's role is to fulfill these needs and to provide the service most effectively. As such, it is imperative to assist survivors in getting the support they need and becoming aware of the abuse. Importantly, the CA's non-judgmental nature aids survivors to disclose sensitive information (i.e., DV) [188]. Moreover, a lack of knowledge and a high need for resources are the most typical reasons for seeking DV services [297]. Therefore, we recommend starting by identifying the type of abuse experienced by the survivors, as a way to promote their awareness and educate them, after ensuring their safety. Figure 3.8 depicts a visual representation of the major phases of the conversation flow of the CA.

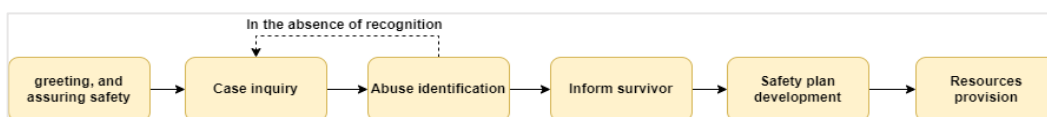


Figure 3.8: Major phases of the CA's conversation flow

When the CA ensures the survivor is safe, it asks the survivor to describe what happened, and so from that information, the CA will infer what kind of abuse was perpetrated on them. As part of this screening procedure, the CA will explain the type of abuse, what forms it might take, and how it can be prevented in the future. Following this, they will be assisted in setting up their safety plan, and then they will be provided with a wide range of resources to pick from according to their needs. Detailed explanations of all phases and subphases of the conversation flow will be provided in the next chapter.

## 4 SAFEHAVEN: DESIGN AND IMPLEMENTATION

### 4.1 Introduction

Based on the survivors' needs, their faced challenges, and design guidelines identified in previous chapters, this chapter attempts to incorporate those guidelines into our design of a system in support of DV survivors. The findings of our interviews show that the main concerns of the DV survivors were safety, awareness, and support. Previous research has also suggested that many DV survivors were unaware of the abuse they experienced or sought to know about the resources and options available to them or sought a safe environment [242]. Therefore, it is of crucial importance that the tool is designed based on these considerations to enable survivors to fulfill their needs adequately.

Based on the design rationale and advice provided by experienced professionals, we developed a CA for survivors of DV. Following the identification of the functions of reviewed CAs and the classification of those CAs into several similar categories (Table 2.3), our proposed CA for DV survivors can perform all of such functions. In addition, it consists of a hybrid approach that allows users to respond using both button-based and text-based responses. This would enable users to interact in a more intuitive and efficient way, while still providing the necessary safety and functions for DV survivors.

To answer RQ3, and 4, this chapter first discusses the process flow of our proposed system. In particular, the system covers the various stages that a DV survivor might go through. Further, we describe the key features and functions that SafeHaven offers to assist DV survivors. Then, we operationalize the design guidelines in our system and describe the major module and sub-



module of our system. Lastly, we discuss the CA implementation and present our evaluation plan.

#### 4.2 SafeHaven – A Dialog Flowchart and Key Features and Functions

Having a deeper understanding of the flow of the conversation would be helpful. Therefore, in this section, we discuss the process flow of our conversation design followed by key features of the CA. The process flow demonstrates a diagrammatic representation of the solution along with a detailed description of the steps crucial to solving the problem. It simultaneously provides a comprehensive overview of the processes and arranges the processes chronologically. In our study context, we have broken down our process flow into five major modules, namely, greeting and assuring the safety, screening and identifying abuse, developing the safety plan, survivor stories and referral resources (See figure 4.1).

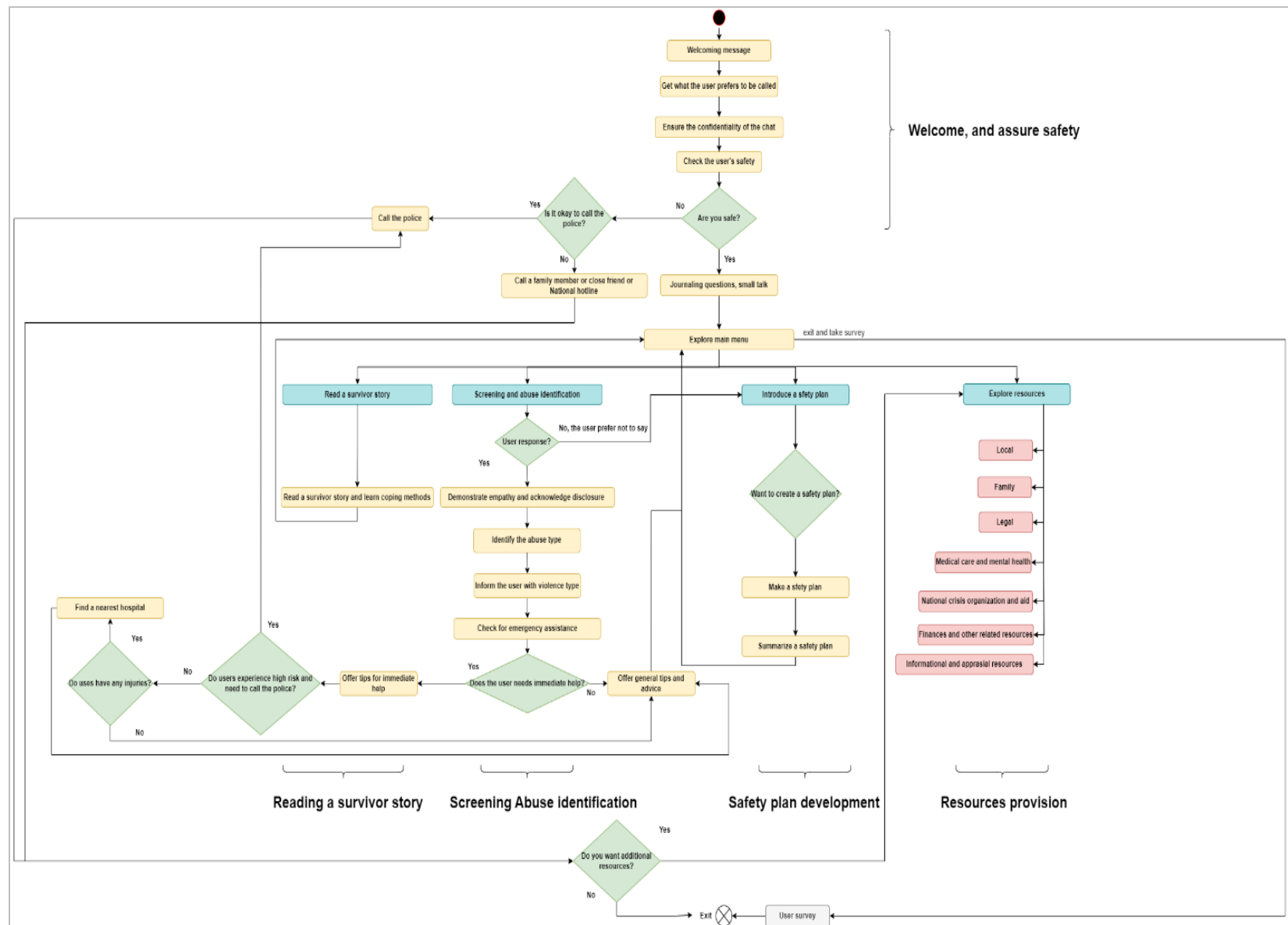


Figure 4.1: Flowchart of a dialogue session

The conversation between the user and SafeHaven starts with a welcoming message, followed immediately by assuring the safety of users. Upon confirming that the users are safe, SafeHaven will proceed to the main menu where the user will be able to browse different options. For instance, the user will be able to choose the survivor story module, where they will be able to read the story of a survivor and learn what coping methods have been used to cope with their trauma. The user will also be able to screen for abuse where they can identify the type of abuse they encounter. In addition, SafeHaven will introduce a safety plan to DV survivors and ask whether they are interested in creating one. Finally, SafeHaven will provide a variety of referral resources and tips that suit the user's needs for him or her to choose from. It is worth noting that the user does not need to follow the entire process flow but skip some phases while interacting with SafeHaven. We describe each module in detail in Section 4.3.

A series of key features and functions are offered through SafeHaven to help DV survivors.

- **Screen and abuse identification:** In cases of abuse that are complex, vague, or unclear, survivors may find it difficult to identify the type of abuse that they have been subjected to or might not feel they are being victimized. At the same time, some survivors may be misinformed when they search for information. Therefore, SafeHaven aims to detect the type of DV based on the victim's self-disclosure through keyword recognition in the user's input.
- **Risk assessment:** With the assistance of SafeHaven, the survivor will be able to assess their danger. Upon the survivor's completion of the assessment, SafeHaven will report the survivor's level of danger as well as the likelihood of subjecting to serious harm or even murder.

- Priority setting: It assists the survivor with setting their safety priorities on their abusive relationships.
- Referral resources: SafeHaven will direct the user to a variety of resources (e.g., housing, legal, family counseling, childcare, and financial assistance) that would help in meeting their tangible needs.
- Provide counseling tips: There is a rich library of information and guides concerning DV available to survivors through SafeHaven. Survivors can learn about their rights in the relationship through SafeHaven. Specifically, they will gain a better understanding of some of the major family and legal issues like the importance of gathering evidence and why evidence is crucial, divorce issues, protection orders and how to file it, and reporting abuse. Further, SafeHaven educates survivors about the types of evidence they should consider gathering after an incident of abuse has taken place, and what would be the follow-up after obtaining the evidence. Besides, the user will also learn about the signs and indicators of potential abuse. For instance, how can you tell if someone is being abused emotionally?
- Privacy and safety: Several measures have been adopted to enhance the survivor's privacy and safety. As an example, we have designed an escape feature that enables survivors to exit a chat session quickly in case others become aware that they are seeking information of this nature. The survivors can exit the chat by clicking on the exit button in the persistent menu at the bottom right or by typing "exit," which will bring up the exit button that will navigate them to a neutral site (e.g., Google.com). In addition, SafeHaven instructs survivors on how to clear or archive conversations whenever they are concerned about others being able to access their chat.

- Offer emotional support: SafeHaven provides emotional and social support in the form of preventing self-blame, emphasizing that it is not the survivors' fault, understanding the situation, and reframing it without judgment. Furthermore, to provide comfort to survivors, it sends supportive messages and sympathetic phrases periodically.
- Communicate in a simple language: SafeHaven provides simple and intuitive language that is easy to understand. Rather than relying on lengthy details, it utilizes short and summarized responses and employs words that everyone is familiar with, creating conversations in the form of one-to-one interactions. SafeHaven also enhances the readability of content by incorporating visual images and emojis.
- Offer retrieval modules: SafeHaven can recognize the survivor's inquiry for information or service, linking them to the appropriate module (submodule), and obtaining the relevant information.

### 4.3 Detailed Design

We have created a CA that translates design guidelines into practical forms. Our CA was developed using the ManyChat tool. ManyChat was chosen for several reasons. First, it allows users to build a CA on Facebook Messenger or Telegram, and the NLP in the system enables training information to be applied to a broader set of terms and phrases. With the popularity of Facebook Messenger and Telegram, ManyChat has opted to focus on it as it is one of the most common messaging platforms globally, with over 1 billion people using Messenger, and 700 million active users using Telegram [298], [299] . Additionally, this system enables the end user to practice different interactivity features, like typing, clicking buttons and links, browsing catalog carousels and categories, and sending and receiving messages to and from the bot. Further, at the design level, the user will be shown elements of text, buttons, carousel, image, and

other design features that may have a great deal of effect on the user's overall experience. ManyChat offers various automated and intelligent features like keyword recognition, personalization, a variety of automation options to cater to the needs of users, location detection, supporting multiple conversation paths, and the ability to redirect users to specific websites or places.

In this section, we provide detailed descriptions of each of the four main modules of SafeHaven: greeting and assuring safety, screening and identifying abuse, developing the safety plan, referral resources and survivor story.

#### 4.3.1 Greeting and Assuring Safety

SafeHaven greets the user, then introduces the purpose of this chat session, as well as how it aims to enhance the user's awareness of the issue of DV and provide assistance in responding to it. The welcome message also emphasizes confidentiality and privacy of conversation. It also asks the user what he/she prefers to be called. Next, the users is informed of some safety strategies to ensure their safety. These strategies are designed to protect users' privacy and minimize their risk of being identified or hacked by their abusers. For instance, the user learns how to exit the chat session and be directed to a neutral website (i.e., [www.google.com](http://www.google.com)). Further, the user will be informed about how to archive or delete chat conversations (see figure 4.2).

Following the welcome message and greeting, we designed a conversation in which the CA asks survivors about their safety and well-being, and whether they are available to communicate at that moment. If they are not safe, SafeHaven strongly recommends that the individual takes some safety precautions, such as calling 911, reaching out to close family and friends, or contacting an advocate of the National Domestic Violence Hotline (See figure 4.3). On the other hand, if the user ensures they are safe, SafeHaven then asks some journaling and opening

questions, which will help build rapport and relieve stress, as well as give a better understanding of the users' needs by understanding how they feel and what they do on a daily basis [187], [300].

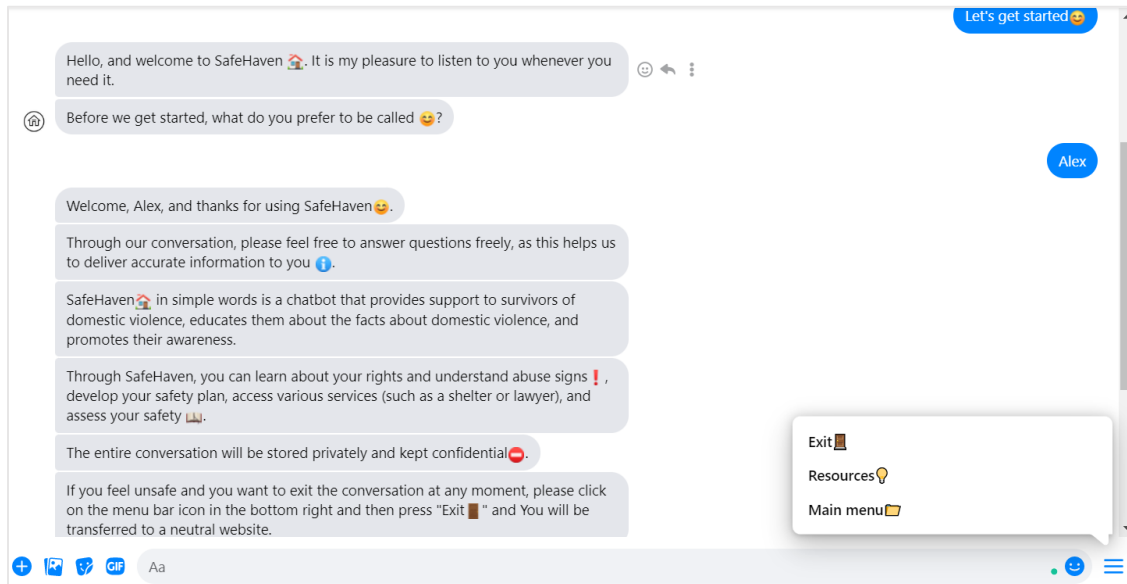


Figure 4.2: Welcoming message

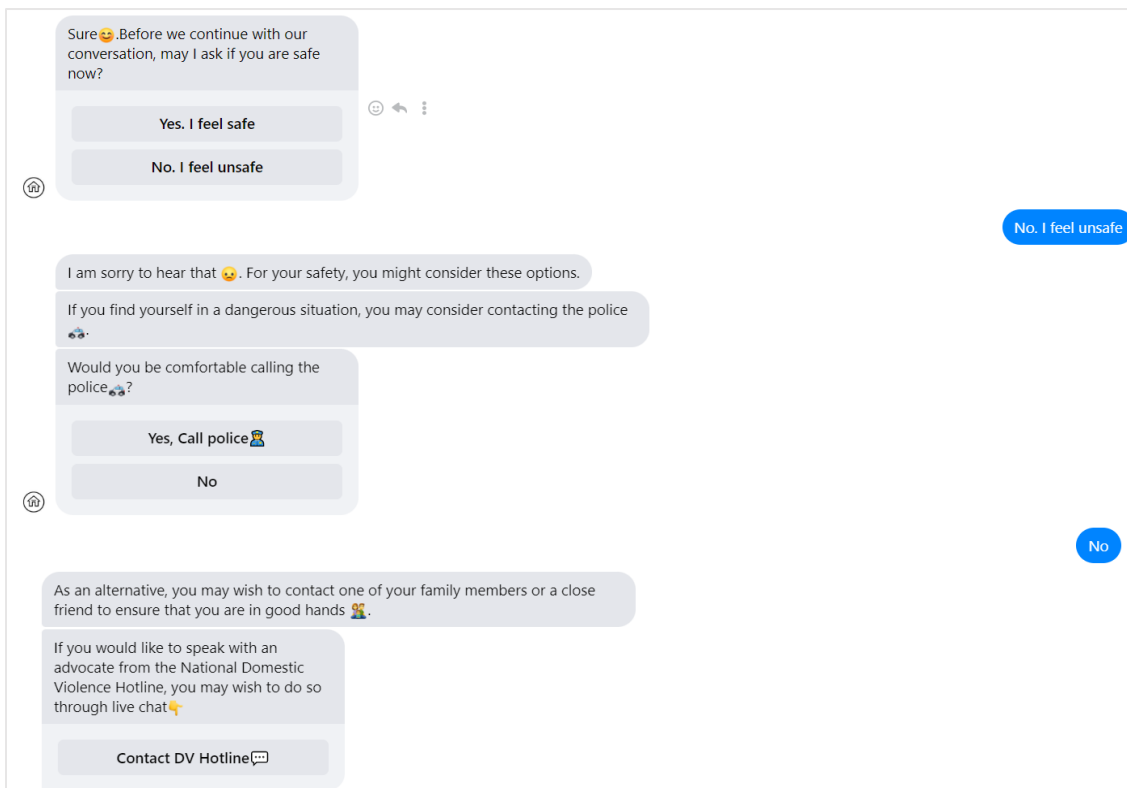


Figure 4.3: Assuring safety

#### 4.3.2 Screening and Abuse Identification

One of the consensuses from our interviewed domain experts is that the CA can serve as a starting point as well as a safe space for survivors to become more aware of their situations. Thus, we have developed a module that is focused on screening and abuse identification. In particular, the module assists the survivor in gaining a better understanding of the type or types of abuse that he or she has experienced, while also increasing their awareness by offering tips and prevention strategies to prevent future abuse. In the first step, SafeHaven asks about what happened to the survivor and their experiences (Figure 4.4). If the survivor is not interested in sharing their experience, they can enter "skip" in the input field to proceed with the development of their safety plan, or to access referral resources. In the event that the survivor writes about his/her experience, SafeHaven initially expresses empathy and offer emotional support, and then determines the type(s) of abuse (e.g., emotional abuse, physical abuse, or financial abuse) by using a keyword-based method and outlining their various forms and how they relate to other abuse types (Figure 4.5). Afterward, SafeHaven ensures the user's safety and checks whether they are exposed to any high-risk situations or dangers. When a user indicates a need for immediate assistance, SafeHaven offers them a few safety tips and then connects them to the appropriate first responders for further assistance. For example, SafeHaven suggests that they may call the police, find a hospital near them, or contact the National Domestic Violence Hotline. If the user indicates they are safe, SafeHaven provides them with prevention strategies and tips on how to prevent such abusive actions in the future. As an example, it recommends that they inform a trusted person of the abuse incident, develop a safety plan, identify support services, and so on (Figure 4.6).



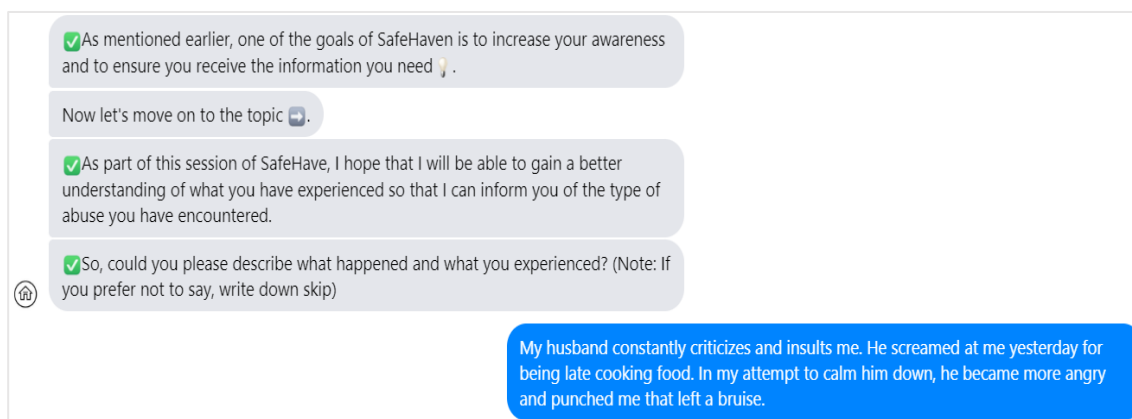


Figure 4.4: Abuse screening

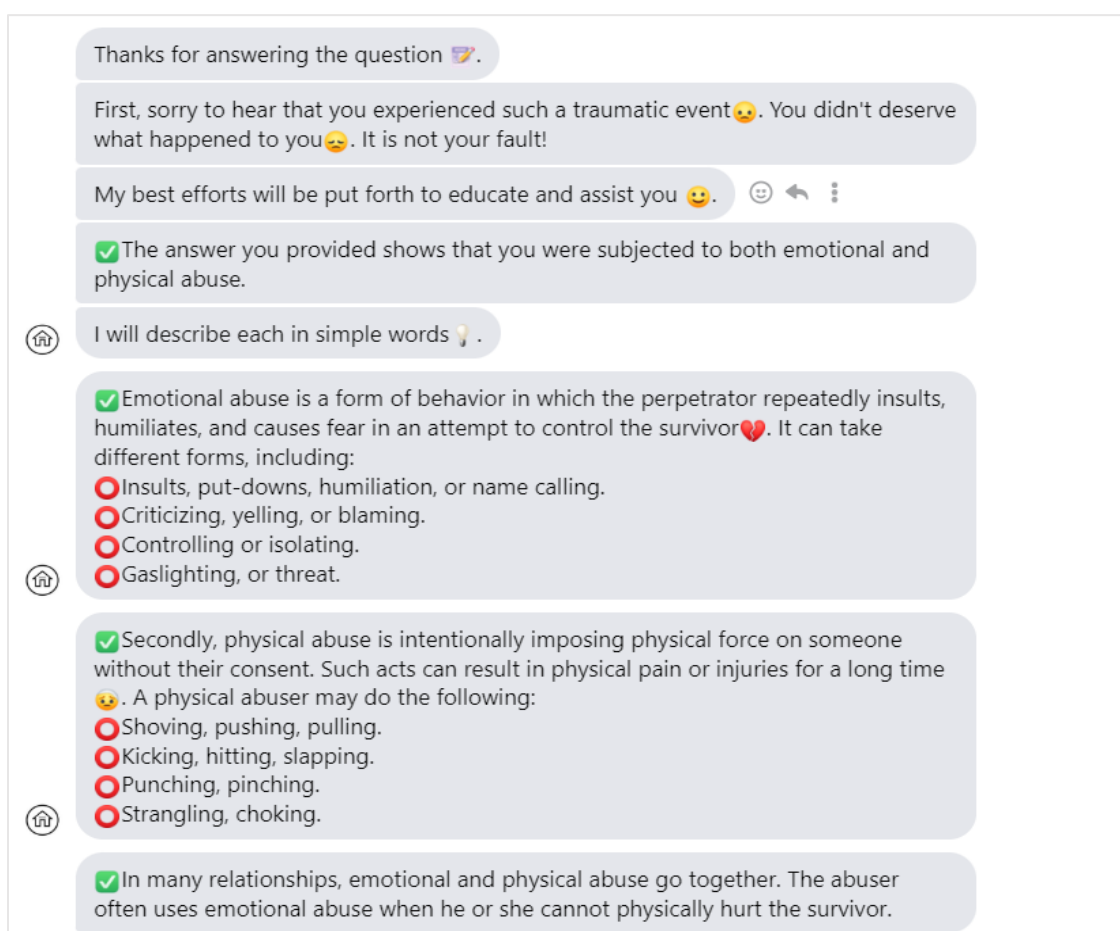


Figure 4.5: Abuse identification

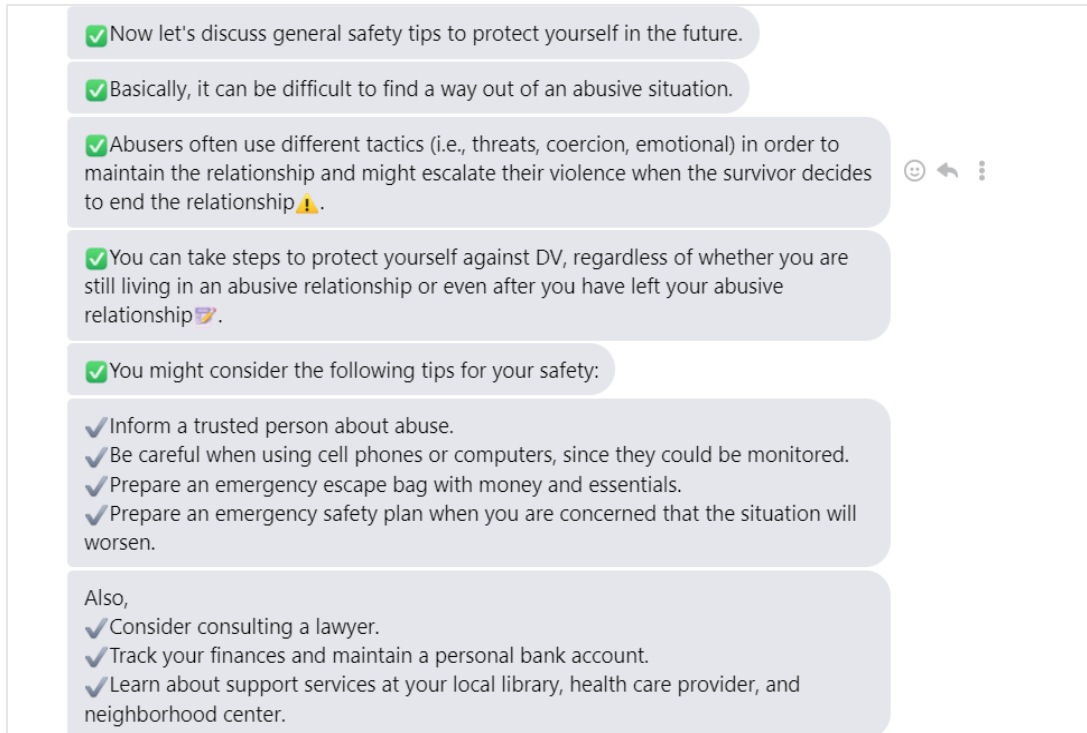


Figure 4.6: General safety tips

#### 4.3.3 Safety Plan Development

This module provides the user with the opportunity to develop a safety plan. In simple words, a safety plan is a series of steps designed to reduce the possibility that a survivor will be negatively affected by their abusers. Throughout this guide, they will find relevant information about their family and life that will allow them to improve their safety when dealing with their children and their partner, or when staying at home or in other places where they spend time on a daily basis. A survivor can use this plan to make decisions in advance, regardless of how much trauma they are experiencing. As they continuously confront any abusive actions, they have clearly outlined the steps that they can take.

In our design, we simulated the interactive safety plan from the National Domestic Violence Hotline (<https://www.thehotline.org>). The plan consists of several sub-phases, including ensuring the safety of survivors at home, at school, with their children, and when they interact with their

partners. For instance, the bot asks them where they could go for safety if they feel uncomfortable at home, or what items (i.e., ID, cash, passport) they would take with them if they had to leave home rapidly. In addition, when they are asked about the safety of their children, they are asked what code might be shared with their children if they feel insecure. As part of the process, the user is asked a series of questions that will guide them in determining their safety options (see figure 4.7). Upon completion of all questions, they can review their personalized safety plan in a structured-looking manner (see figure 4.8). The CA presents answers to the survivors and informs them that they may copy, screenshot, or print them for later reference. Lastly, as it is important to note, SafeHaven recommends that users include accurate and comprehensive information in their safety plan, maintain the safety plan in a safe and accessible location, and give a copy of the safety plan to someone they trust.

Here are questions to help you organize your plan. 📝 🗑️ ⋮

First, has anyone in your family been aware of your relationship? 🗑️

No

What phrase or word can you use as a code in a phone call or text message to ask your family, friends, or neighbors for help without your partner's knowledge? 🗑️

I wish it would be sunny!

When you're alone at home, who can you stay with? 🏠? (Keep this person's phone number on your emergency card.) 🗑️

ALAN, Phone #: (123)456-7890

Where else could you go for safety if you're uncomfortable at home? 🗑️

To my friend's home (Adam)

What items would you take with you if you have to leave your home quickly (i.e., your ID card, cell phone 📱, debit card 💳, car key 🗑️)? 🗑️

ID, Cash, Laptop, cell phone, phone charger

Figure 4.7: Developing a safety plan

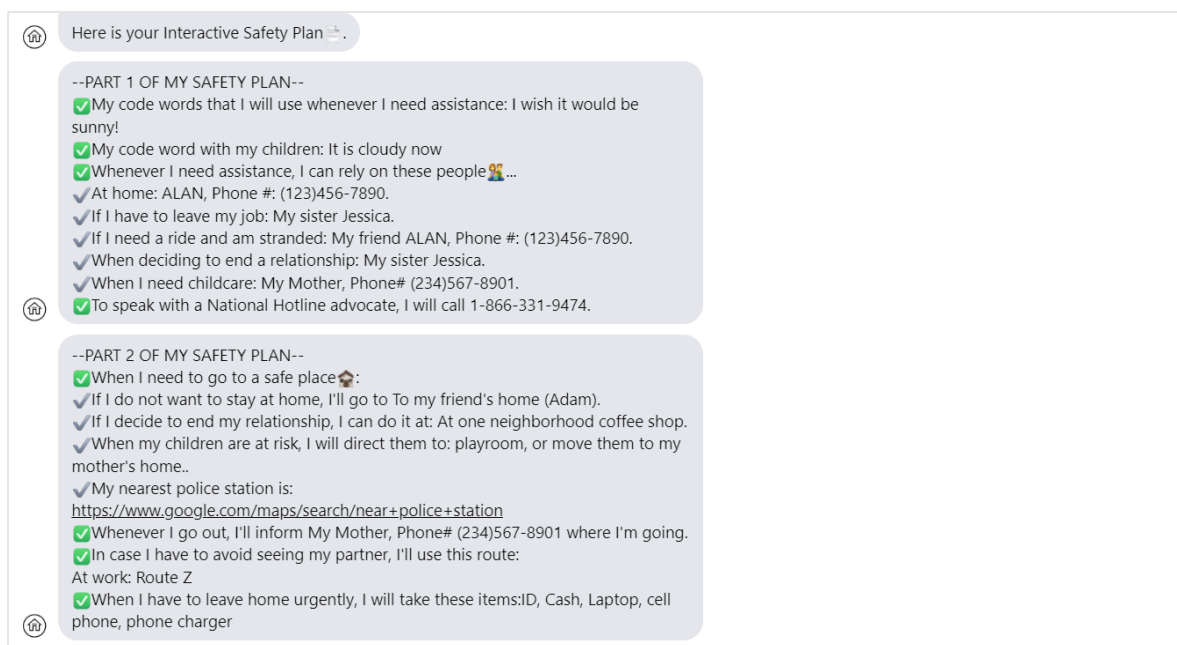


Figure 4.8: Personalized safety plan

#### 4.3.4 Survivors' Stories

This module discusses the stories of DV survivors and how they coped with challenging circumstances. These narratives offer valuable insights into the direct experiences of survivors, aiming to shed light on different aspects of situations of DV. This module describes strategies to assist survivors on how they may engage with local domestic violence programs, elicit assistance from their social networks, or seek legal counsel, thereby fostering an environment that facilitates empowerment and positive change. (Figure 4.9).

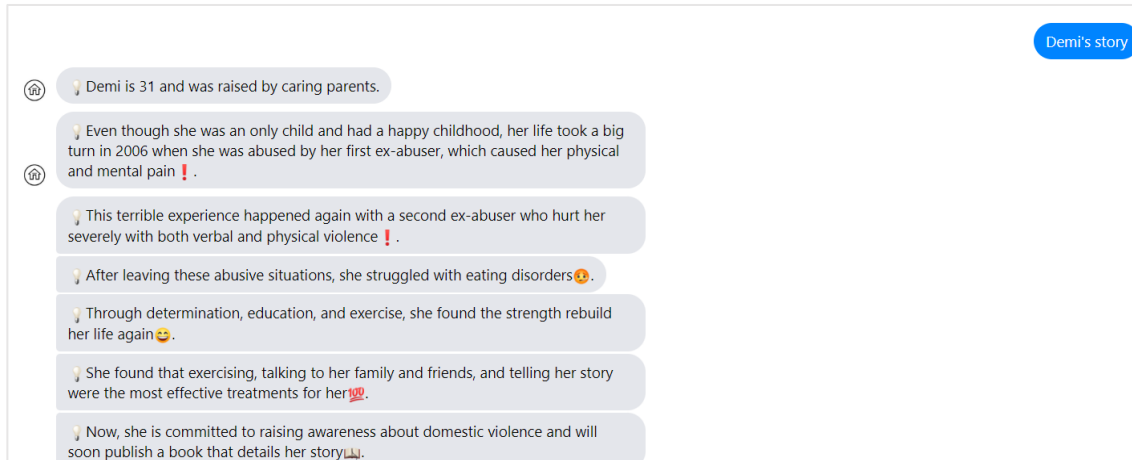


Figure 4.9: An example of the survivor's story

#### 4.3.5 Referral Resources

To develop the resources module, we created a list of DV support services. In particular, we selected seven main types of support resources, including local DV services, family services, legal services, medical and mental healthcare services, national crisis organizations, financial aid resources, and informational resources. In designing SafeHaven, we incorporated several resources from national and non-profit organizations' websites that are dedicated to helping DV survivors. Examples of these organizations/services are The National Domestic Violence Hotline, The National Coalition Against Domestic Violence, The National Network to End Domestic Violence, WomensLaw project (Womenslaw.org), 211.org, DomesticShelters.org, etc. It is essential to have these facilities available to assist survivors throughout the entire process of addressing their incidents. Survivors can explore a list of available support services and the options for each type of the service, and then choose an option from a carousel menu under a specific type of support service (See figure 4.10). To fit into the space of CA's buttons, some support service options are labeled with their shorthand notations (i.e., NNEDV instead of National Network to End Domestic Violence). After the survivor has reviewed an option, the CA

allows them to continue to explore other referral or informational resource options or navigate back to the main menu. We will describe each type of support service in detail next.

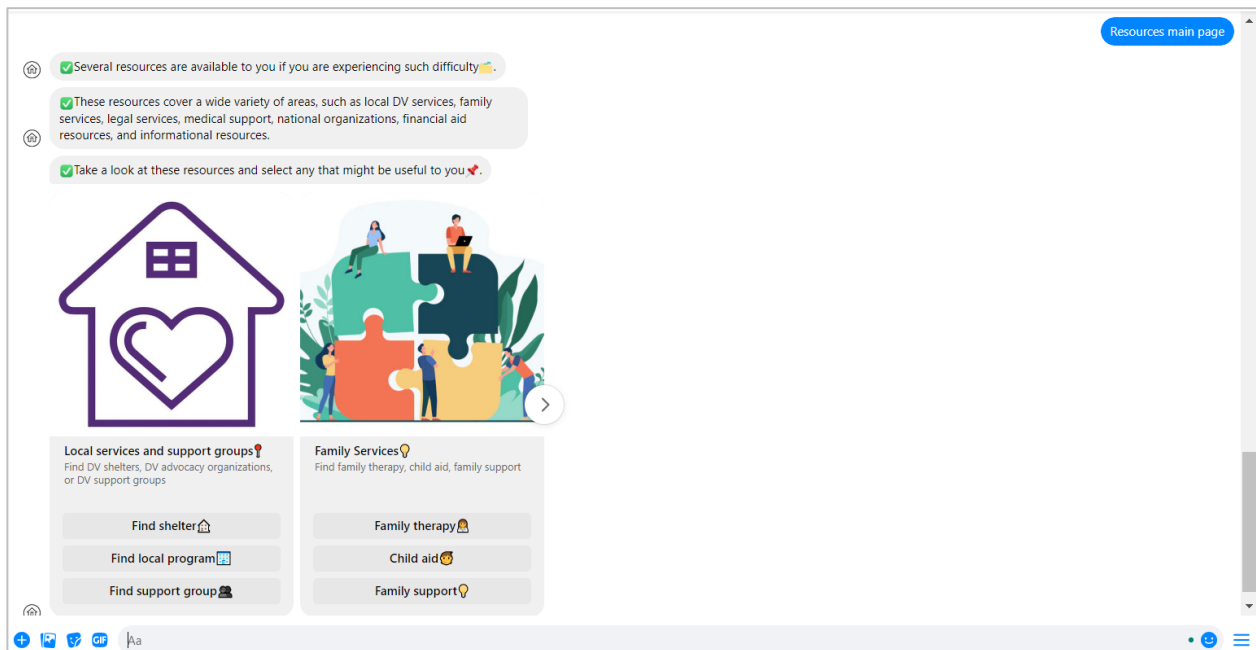


Figure 4.10: Resources module

#### 4.3.5.1 Local Services and Support Group

The local services comprise three main categories: finding a DV shelter, finding an agency/local program providing domestic violence services, and finding a domestic violence support group.

- Find a domestic violence shelter: This submodule is designed to assist survivors in locating domestic violence shelters in their area. First, survivors learn how shelter contributes to their safety and the safety of their belongings. SafeHaven then recommends that survivors take into consideration several factors before leaving home and seeking shelter. This involves knowing when to leave, where to go, and what to take, as well as how to keep the details of your plan secret from your abuser. Following that, SafeHaven asks the survivor to share their location to determine a

nearby shelter (see figure 4.11). Upon sharing the location, SafeHaven refers survivors to the Domesticshelters.org website, where they can find a list of nearby shelters with detailed information, including contact information and service offerings (see figure 4.12).

- Find a local domestic violence program: SafeHaven would connect the survivor with local programs and agencies that provide support services. As a preliminary step to sharing local programs, SafeHaven will explain the importance of local programs and how professionals' recommendations and referrals can potentially mitigate the effects of the DV. Following that, SafeHaven will ask the survivor to share their location. Once the location is shared, it would refer them to the Domesticshelters.org website, which would display a list of all available programs in the city they live in. As a further recommendation, SafeHaven would also refer the survivor to another website, The National Domestic Violence Hotline, which provides an updated list of services and information about the local programs in the survivor's area.

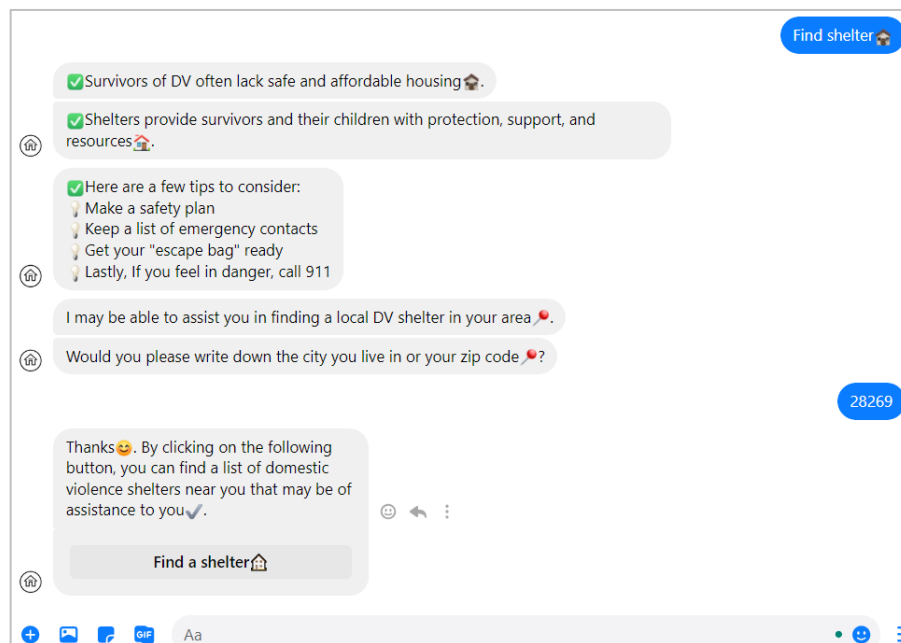


Figure 4.11: Find a shelter

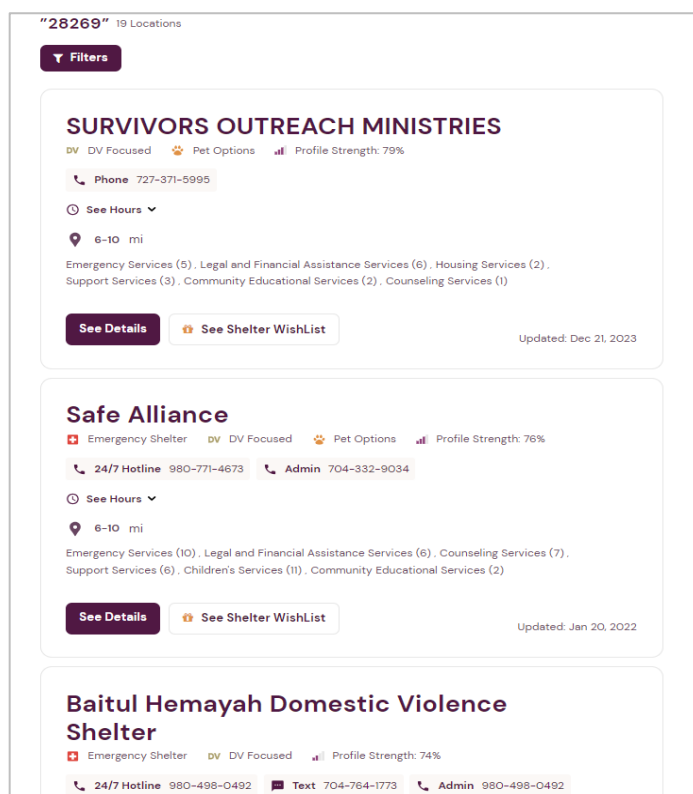


Figure 4.12: List of nearby shelters (Domesticshelters.org)

- Find a domestic violence support group: Communication not only allows individuals to make sense of their experiences, but also provides support, acceptance, and reassurance from others who have experienced similar incidents. In light of this concept, SafeHaven assists survivors by connecting them to different online DV support groups, including social media groups. Firstly, SafeHaven discusses the importance and benefits of participating in such groups as a means of alleviating stress and isolation, as well as gaining empowerment. Then, it offers several options of online support groups that the survivor might reach out to, such as DomesticShelters, Fort Refuge, Hope Recovery, and Love Is Respect. In addition, it also refers them to some social media groups that concentrate on supporting DV survivors like Break The



Silence Against Domestic Violence (BTSADV) group on Facebook, and “r/domesticviolence” subreddit on Reddit.

#### 4.3.5.2 Family Services

Under this category, we have designed three sub-modules with the potential to support survivors of DV with family-related services, including family therapy, child aid, and family support.

- **Family therapy:** One of the most important features we implemented in the CA is the capability to refer users to a family therapist. The purpose of this sub-module is to assist survivors in finding family therapists in their local area. Initially, SafeHaven emphasizes the critical role that a family therapist plays and how family counseling can benefit the survivor and their family. Also, the survivor can learn what point to seek the assistance of a family therapist. In addition, SafeHaven also offers recommendations on how to find a therapist (i.e., consult your primary doctor, and research their degrees and credentials). As a final step, the CA would ask the survivor for their location to locate a family therapist in their area (figure 4.13). Following the sharing of the location, the CA would use the shared location to direct the user to the website of the American Association for Marriage and Family Therapy, where they could find a list of family therapists in their area (figure 4.14).
- **Child aid:** As part of this submodule, SafeHaven introduces the risk of child maltreatment as well as its physical and psychological consequences on children and their families. Additionally, SafeHaven provides advice to survivors, especially if they are parents, regarding how to handle this type of abuse. By way of example, the CA suggests that parents learn how their children behave, educate them about their rights,

and report abuse if necessary. The CA then refers the survivor to the ChildHelp hotline, where they can learn about child abuse, reporting abuse, parenting styles, etc. In addition, the CA offers a referral option to the Child Welfare Information Gateway website, which provides awareness programs and resources geared toward working with children and adolescents suffering from DV.

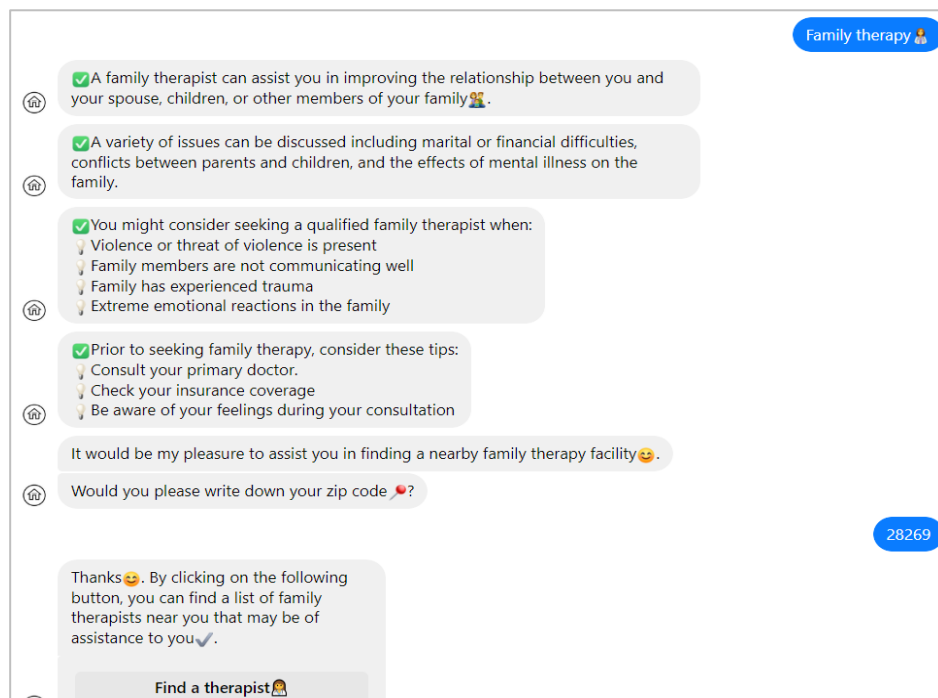


Figure 4.13: Find a family therapist

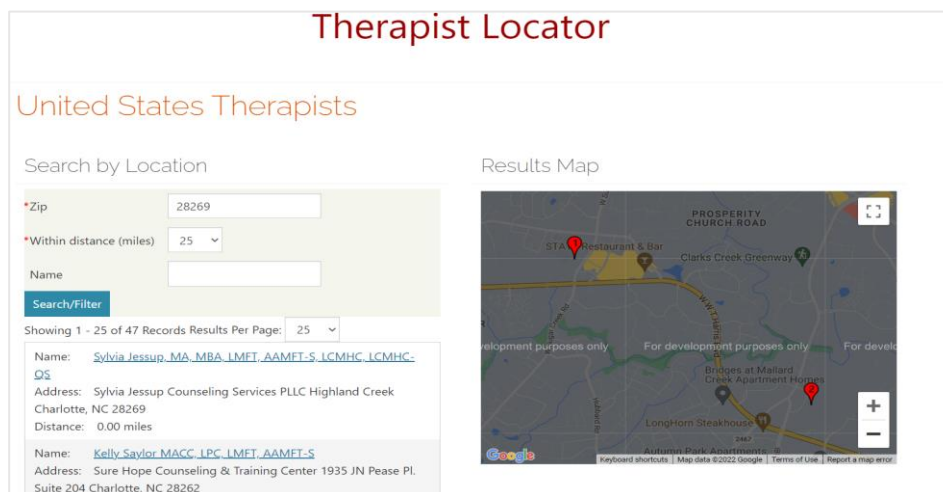


Figure 4.14: List of family therapists (aamft.org)

- **Family support:** This submodule explores the role of family support and preservation programs and how they contribute to the continuum of family-centered services targeted at strengthening and supporting families. In this regard, the CA refers the survivor to a couple of options that offer these supportive programs, including 211.org, the National Family Support Network (NFSN), the National Parent Helpline, and Just in Time Parenting.

#### 4.3.5.3 Legal Services

In this category, we have designed three sub-modules that are intended to support survivors of DV with legal-related issues, namely finding a lawyer, law help, and finding the nearest courthouse.

- **Finding a lawyer:** In the legal context, one of the survivor's most important needs is to find a lawyer. In this regard, SafeHaven provides a service that refers the survivor to a lawyer. In its initial message, SafeHaven emphasizes the importance of a lawyer in DV cases and demonstrates how a lawyer's experience and knowledge can benefit the survivor from a variety of angles, including defending the survivor during the case, knowing legal issues and court procedures, and finally, limiting the continuation of violence and mitigating its consequences. Following this, it asks about the survivor's location to locate a lawyer in the area. After receiving the shared location, it will direct the user to the FindLaw website, where they will be able to find a list of law firms specializing in DV in their area. In addition, SafeHaven would also provide a referral to the WomensLaw website, where survivors can find a listing of legal assistance providers in their state.

- **Law Help:** This sub-module was designed as a means of raising awareness regarding legal matters. In particular, the survivor will be referred to a state LawHelp guide where they can identify links to legal aid and public interest law offices, basic legal information, court forms, getting help from a lawyer, self-advocacy tools, and other social services information in their states. In addition, it also assists survivors who cannot afford a lawyer to obtain free legal assistance through legal aid offices.
- **Find a courthouse:** The courthouse is a crucial resource for survivors when filing a protection order, custody, divorce, or any other legal matter. In this sub-module, SafeHaven discusses the role of courthouses in DV issues and assists the survivor in locating a courthouse in their county. In the beginning, it touches on the importance of the court and how it contributes to many issues related to DV, such as child custody, divorce cases, protection orders, parental rights, and others. Additionally, it outlines a few safety tips that may be helpful to the people who are in the courthouse, such as staying with family members or friends and being close to the security guards. In conclusion, SafeHaven asks survivors to share their location so they can locate courthouses in their area (Figure 4.15). Once the survivor shares the location, it will refer them to the Women'sLaw website, where they can locate the courthouse in their county (Figure 4.16).

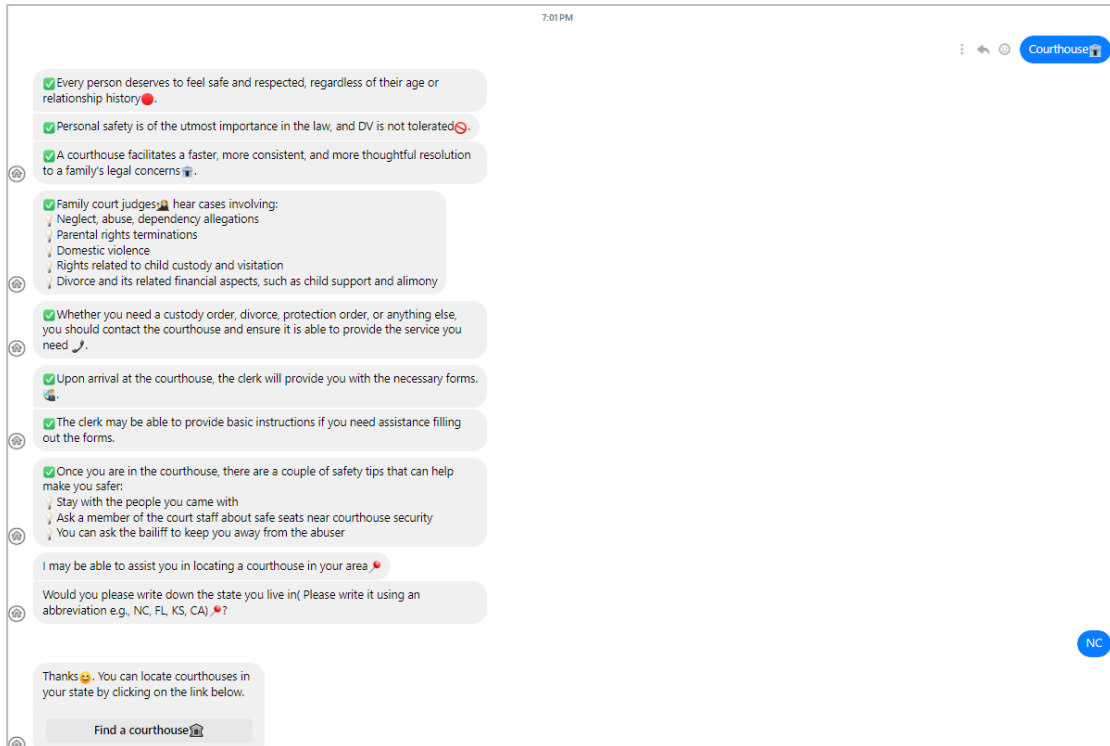


Figure 4.15: Find a courthouse



Figure 4.16: List of courthouses locations (WomensLaw.org)

#### 4.3.5.4 Medical and Mental Health Care Services

We identified three main resources that can provide physical and mental health services to DV survivors, including finding a doctor and hospital, mental health services, and disaster distress.

- Find a doctor/ hospital: Due to the possibility of DV-induced physical and mental harm, finding a doctor or hospital for the survivor is an invaluable function of SafeHaven. This sub-module aims to assist the survivor in locating a doctor or hospital in their area. In its initial discussion, SafeHaven emphasizes the critical role of healthcare professionals and the fact that they have a unique ability to identify problems (i.e., DV that may not be obvious at first. It also discusses how healthcare professionals may be able to assist DV survivors in several ways, including, but not limited to, providing general and individualized screenings as well as maintaining patient confidentiality (Figure 4.17). Following this, it offers a few tips on how to choose healthcare providers that could meet the survivor's needs. Then, it asks the survivor to share the information of their location to help locate a doctor or hospital in their area. It will then ask them whether they want to find a doctor or hospital. If they choose to look up a doctor, then they can specify what kind of condition/specialty they are looking for (e.g., orthopedic, cardiology) (Figure 4.18). The CA then uses both the shared location and user condition and then refers the survivor to the US News Health website, where they can access a list of doctors specializing in the condition they are looking for in their area (Figure 4.19). On the other hand, if survivors are looking for a hospital, they will be referred again to the US News Health website to find a list of nearby hospitals.

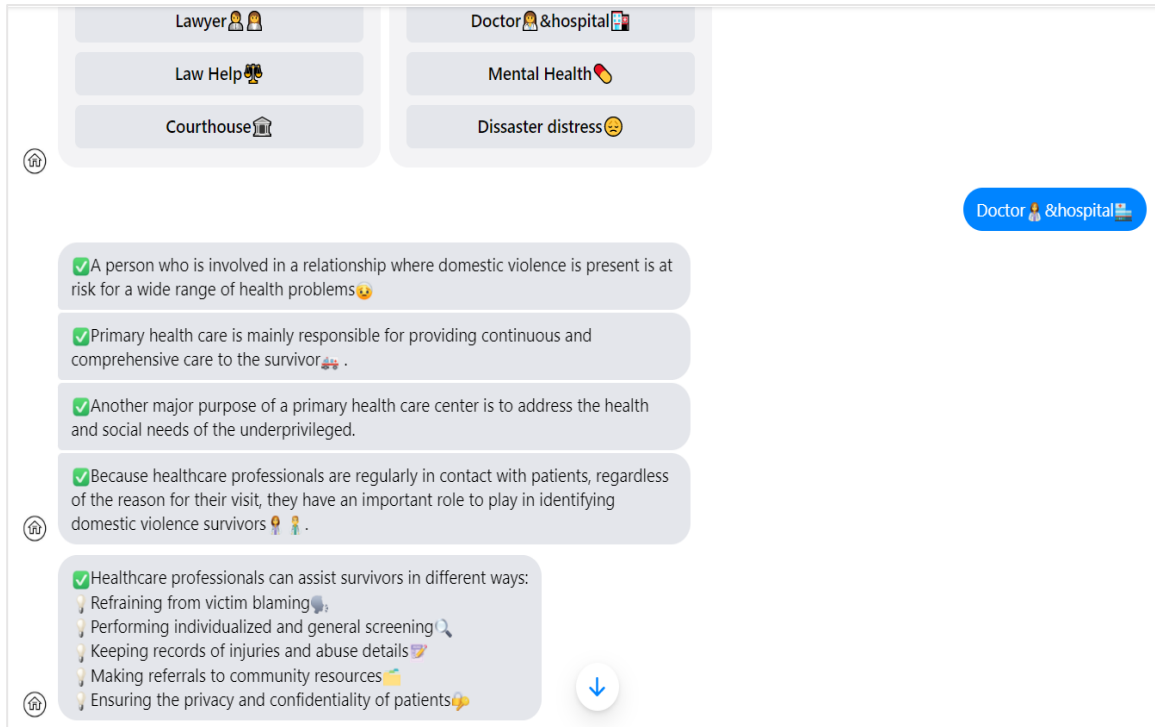


Figure 4.17: Find a doctor/hospital

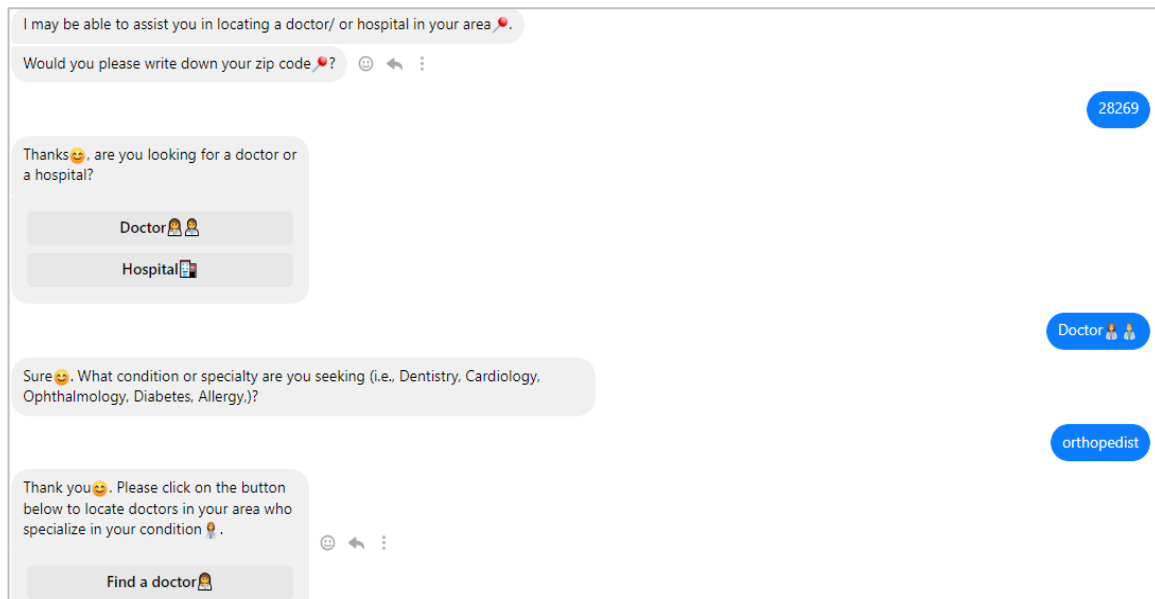


Figure 4.18: Find a doctor (Sharing the location and providing the specialty)

The screenshot shows the US News Health website's doctor search results for 'Orthopedists in 28269'. The header includes navigation links like 'HEALTH', 'Hospitals', 'Doctors', etc., and a search bar. Below the header, the page title is 'Orthopedists in 28269'. A blue banner encourages users to 'Learn more about Orthopedists and find the best doctor for you'. Filter buttons show 'Specialty: orthopedist', 'Location: 28269', and 'Name'. It indicates '293 Doctors' are found. Two doctor profiles are visible: Dr. Abhishek Ganta MD (Sponsored, 6-10 Years of Experience, 1 Language, Accepting New Patients) and Dr. Jonathan J. Paul MD (21+ Years of Experience, 1 Language). Both are Orthopaedic Surgeons. Dr. Ganta's profile includes a 'Book Appointment' button.

Figure 4.19: List of doctors (the US News Health, <https://health.usnews.com/>)

- Mental health:** We learned from our interview study that DV has a negative and severe impact on the lives of survivors, particularly on their mental health. DV can exacerbate mental health problems in survivors, including depression, anxiety, and PTSD. In light of this fact, we have developed a sub-module aimed at educating survivors about mental health, available treatment options, and various referrals that can contribute to alleviating the severity of this problem. SafeHaven highlights first the grave consequences of DV and its impact on mental health. Afterward, SafeHaven suggests a variety of resources that might be of help to the survivor. For example, it refers to a tool developed by the Substance Abuse and Mental Health Services Administration (SAMHSA) that helps individuals locate behavioral health treatment facilities for psychiatric illnesses. Likewise, SafeHaven directs survivors to the Mental Health America website, which provides information on the appropriate treatment



options for them, how to obtain help (local health departments, marriage, and family counselors), and which mental health professional is most appropriate (psychiatrist, occupational therapist).

- **Disaster distress:** During a time of emergency and disaster (such as a DV incident), survivors are encouraged to maintain their emotional wellness to protect themselves and their families. This type of service will help survivors learn about the risks of distress and get suggestions on how to mitigate the impact of such a risk. Specifically, SafeHaven offers some tips on how to deal with distress following a disaster. In addition, it refers the survivor to the Disaster Distress Helpline (DDH), which provides immediate professional assistance to anyone suffering from emotional distress due to a disaster, whether it is natural or caused by human action.

#### 4.3.5.5 National Crisis Organizations and Aid

The national level of services includes national organizations that offer assistance to survivors of DV. We mainly consider three resources in our design, including the National Domestic Violence Hotline, the National Network to End Domestic Violence (NNEDV), and the LoveIsRespect project.

- **The National Domestic Violence Hotline:** The purpose of this sub-module is to refer the survivor to the National DV Hotline. It is an important referral resource that may be highly useful to the survivor in obtaining immediate assistance. SafeHaven describes some of the resources that the hotline offers, such as making a safety plan, searching for local resources and agencies, identifying legal resources available in the survivor's community, learning about healthcare services, and being able to locate providers.

- The National Network to End Domestic Violence (NNEDV): It is the purpose of this organization to produce social, economic, and political change in order to eliminate DV. As part of this sub-module, survivors will learn about some NNEDV projects that may benefit them. To illustrate, SafeHaven refers a survivor to the WomenLaw project, which provides legal information, particularly for those representing themselves in court, safety advice, tips for working with a lawyer, and much more. Furthermore, it makes a referral to the SafetyNet project, which provides survivors with technology safety tips, information, and privacy strategies.
- Love Is Respect: It is part of the National Domestic Violence Hotline's project that aims at preventing problematic relationships and intimate partner violence by providing education, support, and resources. In this submodule, we provide the survivor with an opportunity to explore what resources and supportive services are offered by this project. SafeHaven briefly describes a few of the project's resources, including safety planning, understanding abuse signs, finding local resources, learning tips for a relationship, and learning safety tips at home, work, and online.

#### 4.3.5.6 Financial and In-Kind Resources

This sub-module is designed to connect survivors with different funding and in-kind resources, which may have a great deal of value for survivors, especially during a tragic event. We have designed three sub-modules under this category, including financial resources, food resources, and healthcare resources.

- Financial resources: Given the fact that DV can have a significant financial impact on survivors, it is important that SafeHaven provides survivors with access to financial resources and raises their awareness of financial issues emanating from DV. For this

purpose, we designed four options for financial aid covering housing and utilities, financial education, crime victim compensation, and student assistance (Figure 4.20).



Figure 4.20: Financial resources

- **Housing and utilities.** As part of this sub-module, survivors are able to access a variety of housing assistance resources that will assist them with housing matters such as paying rent and bills. As a first step, the CA checks with the survivor to determine whether they require temporary housing. In this case, they are referred to the DomesticShelters.org website, which provides information on safe temporary housing in their area. Otherwise, SafeHaven provides access to a range of housing assistance services, including rental assistance, mortgage assistance, general utility assistance, and LIHEAP assistance.

Under the option of rental assistance, SafeHaven describes various programs and resources that offer rental assistance funds or can help the tenant negotiate a payment plan with their landlord. By way of example, it refers to a tool developed by the Consumer Financial Protection Bureau that assists

individuals in searching for local rental assistance programs in their area. Furthermore, the CA also introduces the survivor to the Emergency Housing Voucher program offered by the American Rescue Plan Act (ARPA) and designed to assist families and individuals at risk (i.e., DV survivors).

Under the housing and utility assistance sub-module, SafeHaven also provides mortgage assistance. In this option, the CA refers the survivor to several resources and provides tips to assist them in managing their mortgage payments. For example, it suggests that they contact their lender or loan servicer directly to discuss foreclosure prevention options. It also shares a link to the Department of Housing and Urban Development website, where they can obtain additional information about their options. SafeHaven also recommends that survivors contact 2-1-1 to speak with a trained call specialist about local mortgage assistance programs.

General utility assistance is another option that is designed to aid survivors who have difficulty paying their bills on time and allow them to access different utility assistance programs. A variety of utility assistance programs for internet, phone, and water services are shared. For instance, SafeHaven refers the survivor to the Emergency Broadband Benefit Program, which is funded by the Federal Communications Commission (FCC) and offers discounts for broadband services to eligible households. In addition, it refers them to the Low-Income Household Water Assistance Program (LIHWAP), offered by FCC, which provides benefits for low-income households experiencing water bill hardship.

SafeHaven also makes a referral to the Low-Income Home Energy Assistance Program (LIHEAP), which offers support with utility bills and energy costs for low-income households. It explains what this program offers and then connects the survivor to the program's contact information and website.

- Financial education: One of the most devastating effects of DV is the loss of financial resources. It is common for DV survivors to suffer from financial hardships or be unable to manage their finances, which continuously places them at risk. Taking this into account, this option was designed to allow survivors to learn about several financial education programs and gain a better understanding of how to handle their finances, understand credit, and save money. As an example, SafeHaven refers the survivor to the Financial Education Webinar Series developed by the National Coalition Against Domestic Violence (NCADV), which introduces a variety of financial literacy skills survivors should learn, such as basic money management skills, budgeting, planning for holidays, and setting financial goals.
- Crime victim compensation: Some DV effects may be associated with physical or mental health, while others may be associated with hardship on the financial front. This option highlights the crime victim compensation program. As SafeHaven begins by explaining what the program is about, it goes on to explain what costs may be covered by the program. Furthermore, it emphasizes that the survivor should report the crime to the police and cooperate with them, as well as submit a compensation application in a timely manner. As a final step, it directs them to the website of the National Association of Crime Victim

Compensation Boards, where they can review contact information for the victim compensation programs in their respective states.

- Student assistance: One key aspect that survivors might need toward their financial stability is to obtain a degree or a special certification. Therefore, we have designed an option that identifies educational resources available to DV survivors who wish to pursue further education. SafeHaven lists some educational programs that can help the survivor receive a certificate. For example, it refers them to the Sunshine Lady Foundation, a program that offers assistance to women who have suffered and survived mistreatment in the home. Furthermore, as part of the CA's recommendations, it refers survivors to the Women's Independence Scholarship Program (WISP), which helps battered women obtain the education necessary to become employable and financially secure, and in the long run, they may become advocate of supporting DV survivors.
- Food resources: Researchers have found that food insecurity is associated with violence according to three causal factors [301], [302]. First, food insecurity can arise as a result of financial abuse. Second, individuals fleeing violent situations may depend disproportionately on financial aid or low-wage employment to survive. As a result, these individuals might be unable to obtain food on their own. Third, there is a possibility that a food-insecure environment might also contribute to increased violence. Poverty is strong evidence of this. Both IPV and food insecurity are associated with poverty [301], [303]. As outlined by our interviewees, and in acknowledgment of its importance, we designed this submodule to assist survivors and

their families in ensuring their food security through referrals to several resources and programs that offer food assistance. As part of the food resources sub-module, we designed four options, including finding a food pantry, the Women, Infants, and Children Program (WIC), and the National School Lunch Program (NSLP), the Supplemental Nutrition Assistance Program (SNAP).

- Find a food pantry: The purpose of this option is to help survivors find a food pantry in their local area. SafeHaven first discusses the role of food pantries and how helpful they are in providing nutritious food to those in need as well as educating the community about nutrition and life skills to improve the life quality of the community at large. Following this, SafeHaven asks the survivor for their location to locate a nearby food pantry (Figure 4.21). Upon sharing the location, SafeHaven directs them to the [findhelp.org](http://findhelp.org) website, which lists food pantries in the area where they reside (Figure 4.22). As an additional referral, SafeHaven also refers the survivor to the Feeding America website, a nonprofit organization that operates over 200 food banks nationwide and provides food assistance to over 46 million people through food pantries, shelters, and other community-based organizations.
- Supplemental Nutrition Assistance Program (SNAP): This option was designed to refer the survivor to the SNAP program. SafeHaven first briefly describes this program and what it offers, as well as refers the survivor to the SNAP eligibility requirement webpage to determine whether they are qualified. SafeHaven then directs the survivor to the SNAP state directory, where they can obtain information on SNAP's local offices.

- Women, Infants, and Children Program (WIC): It is possible that abuse and its impact may not be limited to the survivor alone but extended to other members of the family (e.g., children). In some cases, the impact can be financial, which may lead to food insecurity. In light of this point, we designed this option to refer the survivor to the WIC program, which provides healthcare and nutrition needs to low-income pregnant women and breastfeeding mothers, as well as children under the age of 5. As a first step, SafeHaven describes the program and the services it provides. Furthermore, SafeHaven provides information on how to apply for the WIC program and provides a toll-free number for WIC. SafeHaven also provides a referral to WIC's PreScreening Tool, which helps applicants determine their eligibility for WIC. Finally, SafeHaven asks the survivor for their location to locate a WIC program in their area. Once SafeHaven gets the location, it will refer them to the [wicprograms.org](http://wicprograms.org) website, where they can find a list of WIC clinics, offices, and programs in their area.
- The National School Lunch Program (NSLP): In addition to the survivor, their children might also be at risk of being abused. A recent study suggests that persistently food-insecure households are six times more likely to expose their children to violence and/or victimization in their homes than food-secure households [304]. Providing food security to children is a crucial component of their well-being. With this in mind, we designed this option to assist the survivor in ensuring food security for their children. In this option, SafeHaven introduces the National School Lunch Program (NSLP), which provides nutritious meals to children as part of their regular school day. As a first step,



SafeHaven provides a brief description of the NSLP program. Moreover, it discusses the eligibility requirements for the program and how to apply for it. Lastly, SafeHaven refers the survivor to the NSLP website to find out more information about the program.

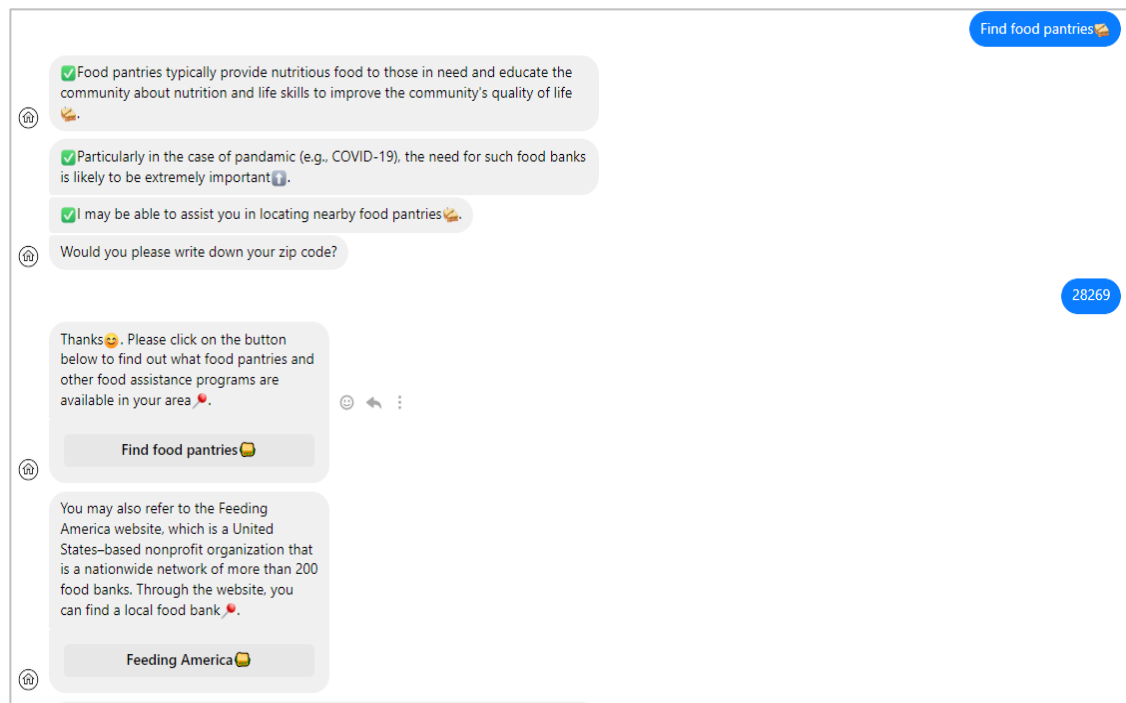


Figure 4.21: Find a food pantry

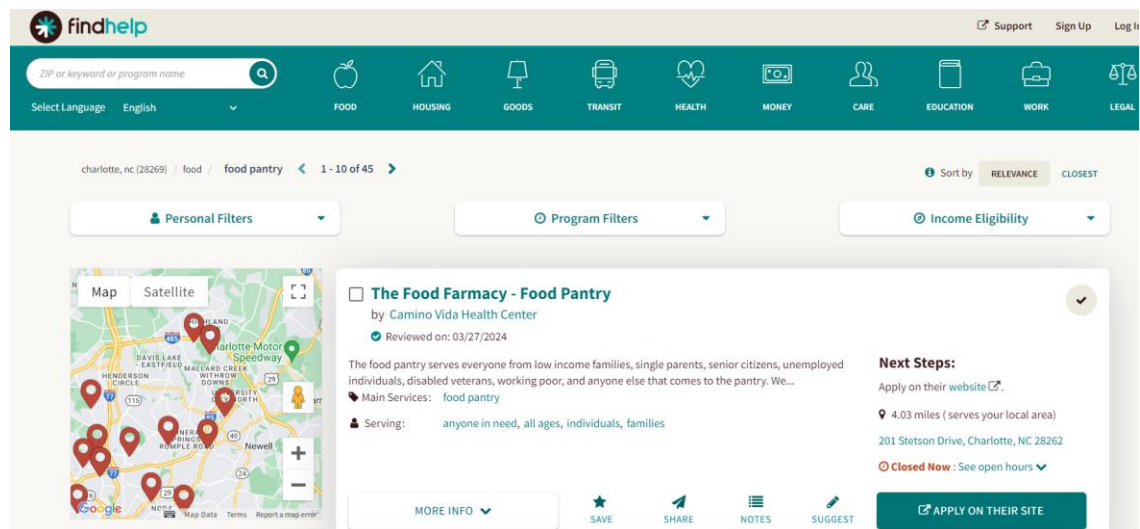


Figure 4.22: List of nearby food pantries (findhelp.org)

- Healthcare resources: Survivors may find it challenging to navigate the healthcare system, particularly when bills begin to pile up. There may be a need to pay for a prescription, visit an emergency room, or find low-cost mental health treatment. Thus, we have designed this sub-module to assist the survivor in discovering programs and resources that may be of assistance with their healthcare expenses. In particular, we have designed five options, including health insurance aid, low-cost healthcare, Medicaid and CHIP, medication payment, and mental health (MH) and substance use disorders (SUD).
  - Health insurance aid: An essential item that survivors and their families might need to obtain essential medical care, preventive care, and prescription medications is health insurance. In this option, SafeHaven offers the survivor different solutions so they can find health insurance plans. For example, it refers them to an affordable marketplace offered by the government (Healthcare.gov) where they can find health insurance plans in case that they are not able to obtain health insurance through their employers. Furthermore, it also introduces them to the Consolidated Omnibus Budget Reconciliation Act (COBRA), which allows employees and their families to continue to receive group health insurance benefits for a limited time under certain circumstances such as death, divorce, and other life events.
  - Low-cost healthcare: Even with health insurance, some plans may not provide adequate coverage for all healthcare needs, or it may be difficult for survivors to find a healthcare provider in their area who will accept their policy. In this option, SafeHaven helps the survivor find low-cost healthcare. For example, it

assists the survivor in locating a nearby healthcare center that offers medical services based on their financial ability through a referral to the Health Resources and Services Administration (HRSA) website, which is dedicated to improving access to healthcare for people without insurance, isolated, or suffering from medical conditions (Figure 4.23, 4.24).

- Medicaid and CHIP: Another referral made by SafeHaven is Medicaid and CHIP. Millions of Americans are covered by these programs, including low-income adults, pregnant women, children, elderly, and disabled individuals. In the CA, both programs, eligibility requirements, and coverage are briefly described. Further, it directs them to the state agency's contact information if they have any Medicaid-related questions.
- Medication payment: Despite having health insurance, there might be a possibility that the survivor struggles to pay for medications prescribed by their physicians. We designed this option to assist survivors in exploring resources that may assist in paying for their prescriptions. As an example, SafeHaven offers several helpful options, such as the FamilyWise Prescriptions Savings Card, which provides discounts on all FDA-approved prescription medicines not covered by insurance, Medicaid, or Medicare. It continues by referring the survivor to RxAssist, an online resource center for medication assistance that offers information about pharmaceutical company programs and other resources that may help them reduce the cost of their medication.
- Mental health and substance use disorders: As mentioned earlier, those who have suffered DV are highly vulnerable to the danger of mental illness as a

result of their trauma. The survivor might suffer with insurance providers as they do not provide sufficient coverage for the treatment of mental health disorders or substance abuse disorders. Given these considerations, this option was designed to discuss this issue as well as identify what solutions are available. By way of example, It refers the survivor to the [mentalhealth.gov](https://www.mentalhealth.gov) website so that they may understand the protection that is available to them under the Mental Health and Substance Use Disorder Coverage Parity Laws and what may be covered and what might not be covered. In addition, it refers the survivor to the Open Path Psychotherapy Collective, which is a non-profit organization that offers low-cost mental health services in-office, where they can find a therapist in their area. It asks the survivor for their location to locate a therapist in their area. Once SafeHaven has confirmed the location, it refers them to the [openpathcollective.org](https://openpathcollective.org) website, where they can find a list of mental health therapists nearby.

Low cost healthcare 🏡

✅ In spite of having health insurance, many plans do not cover everything, and finding a provider that accepts your insurance may be difficult 😞.

✅ The Health Resources and Service Administration (HRSA) supports approximately 1,400 health centers which provide medical services on a sliding scale based on your ability to pay 💵.

✅ A wide range of health services are available at these health centers, including medical care, dental care, vision care, and behavioral health services.

I may be able to assist you in locating a health center in your area 📍.

I would like to confirm that your zip code is 19713. Is that correct?

Yes

No

Thanks 😊. Please click on the button below to find out healthcare centers available in your area 📍, and offers medical options on a sliding scale based on your financial ability.

Find health center 📍

I hope you find this information useful 😊. You might refer to other healthcare aid programs, or relevant resources based on your needs 🙌.

[Healthcare aid](#)
[Resources main page](#)
[Main menu](#)

Figure 4.23: Low-cost healthcare

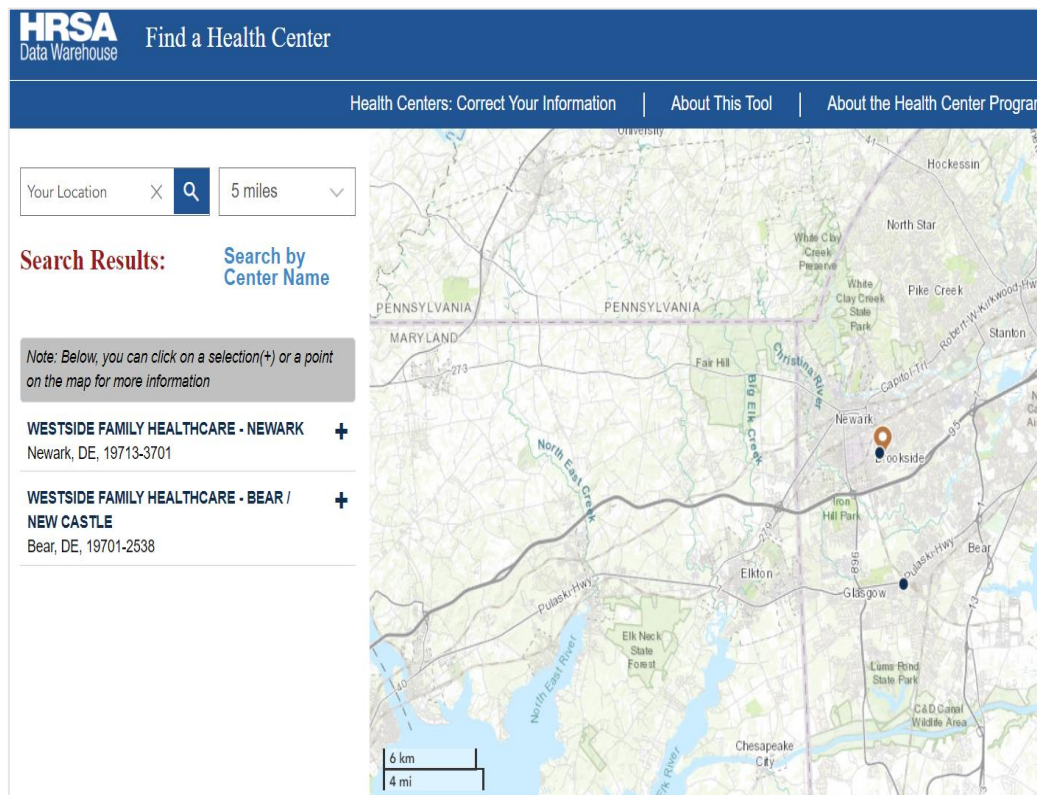


Figure 4.24: List of health centers (findahealthcenter.hrsa.gov)

#### 4.3.5.7 Informational and Appraisal Resources

A common recommendation from our expert interviewees was the potential of CAs to provide survivors with information and guidance. Through education about DV, their rights, and safety tips, SafeHaven can better equip survivors with the ability to recognize the dangers associated with DV, as well as take the necessary precautions to prevent abuse from occurring again. From this standpoint, we designed this sub-module to educate the survivor about DV and offer knowledge and advice regarding the survivor's relationship, safety, assessments, and other relevant topics. More specifically, within this sub-module, we have designed six options, including understanding DV, abuse signs, knowing about rights, technology safety, risk assessment, and priority setting.

- Understanding DV: We designed this option to help survivors understand the topic of DV because many survivors might be unaware of this issue, believing it to be a normal phenomenon in all houses and relationships and that it does not constitute a serious issue. This option provides the survivor with some informational resources, including the definition of DV, facts about DV, DV forms, the Power and Control wheel, and the cycle of violence.
  - What is domestic violence: In this option, the survivor will gain knowledge regarding DV. Specifically, SafeHaven provides information to the survivor regarding what DV is, what types it can take, and who would be most vulnerable to such violence. Additionally, the survivor will be informed of the dangers associated with this problem and that it may occur in any relationship and may expand to affect other family members (Figure 4.25).

- Forms of domestic violence: We created this option as a means of educating the survivor about the forms of DV and what actions and behaviors are used by the abuser to exert control. To clarify, SafeHaven will list all major forms of abuse (i.e., physical, emotional) and explain each form in detail with relevant examples.
- Facts about domestic violence: DV is widespread, impacting millions of people in the U.S. There is one detail that is critical to be aware of—one that is often overlooked—that anyone might be subject to abuse, regardless of their age, income, ethnicity, gender, or any other characteristic. We designed this option to educate the survivor about the statistics regarding DV and the extent to which it is prevalent. All of the statistics have been compiled from various government agencies or national domestic violence organizations such as the Department of Justice, the Centers for Disease Control and Prevention (CDC), the National Coalition Against Domestic Violence (NCADV), and the National Network to End Domestic Violence (NNEDV). Through textual and imagery content, SafeHaven explains several key facts about domestic violence across different types of violence and explains how they are common regardless of age, gender, or other factors.

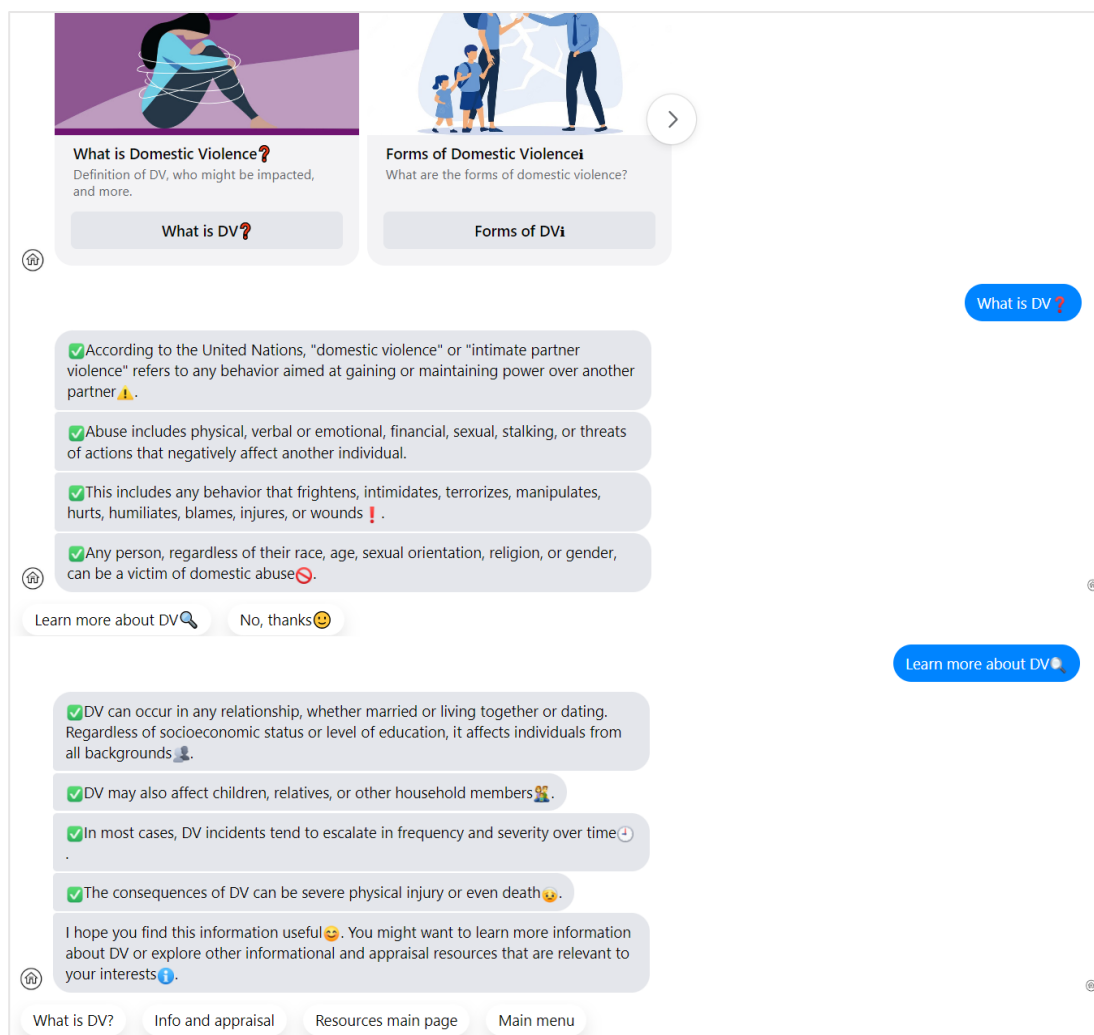


Figure 4.25: What is domestic violence?

- The Power and Control Wheel: Following the recommendations of many of our interviewees and considering the importance of providing informational support to DV survivors (DP5), this option aims to educate survivors about the Power and Control Wheel and what key points they should take away from it. First, SafeHaven explains to the survivor what the Power and Control Wheel is. In addition, it visually presents the wheel and lists all the tactics abusers use to gain control and power over their victims, either in the inner ring (i.e., using intimidation, minimizing, denying, and blaming, using male privilege) or the

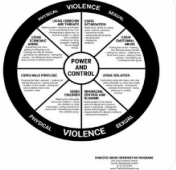


outer ring (i.e., physical and sexual violence) of the wheel. Furthermore, SafeHaven demonstrates how the wheel can be used in different settings for describing battering (i.e., law enforcement or prosecutor training) (Figure 4.26).

- The cycle of domestic violence: There has been a reference to the fact that DV is a crime with the highest repeat rate of all crimes [305]. This means that the phases of violence will be repeated several times when violence occurs. In this option, the survivor will learn about the cycle of abuse and its stages. Upon introducing the cycle of abuse theory, SafeHaven discusses why the abuse continues to occur and how it can be stopped. In addition, SafeHaven outlines the stages of the cycle (i.e., tension building, incident, reconciliation, calm) and defines each stage.

**Power and Control**

- ✓ The power and control wheel is a tool that is used in the DV field to gain a better understanding of how abusers gain power and control over their victims.
- ✓ A wheel is a visual representation of the tactics that abusers employ to keep their victims in relationships.
- ✓ This wheel was developed by Domestic Abuse Intervention Programs as part of "Duluth Model".
- ✓ It focuses on training and education, and aims to shift blame to abusers and support victims.
- ✓ Here is a diagram that illustrates these tactics.



- ✓ As you can see, the inner ring represents subtle, continuous behavior over time, while the outer ring represents physical and sexual violence.
- ✓ The wheel is composed of the following components:
- ✓ Outer ring:
  - Physical and Sexual Violence.
- ✓ Inner ring:
  - Using Intimidation
  - Using Emotional Abuse
  - Using Isolation
  - Minimizing, Denying, and Blaming
  - Using Children
  - Using Male Privilege
  - Using Economic Abuse
  - Using Coercion and Threats
- ✓ In many settings, the wheel can be found in guides, books, scholarly publications, and on the walls of organizations that are working to prevent DV.

Figure 4.26: The Power and Control Wheel

- **Abuse signs:** Although abusers do not have a typical, identifiable personality, they usually exhibit certain characteristics and behave in a particular manner that may indicate a warning sign of possible mistreatment. In the design of this option, we seek to assist survivors in understanding DV signs. As part of this option, SafeHaven explains these signs of the major types of abuse, such as emotional, physical, financial, sexual, stalking, control, and coercion abuses. In addition, it offers them various suggestions to mitigate its impact and prevent it from recurring in the future.
- **Know my rights:** There is an important aspect that survivors might be in high need of, which is knowledge of their rights to protect themselves and their dependents. Our objective in designing this option was to inform survivors about several of their rights.

Specifically, the survivor will be able to learn about: gathering evidence, protection order, divorce, reporting abuse, and child safety.

- Gathering evidence: Because the majority of DV incidents take place in the private sphere, it is not surprising that only 16 percent of all cases investigated by police patrols and 19 percent of cases investigated by special units were supported by witnesses [306]. Consequently, the only evidence that a survivor may have and that could be helpful to them is most likely their word, which could be denied by the abuser. Evidence can be the determining factor in whether DV will recur or cease or whether a protection order should be granted or not. We developed this option to explain the importance of gathering evidence when pursuing legal action or obtaining assistance. SafeHaven first explains the importance of having evidence, especially in court hearings or negotiations in separation agreements. In addition, SafeHaven instructs survivors when evidence can be gathered. For example, it teaches a user to consider evidence when they are financially marginalized, has documented that they provide the majority of childcare, has medical reports of stress-induced injuries and illnesses, or have eyewitness testimonials. Additionally, SafeHaven instructs survivors on where they can find evidence. For instance, it teaches them that evidence can be found in medical reports, text messages, emails, letters, photos, videotapes, call logs, or the survivor's or abuser's social media accounts. Lastly, SafeHaven advises survivors to be cautious when gathering evidence and ensure the abuser is unaware of what they are doing (Figure 4.27).

- Protection Order: It is a court order that directs the abuser to stop stalking and harassing the survivor and stipulates a set of restrictions on the abuser's behavior. The acquisition of a protection order is a crucial step in protecting the survivor and their dependents from further mistreatment. The purpose of this option is to educate the survivor to make them more aware of the protection order. As part of the initial step, SafeHaven explains what a protection order is and who can apply for it. Following this, it describes how a protection order protects the survivor in several ways, such as requesting the abuser to refrain from contacting the survivor and remain away from any location specified by the survivor, as well as allowing the survivor to have temporary custody of the child. SafeHaven outlines the steps for survivors to follow to obtain a protection order. Finally, SafeHaven refers the survivor to the WomensLaw.org website to learn more detailed information on obtaining a protection order in their state.

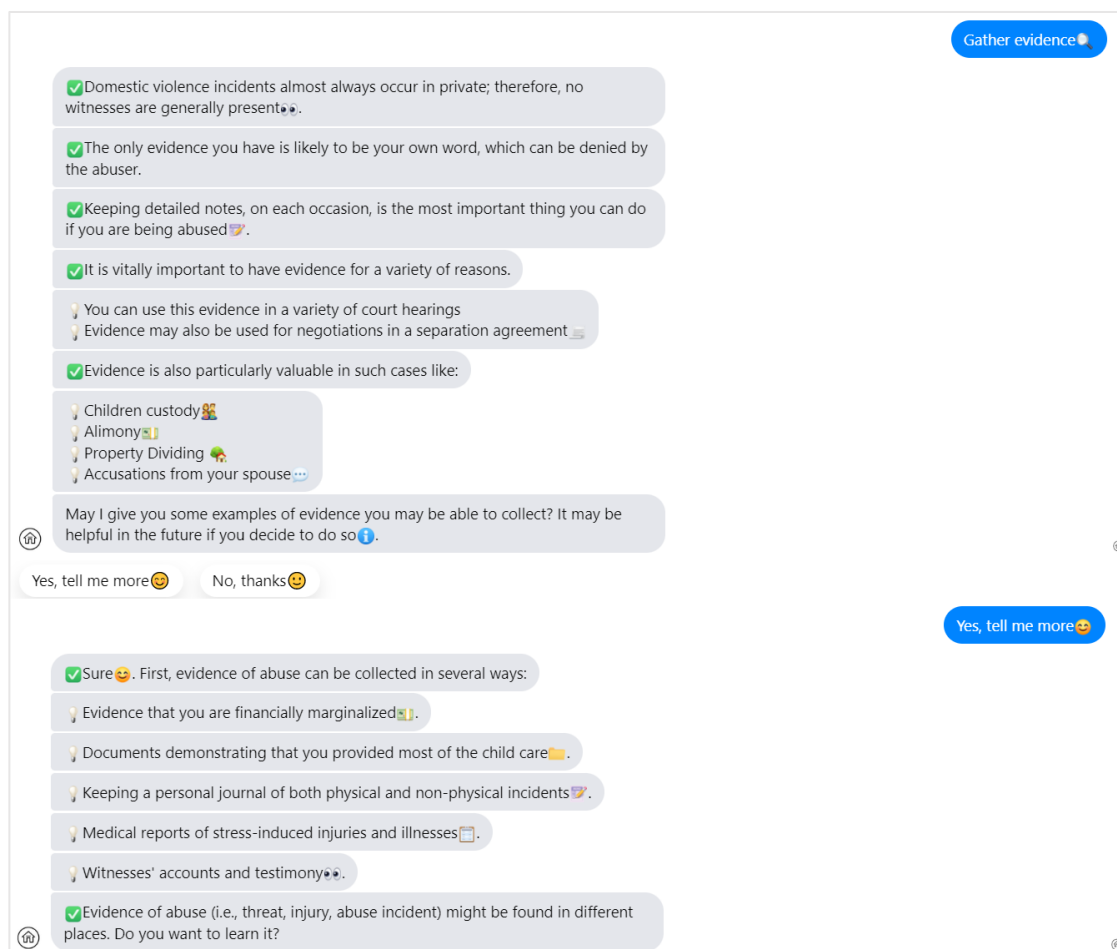


Figure 4.27: Gathering evidence

- **Report abuse:** In this option, the survivor will learn how to report a DV incident. As a first step, SafeHaven encourages survivors to report their cases to the police if they are in a safe place and provide information about the situation, the perpetrator, and the violent act. However, in the event that they encounter obstacles in contacting the police, SafeHaven recommends that they consider contacting a victim services unit or local DV service or visiting a hospital or police station to report the incident. Besides, SafeHaven suggests a few tips for reporting, such as bringing a family member while filing a report, reaching out to a domestic violence advocate, and learning about their legal

rights. Lastly, when a child is involved in a case of child abuse, SafeHaven offers guidance regarding how to report the abuse, which includes calling the police, contacting the local child welfare office, or reaching out to the Childhelp National Child Abuse Hotline ([childhelpline.org](http://childhelpline.org)) (Figure 4.28).

- Divorce: There is evidence that domestic violence is among the primary "last straw" reasons that lead to divorce [307]. This option aims to provide the survivor with a better understanding of the divorce process. SafeHaven first outlines the basic steps involved in filing for divorce in most cases (i.e., meeting the residency requirements, having a legal reason, and filing a divorce), even though each state has its divorce laws. Afterward, SafeHaven discusses many considerations and topics regarding child custody and support, like the final custody order or the temporary child support order. In closing, it refers them to the WomensLaw.org website, where they can learn more details on the topics of divorce, custody, and child support laws in their state.
- Children safety: Despite the widespread belief that DV occurs only in adults, the truth is that children are also exposed to abuse or witness abuse in some way, and this might increase their risk of developing physical, social, emotional, and behavioral problems, as well as mental disorders, substance abuse, and academic difficulties [308]. In this option, we aim to educate survivors about enhancing their children's safety as well as reducing their stress during a DV incident. As an initial concern, SafeHaven discusses the detrimental effects that a child might experience as a result of being exposed to or witnessing a DV event. After that, SafeHaven provides suggestions to

survivors for protecting themselves and their children. The suggestions include establishing a code phrase with children, seeking temporary shelter (i.e., a shelter), and preparing a safety plan. In the same vein, SafeHaven advises the survivor to consider some practices for the sake of the safety and well-being of their children, such as ensuring the children understand the abuse is not their fault, listening to them, providing information and support to assist with their feelings, and helping them develop positive coping skills.

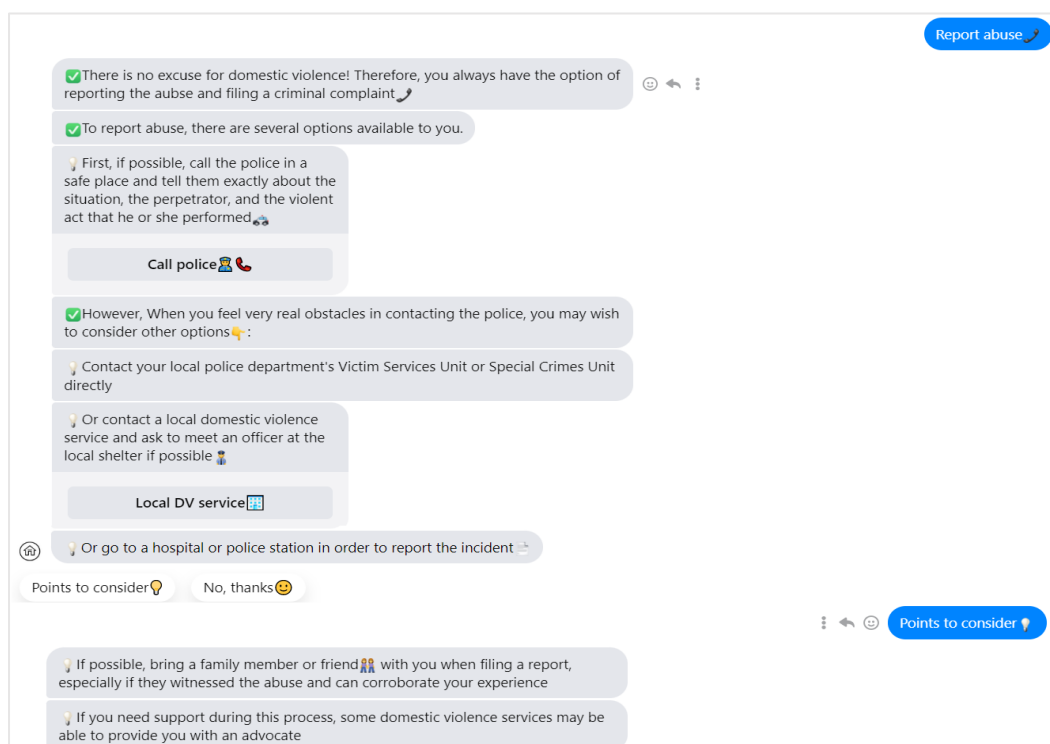


Figure 4.28: Reporting abuse

- **Technology safety:** The use of technology can assist survivors and their children in fleeing from violent abusers, but it is important to keep in mind that technology can also compromise a survivor's privacy and safety. DV survivors are often harassed or stalked by abusers who are exploiting technology as a way to control survivors and manipulate their actions. Technological advancements occur rapidly, making it

challenging to remain informed of the implications, advantages, and risks. For example, spyware programs, which were designed originally to assist parents in keeping their children safe when using the computer, are often used by abusers to monitor their survivors' computer activities. Further, GPS tracking devices can provide abusers with information regarding the location of their victims. Furthermore, SMS messages, which serve as a means of communication, may be misused by abusers to harass or threaten their survivors. Keeping what we discussed above in mind, we designed this option to inform survivors on how to increase their safety while using technology to mitigate their risk of abuse from abusers. Besides providing information about the various forms of technology abuse and how to choose online assistance safely, this option provides tips for technology safety plans, abuse documentation, smartphone safety, and online safety. All contents in this option are sourced from the National Network to End Domestic Violence (NNEDV) and Womenslaw.org websites.

- Technology safety plan tips: This option is designed to raise awareness among survivors about tips and precautions to consider when developing a technology safety plan. This includes prioritizing safety, recognizing abuse, enhancing security, and maintaining privacy. As part of the prioritize safety component, SafeHaven offers survivors tips about how to prioritize their technology safety. For example, it advises them to consider using a safer device if they suspect that their device is being monitored. Further, it suggests they visit the National Domestic Violence Hotline or the National Network to End Domestic Violence



(i.e., the Safety Net project) to learn more about solutions, local resources, or technological advice.

Under the recognizing abuse component, SafeHaven provides survivors with tips about how to identify abuse. By way of example, it recommends that they observe patterns and consider the kinds of technologies that may be used to stalk or harass them. Moreover, it recommends that they document the incident, as well as report it to the police.

As far as enhancing security is concerned, SafeHaven suggests some steps that can be taken to enhance the survivor's security and reduce their risk when using technology. To illustrate, SafeHaven suggests that survivors update their passwords and usernames if they suspect their online accounts have been compromised. Furthermore, it recommends that they obtain a new device with an account that the abusive person does not have access to, especially if it appears that their device is being monitored. In addition, SafeHaven advises survivors to ensure their location is protected, check whether location sharing is enabled or disabled on their devices, and adjust accordingly.

In the same context, we created a component that discusses how to maintain the survivor's privacy online or offline and how to keep their data as secure as possible. In this regard, SafeHaven offers several recommendations. For instance, it suggests that the survivor ensures that their address is protected and kept confidential. Further, survivors are advised to avoid providing more information than necessary and opt out of information collection. Additionally,

it suggests that survivors take control of their online and offline privacy (i.e., remove history, cookies, and temporary internet files periodically).

- Abuse documentation: We created a separate option to explain document abuse because it is a vital component of legal proceedings or seeking authorities' assistance. Although survivors may learn how to gather evidence in the option of "gathering evidence" under the "know my rights" sub-module, this option focuses on documenting technology-related abuse and what precautions they need to follow. First, SafeHaven emphasizes the importance of abuse documentation for several reasons, including pursuing legal action, warning them about monitoring, and assisting in identifying patterns of abuse. Furthermore, it provides several tips on how to document tech-related abuses. For example, SafeHaven recommends that the survivor record the incident and gather pertinent information (i.e., date, time, location, the technology used, etc.). In addition, it suggests that they should understand what technology the abuser might use (i.e., computer, hidden camera, GPS tracker, etc.). As a final step, SafeHaven describes what should be documented and offers tips in this regard, and this includes email, text messages, social media harassment, phone harassment, and impersonation of telephone numbers. For instance, when the abuser sends an email containing harassment, SafeHaven instructs the survivor to print or take screenshots of the email content, as well as save the email headers that contain the email address.
- Choosing online help safely: It is often challenging to obtain assistance, find alternatives, and escape from the abuser, especially when there are fears of

retribution from the abuser. With in-person services, there are many options. There is a possibility, however, that not all these services might be accessible to the survivor at the precise moment they are needed. Several online options could help in such a situation, such as joining online support groups, reaching out to hotlines via phone, text message, or chat, or contacting a family member or friend via phone, or social media. To maintain the survivor's privacy and safety, this option aims to educate them about safety precautions when seeking online assistance. SafeHaven offers several suggestions in this regard. For example, survivors are encouraged to utilize technology to enlist help or connect with others and consider a safe option when doing so. Moreover, SafeHaven recommends that survivors clear the threads of text messages on their devices so that others will not be able to view them. In the same context, SafeHaven advises the survivor to delete their browsing history and cookies as a means of covering their tracks so that chat messages or online support group sites are not stored on their devices.

- Forms of technology abuse: The term technology-facilitated abuse describes a form of control by using technology to harass, intimidate, or stalk another individual. Abuse of this nature can be perpetrated using a wide range of technological means, such as smart devices, online platforms, and mobile applications, and information about how to use these technologies is readily available online. To this end, we created this option to educate survivors about the various forms of technological abuse. With examples, SafeHaven discusses a variety of forms of technology abuse, such as cyber-surveillance,

cyberstalking, online harassment, recording, spoofing, impersonation, GPS tracking, electronic surveillance, and abuse involving texts, photos, and video (Figure 4.29).

- Phone safety tips: Smartphones today offer a variety of useful features, such as the ability to store information, share locations, browse the web, and access personal accounts. An abuser, on the other hand, may view it from a different angle, in which they misuse smartphones to exert power and control over their victims. Survivors of abuse must be equipped with the necessary knowledge and skills to respond to tech abuse. Throughout this option, SafeHaven provides some tips to increase survivors' privacy and safety when using smartphones. For instance, it teaches them to secure their phone with a passcode (i.e., password, pattern). In addition, SafeHaven recommends the survivor keep an eye on downloaded apps from time to time and delete any unfamiliar apps. Furthermore, SafeHaven urges the survivor to download anti-spyware and anti-virus software (Figure 4.30).
- Online safety tips: There is no doubt that the Internet serves as an invaluable tool for the sharing and learning of information. Those who have experienced DV can benefit greatly from it. It can be an indispensable resource for gaining access to support and information, as well as connecting with family, friends, advocates, and professionals. In contrast, abusive partners can use the Internet to initiate, continue, or escalate abusive behaviors, thus making your online safety all the more essential. There is an increasing threat to DV survivors from cyberstalking or using the Internet to harass or spy on them. In designing

this option, we aim to increase awareness of the survivor while they are online. SafeHaven offers several tips in this regard. For example, it recommends that survivors avoid providing personal information beyond what is required when setting up accounts or profiles. In addition, it advises that they clear history, cookies, and saved passwords from their web browser, or use incognito mode when surfing the Internet. Moreover, SafeHaven urges the survivor to make sure their devices are regularly scanned with antivirus software that is kept up to date.

Forms of tech abuse !

✓ In the context of domestic violence, technology abuse is a common tactic used by abusers. 📱

✓ Technologies such as social media and texting are used by harmful partners to gain control and power 😬

✓ Technology abuse takes many forms, including:

- 🔍 **Cyber-surveillance**: Internet-connected devices are used for monitoring. An abuser may use cyber-surveillance technology to photograph, or videotape you; log your activities, and use the connected devices to access your email.
- 📱 **Cyberstalking and online harassment**: Abusers misuse social media, emails, or apps to harass and stalk victims (i.e., send threats, doxing, cyberbullying). Abusers tend to commit online harassment to make victims uncomfortable, ashamed, or scared.
- 📺 **Recording**: The abuser may record your conversations or videotape you to gain more information about you.
- 📞 **Spoofing**: It may be used by an abuser to pretend to be someone else to ensure you pick up the phone.
- 👤 **Impersonation**: For example, establishing a fake social media account or email account under your name and spreading false or harmful information.

Learn more forms 📖 No, thanks 😊

Learn more forms !

Absolutely 😊

📍 **GPS monitoring**: The abuser might misuse GPS technology to control you. It is possible for an abuser to use GPS to track your movements and misuse that information.

📺 **Electronic surveillance**: It involves misusing cameras, recorders, wiretaps, social media, or email. Monitoring software (also known as spyware) can also be misused to track your activities secretly on a computer, tablet, or smartphone.

📱 **Abuse Involving Texts, Photos, and Videos**: The abuser may utilize images or videos shared through sexting to maintain power and control (i.e., blackmail).

✓ You may also find the figure below helpful for more details 📖

To learn more about abuse using technology, please refer to the buttons below 📖

Figure 4.29: Forms of technology abuse

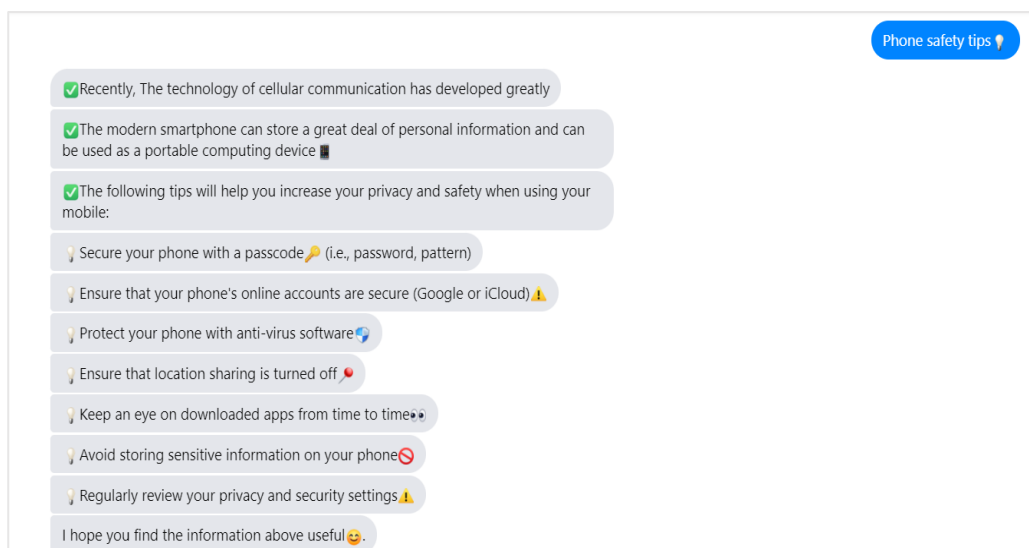


Figure 4.30: Phone safety tips

- **Risk assessment:** As part of the informational and appraisal sub-module, the 'Risk assessment' option enables the survivor to complete the Danger Assessment (DA), a 20-item validated and comprehensive risk assessment comprised of severe or fatal violence risk factors [241], [309]. There are several items in the assessment that pertain to the nature or severity of the abuse experienced by the survivor in her or his relationship with the abuser, including threats of violence, forced sexual contact, or strangulation; several factors are reflective of the survivor, for example, whether she has a child or has a previous attempt or threat of suicide; some factors reflect characteristics of the abuser, including issues related to employment, substance abuse, or suicidal thoughts.

There is a 2-point scale used to assess the presence of the 20 risk factors (i.e., 0 = no, 1 = yes). Among the items rated, some are based on lifetime presence, while others are based on their existence in the previous year (Figure 4.31). The ratings of the items are summed using a simple unit-weighting approach to determine the total score between 0 and 20. The total scores can be categorized into four levels of danger that reflect the

likelihood of life-threatening violence. These levels are categorized as follows: 0–7 = variable danger, 8–13 = increased danger, 14–17 = severe danger, and 18 or more = extreme danger. The survivor will be informed of their score as well as their level of danger after completing the assessment. Furthermore, they will receive suggestions regarding intervention measures and guidelines that may help mitigate their risk. As an example, if the survivor's danger assessment score is 10, this means that her/his danger level is “increased danger”. SafeHaven first informs the survivor of his/her danger level, and then provides several recommendations related to the survivor's level of danger, including continuing to monitor and plan for safety, remaining vigilant for any further indications of danger, consulting judges, and seeking advice from high-level supervision (Figure 4.32).

✓ The purpose of the risk assessment is to determine the level of danger that an individual is in when he or she is abused by a partner 🗨️  
 ✓ This assessment consists of 20 items and is scored based on yes/no responses to risk factors associated with intimate partner homicide ⚠️  
 ✓ Among the risk factors are prior death threats, the employment status of the partner, and access to guns by the partner 🚫

🏠 would you like to proceed?

Yes

Thank you 😊. It is crucial to understand the danger of abuse and to become aware of the risk factors that are relevant to you !

Now, Let's start with questions ➡️

1) Has the physical violence increased in severity or frequency over the past year?

Yes

No

🏠 2) Does your abuser own a gun?

Yes

No

Yes

No

Figure 4.31: Risk assessment- Questions

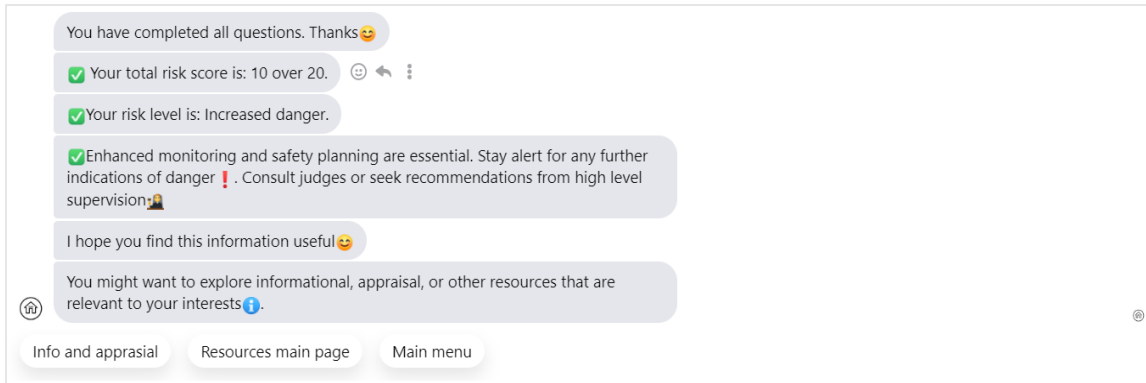


Figure 4.32: Risk assessment- Risk score and level

- **Priority setting:** Among the features of SafeHaven is an option that allows survivors to identify their top priority for safety when it comes to abusive relationships. Through this option, survivors may weigh the importance of factors to guide their decision-making [109]. For the survivor of an unsafe relationship, there are several factors to consider: my safety, my child's well-being, having resources, and my feelings toward my partner. SafeHaven presents these four factors to the survivor two at a time, in which the survivor prioritizes each factor according to its perceived importance based on pairwise comparisons, such as comparing the importance of "my safety" to "having resources." In the comparison, they will be shown five selective buttons with different percentages (i.e., my safety: 90% vs. having resources: 10%, my safety: 75% vs. having resources: 25%, my safety: 50% vs. having resources: 50%, my safety: 25% vs. having resources: 75%, my safety: 10% vs. having resources: 90%), and they need to choose one that they consider more significant (Figure 4.33). After conducting the pairwise comparisons, mathematical calculations will be used to determine preference weights [310]. Finally, an outline of the survivor's priorities will be presented to him or her. In addition, SafeHaven would provide the survivor with a few referral resources and guidance to facilitate the fulfillment of their needs following their top



priorities. For instance, in the event that the survivor's highest priority is their safety, SafeHaven will refer them to safety-related resources and tips such as developing a safety plan, general safety tips, technology safety, or reaching out to the National Domestic Violence Hotline (Figure 4.34).

✓ The purpose of this part is to assist you in determining what factors are most important to you, and as a result, you can make the right choice in the future 🙌

✓ There will be a comparison of a number of factors, and you choose the factor (in %) that you believe is more important 🙌

Would you like to start?

Yes

No 😊, go to resources

Yes

Sounds good 😊. We want to compare now which of the below factors has higher priority to you 1

- Your concern about safety ⚠️
- Your children's well-being 🧒🧒
- Having resources 📁
- Feeling for your partner 👫

As part of each comparison, you have to choose one option based on your priority for the two factors 🗳️

Let's begin now 🙌

✓ How important are these two factors to you (your concern about safety ⚠️ vs. having resources 📁)?

90% ⚠️ vs. 10% 📁 75% ⚠️ vs. 25% 📁 50% ⚠️ vs. 50% 📁 25% ⚠️ vs. 75% 📁 10% ⚠️ vs. 90% 📁

90% ⚠️ vs. 10% 📁

✓ How important are these two factors to you (your concern about safety ⚠️ vs. your children's well-being 🧒)?

90% ⚠️ vs. 10% 🧒 75% ⚠️ vs. 25% 🧒 50% ⚠️ vs. 50% 🧒 25% ⚠️ vs. 75% 🧒 10% ⚠️ vs. 90% 🧒

50% ⚠️ vs. 50% 🧒

Figure 4.33: Priority setting

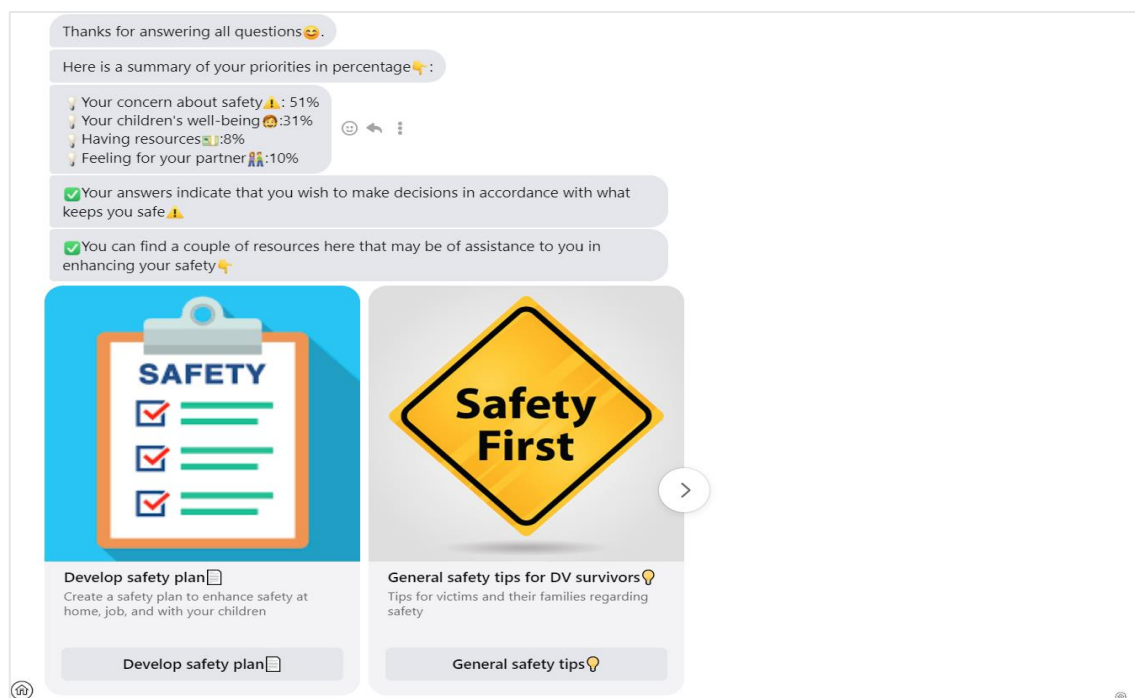


Figure 4.34: Priority setting - summary of priorities

#### 4.4 Implementation

As part of our CA development, we used Google Dialogflow and ManyChat. Using a hybrid approach, we have implemented SafeHaven in a manner that combines two modules: a rule-based approach and a retrieval-based approach. A rule-based approach, which aligns finite state-based system, guides user through a predetermined decision tree. This approach allows users to interact with the conversation by following the scripted conversation flow and selecting appropriate responses. As a result of the design section, we created five major modules and 68 submodules/sub submodules. An integrated menu of main modules and submodules (e.g., referral resources, abuse screening, safety plan development, informational and appraisal resources, survivors' stories) was developed to provide survivors with well-organized information.

As part of the retrieval-based approach, we integrated SafeHaven with Google Dialogflow. It is designed to integrate frame-based and agent-based systems simultaneously, with Google

Dialogflow handling intents and entities, filling out necessary information during interactions (frame-based), while machine learning and NLP generate responses based on context and intent (agent-based). To make the integration process work, programming had to be done to handle API interactions between ManyChat and Dialogflow in order to facilitate data exchange immediately. As users engaged with SafeHaven, their inputs were forwarded as API requests to Dialogflow. There, the data is analyzed, intents and entities are matched with a pre-trained model, and appropriate responses are generated. By implementing this configuration, users were able to smoothly move between different modules, while database searches were used to retrieve specific information in response to their actions (Figure 4.35).

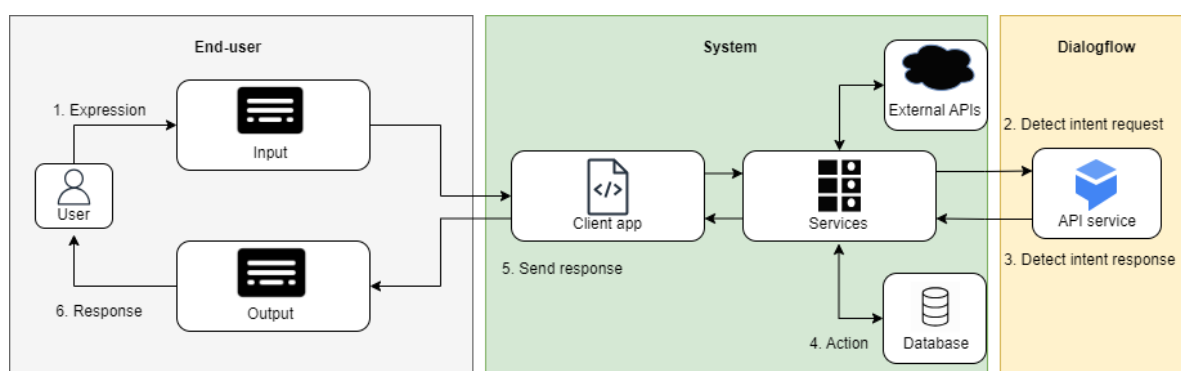


Figure 4.35: The process flow in interacting with the API

Our goal was to boost the user's perception that they were conversing naturally with the CA. Our agent is trained to build a machine learning model in Dialogflow using training data. Training data is composed mainly of intents, intent training phrases, and entities, which can be used to label data for machine learning. Since Dialogflow utilizes Natural Language Processing (NLP), even if the user does not use precise keywords to communicate with the CA, Dialogflow is still capable of recognizing their intent. In our context, we train our agent with related entities and phrases and then connect them to the corresponding 5 modules or 68 submodules/sub submodules. For instance, when the user asks the CA, 'I would like to obtain a protection order.'

or 'What is the process for obtaining a restraining order?', the CA would interpret this inquiry and then classify it as an intent of filing a protection order, as it compares the inquiry to the training phrases with an entity in a training phrase, to find the best match. Then, the CA would connect the user to the submodule "Protection Order" so they can follow the submodule conversation flow to obtain further details.

Besides providing further details, the CA provides the option for users to switch between conversation topics. For instance, suppose that the user is in the "Protection Order" submodule. As he/she receives information regarding the protection order and how to obtain it, he or she becomes interested in financial assistance. In such a scenario, the user may ask the CA, "I am looking for financial assistance." In this instance, Dialogflow categorizes the inquiry as "financial assistance" intent and switches the user from the "Protection Order" submodule to the "Financial Assistance" submodule. If the user intends to switch between topics, it is recommended that they allow the CA to complete the conversation flow, or respond when asked to, to avoid the possibility of mingling responses, due to the constraints of ManyChat.

#### 4.5 Hypotheses Development

Nowadays, a growing number of people are looking to the Web to learn about healthcare options and to find support and resources related to various health and human services issues. In light of the fact that teens' concerns about confidentiality seem to be a significant barrier to obtaining information, and help directly from others, teens may benefit from the anonymity of the Internet, which would allow them to discuss sensitive topics online that may not be shared with parents, healthcare providers, school officials, or friends [311]. For example, search engines (such as Google) are frequently used to find information relating to sexual health [312]. In the context of DV, online search is often the first point of contact for survivors seeking information

and guidance [38], [313], [314]. The use of the Internet has been documented as a way to promote empowerment and benefit survivors of DV through informational support [106], [114]. However, there are some potential concerns that may arise when survivors seek social assistance services on the Internet [114]. For example, one disadvantage of seeking assistance online is the possibility of accessing inaccurate websites and unreliable information. Since anyone can post information online, it can be difficult to determine which sources are credible and which ones contain false or outdated information. Further, there is no systematic quality control of Internet information. In addition, there is a concern regarding a loss of privacy, as there is the possibility that information might be archived and searched using internet search engines, or that the abuser might search for web history, which may pose some risks for the survivor.

Motivated by prior work, we anticipate that the CA will address such limitations. It has been demonstrated that CAs provide accurate information, guidance, and tips in a timely manner for survivors of DV [29], [30], [38], [41]. By leveraging the power of AI and machine learning, CAs can better detect and analyze self-disclosed information by users, which can provide valuable insights into how DV survivors can be best supported, assist with identifying risks and provide advice [30]. In addition, the CA proves its capability of providing emotional support, while the Internet may lack this [29], [38]. Through its interaction, the CA demonstrates its ability to offer support without imposing any judgments and securely stores all conversations. AI and machine learning enable CAs to understand the user's language and emotions, which allows them to tailor their support to the individual's needs. This is much more personal and effective than the generic advice often found on the Internet. Further, CAs may provide instrumental support by facilitating access to and referrals to a variety of resources [29], [30]. This can be especially beneficial for those with limited knowledge of the resources available to them. In this regard, we argue that:

H1: CA-based DV intervention is perceived to be more effective in providing emotional support compared to the search engine counterpart.

H2: CA-based DV intervention is perceived to be more effective in providing informational and instrumental support compared to the search engine counterpart.

On November 30, 2022, OpenAI launched ChatGPT, a generative language modeling tool that enables the user to converse with a machine about a variety of topics [315]. Having become a valuable resource across various domains, it has gone beyond its basic role as a chatbot and has emerged as an advanced language model. ChatGPT offers a wide range of services to meet the needs of users, ranging from answering complex inquiries to generating creative content, solving problems, and facilitating interactive conversations. With its ability to comprehend natural language and adapt to a variety of contexts, it makes an ideal companion for anyone seeking guidance, information, or support.

The CA-based intervention of DV designed to offer support is created with a specific target group: DV survivors. Its primary objective is to provide various support services (emotional, informational, instrumental) tailored specifically to this group's needs. ChatGPT, a general language model developed by OpenAI, is capable of understanding and generating text similar to that of a human across a wide variety of topics without the focus on a specific group or subject area. Furthermore, it is likely that CA-based interventions for DV are equipped with a knowledge base and resources related to DV, survivors' services, and support programs. This can include information concerning resources, hotlines, counseling programs, and various other supports. Furthermore, emotional support techniques may have been incorporated into the CA to meet the psychological needs of survivors. The ChatGPT system, despite being a powerful language model, may lack such targeted knowledge and resources, relying instead on patterns learned

from a substantial amount of internet data to generate its responses. The CA-based intervention for DV is developed with a thorough understanding of its sensitive nature, ensuring it does not trigger or retraumatize survivors. It may utilize techniques such as listening actively, validating the feelings of others, and encouraging them. Conversely, since ChatGPT is a general model, it might not be able to recognize and respond to the sensitive needs of survivors to the same extent as a specialized model. Thus, our hypothesis posits that the CA-based intervention of DV will be perceived as more effective in providing informational, emotional, and instrumental support when compared to ChatGPT.

H3: CA-based DV intervention is perceived to be more effective in providing emotional support compared to ChatGPT counterpart.

H4: CA-based DV intervention is perceived to be more effective in providing informational and instrumental support compared to ChatGPT counterpart.

Despite their capability to scour vast quantities of data, traditional search engines often return a deluge of information that may prove overwhelming to users, particularly those in vulnerable states seeking sensitive information (e.g., DV). In contrast, CAs tailored to these groups are developed with a deep understanding of the context and their needs. Through the interactive capabilities of an AI CA, information can be searched and engaging experiences can be driven [316]. A CA could provide a direct and concise response without requiring data seekers to wade through a lengthy search results list [317]. In Brandtzaeg and Flestad's study [318], it was found that users engage with CAs primarily due to their abilities to deliver quick, easy-access, timely, and consistent responses, highlighting that a chatbot's effectiveness is deeply intertwined with the better information quality it provides to users. This is particularly evident in specialized CA applications where the bots are customized to meet the needs of a user. As an example, It was

demonstrated by Maeng and Lee [38] that a CA developed specifically for survivors of image-based sexual abuse provided information that was more accessible, concise, and better organized than what could be found via conventional internet searches. Hence, we assume that,

H5: CA-based DV intervention is perceived to provide better information quality compared to the search engine counterpart.

Despite the extensive information available on ChatGPT, consistency, accuracy, and information quality should be considered when using it for sensitive topics (e.g., DV, mental health). Interactive and context-sensitive CA-based systems may enable a more nuanced and supportive interaction, which is particularly advantageous for DV scenarios where high-quality information is required. For example, based on recent research, it was evident that ChatGPT has not discussed any aspects of dealing with the issue of sexual violence [319]. In the same context, a study by Walker et al. [320] examined the reliability of medical information provided by ChatGPT in comparison with clinical guidelines and concluded that while ChatGPT is capable of providing medical information comparable to that available online, its quality is limited, highlighting the need for caution when utilizing it for healthcare purposes. Similarly, as part of a study examining ChatGPT's utility in mental health and substance abuse education, it was discovered that while the model generated information that appeared helpful on the surface, closer inspection revealed a number of shortcomings, including the outputs often lacked nuanced information and quality evidence, contained biases, and consequently a thorough human review and oversight was still required for all outputs [321]. Thus, a CA-based intervention designed specifically for sensitive topics (e.g., DV) could provide superior information quality and more beneficial outcomes than ChatGPT currently provides. We hypothesize,



H6: CA-based DV intervention is perceived to provide better information quality compared to ChatGPT counterpart.

Among the key design considerations of social technologies is anonymity, which promotes support provision and seeking in socially stigmatized situations. Despite some degree of anonymity on the Internet that facilitates information browsing and searching, as well as social networking, there are a number of adverse outcomes and consequences like cyberbullying [195], lack of accountability [27], and inaccurate and misleading information [114]. A CA allows users to interact with it without revealing their identity or personal information. Using this method may be advantageous in circumstances in which users are concerned about their privacy or do not wish to reveal their identity. While search engines may anonymize the data they collect to some degree, they nonetheless collect and store user information, such as IP addresses and search history, which can be used to track individual users and their online activities. Humans may be more inclined towards CAs, as they offer a completely anonymous environment and will not be recognized by others. As an example, adolescents have been shown to perceive CAs as being more anonymous than search engines when discussing matters associated with sex, drugs, and alcohol. By being anonymous, the survivor will have a greater sense of safety, as they will not have to fear judgment or criticism from others when seeking online support [137]. Hence, we argue that DV survivors will perceive higher anonymity in interaction with a CA than with a search engine. To this end, we hypothesize that:

H7: CA-based DV intervention is perceived to be more anonymous compared to the search engine counterpart.

In CAs that address sensitive matters (i.e., mental health [187], sexual health [215]), the perceived absence of traceability and the ability to exchange anonymous information has

emerged as a significant factor in enhancing user engagement, particularly among individuals who prefer to remain anonymous. Despite the fact that ChatGPT may be a valuable tool, it lacks anonymity. It is a cloud-based service powered by OpenAI. Despite OpenAI's efforts to protect the privacy of users, the use of a centralized cloud service involves some level of data transfer and storage that may compromise the anonymity of users. In addition, the use of an email address when logging into ChatGPT creates a direct connection between the user's identity and their interactions with the tool. In certain instances, this connection could undermine the anonymity that users may otherwise expect when they interact with AI tools. In addition, if users provide personal information to ChatGPT, there is a risk of data linking, potentially allowing AI responses to be associated with individual users, thus undermining their anonymity. On the other hand, CA-based interventions of DV might offer greater anonymity by virtue of their design and features. It is possible to engineer CAs with privacy measures so that users can interact anonymously without disclosing their personal information [215]. For example, Users may not be required to provide personal identification information or provide their e-mail address to use CA-based interventions of DV. A CA-based intervention of DV can also be carried out within a closed system, allowing only authorized personnel to access user data. In this controlled environment, users' data is protected from unauthorized access and is kept confidential and anonymous. Hence, we argue that DV survivors will perceive higher anonymity in interaction with a CA based intervention of DV than with ChatGPT.

H8: CA-based DV intervention is perceived to be more anonymous compared to ChatGPT counterpart.

The medium through which individuals seek information and support on sensitive topics plays a role in determining the effectiveness and trustworthiness of interventions. As a result of recent

advances in AI, CA-based interventions have become increasingly popular and perceived as a trustworthy medium. This perception stems from the personalized, interactive nature of CAs, which simulate a conversational partner, thus providing a semblance of trust and understanding that is often lacking in the impersonal results offered by search engines. This distinction is crucial when addressing delicate matters where trust and confidentiality are paramount. Evidence from literature indicates several factors that contribute to CA trustworthiness. First, it is suggested that CAs, due to their conversational interfaces, foster deeper self-disclosure, an essential component of effective support for those affected by DV. For instance, Lee, Yamashita, and Huang [322] found that participants were willing to share more personal and sensitive information with CAs, indicating a high level of trust in these platforms. Despite being informed that their information would be shared with mental health professionals, participants maintained a high level of disclosure, highlighting the perceived safety and confidentiality of CA interactions. Further, the anonymity and privacy provided by CAs is extremely valuable when dealing with sensitive matters, as users may fear judgment or repercussions if they seek help elsewhere [323]. In the same context, it was evident that AI CA "Jeniffer" was perceived as more trustworthy than traditional search engines, particularly in the context of seeking reliable information during public health crises like the COVID-19 pandemic, and this was due to the CA's foundation on an expert-sourcing framework. Through the direct involvement of domain experts in curating and verifying content, it is ensured that the information provided is not only accurate but also relevant and up-to-date. A difference may exist between this approach and algorithmic retrieval of information by search engines, in which data may originate from various sources of varying quality. Consequently, we hypothesize,

H9: CA-based intervention of DV is perceived to be more trustworthy than the search engine counterpart.

A key benefit of a specialized CA-based interventions over ChatGPT, especially in sensitive topics, is their ability to foster a more customized and engaging interaction, which can have an important influence on the trustworthiness of the tool. In spite of ChatGPT's broad conversational capabilities, it lacks the customized interaction that is crucial to establishing trust when discussing sensitive matters. As an example, Woebot has demonstrated strong user involvement and a noticeable reduction in mental distress through tailored interactions, demonstrating the importance of personalized interactions in strengthening the bond between users and interventions [36]. In addition, CAs with proactive interaction capabilities are able to initiate conversations and offer context-based recommendations, achieving levels of engagement and understanding that ChatGPT may not be able to achieve due to its reactive nature [324]. Especially in domains requiring nuanced understanding and confidentiality, the specialized and context-sensitive design of CAs, enriched with expert knowledge and reliable data, positions them as a more trustworthy alternative [325]. It has been observed that ChatGPT can cause hallucinations and provide false and misleading information. Therefore, users cannot fully trust such a tool and rely on the information provided [326]. Thus, we hypothesize,

H10: CA-based intervention of DV is perceived to be more trustworthy than ChatGPT counterpart.

## 5 EVALUATION OF SAFEHAVEN

### 5.1 Methods

To evaluate the CA, we implemented quantitative (i.e., survey questionnaire) and qualitative (i.e., cognitive walkthrough, interview) methods. We believe both user research methods have a critical role to play in the evaluation of our CA and provide us with a deeper understanding of the user experience. The qualitative method assists us in analyzing users' feelings, thoughts, and behavioral patterns, whereas the quantitative method helps us to analyze the data and present findings in a straightforward and objective manner.

As part of our research, we designed a user study that compared SafeHaven with an internet search and ChatGPT. Since searching online is the most popular means of seeking out information and resources for DV survivors [114], [313], we used it to compare the effectiveness of support provision and other constructs. Further, ChatGPT has recently become increasingly popular as a tool for answering user queries [327], so that we examined how participants perceive the tool in terms of support DV survivors. Thus, in this chapter, we aim to answer the following research question: RQ5. How effective is the CA in providing support to DV survivors? a) How does CA compare with a search engine? b) How does CA compare with ChatGPT?

#### 5.1.1 Task Design

As part of this research, we conducted a user study comparing the CA and Internet search /or ChatGPT on a variety of measures. Considering that DV survivors usually rely on the Internet for guidance and resources [313], we compared SafeHaven with Internet search as a baseline for its effectiveness. In addition, we intend to include ChatGPT as another baseline in our study,

given its increasing popularity as a conversational AI system [327]. We investigated all conditions' advantages and disadvantages in more detail. Throughout the user study, we used hypothetical scenarios based on prior research [328], [329], [330]. These scenarios were designed to convey the impression that they are plausible and realistic [331]. Several scenarios were created that are reflective of real-life situations related to the domestic violence phenomenon. They illustrate more complicated cases of DV in which the abuser exhibits more than one abuse form. These scenarios are crucial as they highlight the intricate relationships between the various types of abuse, as well as the profound effects they can have on victims. These hypothetical scenarios do not inadvertently reinforce stereotypes or victim-blaming attitudes. The scenarios do not place undue responsibility upon the survivors for the abuse they have suffered, and they accurately represent the complexities of abusive relationships. Additionally, these scenarios illustrate how survivors might consider seeking support from hotlines, counseling services, or local resources. The objective of employing hypothetical scenarios in our research is to enable participants to adopt the perspective of the survivor and undertake a user study from their viewpoint. Further, using such scenarios allows us to control the environment in which the study is conducted [332]. Here are examples of hypothetical scenarios:

Scenario 1: Sarah and Mike have been married for nearly six years. The couple's lives were stable at the outset, and their relationship was perfect. Following the birth of their child, Mike began to feel that she was neglecting him. He began criticizing and ridiculing her, calling her names, and unfairly comparing her to other women. In addition, he recently started spying on her phone and tracking her location. He also abused her financially by controlling the joint account between them, as well as refusing to pay the financial obligations of the house, including bills

and other expenses. The matter did not stop there. On one occasion, while they were having dinner, Mike made fun of Sarah's dress, so she politely asked him not to repeat it because it upsets her. An argument ensued, which escalated quickly, and he hit her, resulting in bruises. After this incident, she felt confused, frustrated and feared that more dangerous incidents might occur in the future. She started considering the possibility of seeking assistance from nearby resources but was still confused.

Scenario 2: Marcia and James have been married for nearly four years. At first, things appeared to be going well for the couple. Their lives were stable, and their relationship was ideal. Since moving to a new job, James has been experiencing increased stress and overload. Because of his new job, he often loses his temper. The bad psychological state followed him home, so he got angry with his wife, Marcia, quickly and without reason. He began criticizing and ridiculing her, calling her names, and comparing her unfairly to others. Further, he recently started spying on her social media account and web history and tracking her location. In addition, he abused her financially by managing their joint account and refusing to pay their household expenses, including rent, utilities, and other bills. It did not end there. In one case, while Marcia was cooking lunch, James shouted that he needed a meal immediately. She politely asked him to wait for a few minutes. He lost his temper and pushed her hard. She tried to resist, but he punched her, which resulted in bruises. Following this incident, she started thinking about seeking help and guidance from local resources to create a safety plan and find support during this difficult time. She felt anxious, depressed and feared that future dangerous incidents could happen.

Scenario 3: John and Lisa have been married for the past five years. They enjoyed a good relationship at the beginning, but things began to change when Lisa lost her job and became depressed. During this time, John became increasingly critical and regularly belittled Lisa,

claiming that she was lazy and worthless. Additionally, he would not assist Lisa around the house or with their two children, leaving her to handle everything on her own. One day, while they were arguing, John smashed Lisa's cell phone. Lisa was scared and didn't know what to do. Shortly thereafter, she began to consider safety planning and sought support from local resources.

We randomly assigned each scenario to each system (i.e., SafeHaven, search engine, ChatGPT) to prevent biases and learning effect. We presented these hypothetical scenarios to our prospective participants and asked them to read and imagine themselves in the role of Sarah, Marcia, or Lisa. In the following steps, we asked participants to complete several tasks centered on informational, emotional, and instrumental support. We assume the task list will reflect most of the needs of survivors of DV.

### 5.1.2 Research Instruments

Several research instruments were employed to measure the effectiveness of the CA in comparison with online searches.

#### 5.1.2.1 Social Reactions

The social reaction to disclosure refers to the way informal supports respond in both verbal and nonverbal forms to survivors' disclosures, which is generally categorized into positive (e.g., acknowledging and validating the survivors' experiences) and negative (e.g., doubting and blaming the survivor) reactions [333]. Negative social reactions can exacerbate survivors' psychological distress, while positive social reactions were not related to the psychological distress [334]. In recognition of the importance of this aspect, we intend to use an instrument that can assist us in assessing social reactions. In this regard, we assessed disclosure participants' social reactions by using the revised version of Ullman's Social Reactions Questionnaire (SRQ)



[280]. It was originally developed to assess social reactions to disclosures of sexual abuse, but it has been modified and applied to DV literature as well [71], [334], [335]. The participants were asked how frequently they receive different reactions from different systems (i.e., CA, search engine /or ChatGPT). We chose 14 items from the emotional support and tangible /informational support categories. For each item, the response scale is a 7-point Likert scale (1 = Strongly disagree, 2 = Disagree, 3 = Somewhat disagree, 4 = Neutral, 5 = Somewhat agree, 6 = Agree, 7 = Strongly agree). For the emotional support and belief category, we chose ten items including preventing blame, showing understanding of the survivor's experience, and caring about the survivor's feelings (See Appendix B). Further, we also utilized four items from the tangible aid and information support category, including encouraging seeking counseling and helping to refer to resources (See Appendix B).

#### 5.1.2.2 Information Quality (IQ)

In terms of perceived information quality, it refers to how the information receiver perceives its quality. IQ can be categorized into four sub-categories, including intrinsic, contextual, representational, and accessibility [336]. Intrinsic IQ refers to information itself possessing a quality. Contextual IQ emphasizes that IQ should be viewed in the context of the task. Representational and accessibility IQ highlight the importance of computers being able to store and provide information; that is, information should be presented in a way that is easy to comprehend and manipulate, and that is concise and consistent. Hence, in our study, we aim to use the AIM quality (AIMQ), a questionnaire for measuring IQ, and techniques for interpreting IQ measurements [336]. Various contexts and organizations used this instrument to assess the quality of information. We chose 12 items from the information quality assessment that discuss different aspects, including 1) accessibility: how easily and quickly data can be retrieved, 2)

appropriate amount: data volume is appropriate for the task, 3) believability: the credibility and truthfulness of the data, 4) completeness: adequate breadth and depth for the task, and no missing data, 5) concise representation: how well data is presented in a compact manner, 6) consistent representation: whether the data is presented consistently, 7) ease of operation: how easily data can be manipulated and applied, 8) free of error: the accuracy and reliability of the data, 9) interpretability: clarity and appropriateness of data, 10) relevancy: data relevance, usefulness, and applicability, 11) timeliness: the degree to which the data is current and up-to-date, and 12) understandability: the data can be easily understood. Each item is measured on a scale from 1 to 7, where 1 represents strongly disagree and 7 indicates strongly agree (See Appendix B).

#### 5.1.2.3 Usability

Usability refers to the ability of a particular artifact to serve its intended purpose appropriately [337]. ISO 9241-11 describes usability as having the ability to measure a system's usability based on who uses it, what they use it for, and the environment they use it in. In accordance with ISO 9241-11, usability measures should include the following: a) effectiveness (users' ability to complete tasks and output quality); b) efficiency (effort and resources needed to complete tasks); c) satisfaction (users' subjective responses to the system during use). In our work, we used the System Usability Scale (SUS). Brooke developed SUS in 1996, in what is referred to as a "quick and dirty" usability scale [337]. This method is commonly employed and enables researchers to evaluate the usability of a system quickly and efficiently. It is composed of 10 questions, and participants are asked to choose one of five points ranging from strongly disagree to strongly agree under each question (See Appendix B). Positively worded items contribute a score equal to the scale position minus 1, while negatively worded items will contribute 5 minus the scale position. To calculate SUS, the sum of the item score contributions will be multiplied by 2.5,

which yields a score that ranges from 0 (low perceived usability) to 100 (high perceived usability), where a score greater than 68 is considered above average. In recent years, this measure has been employed to assess the usability of CAs [338].

#### 5.1.2.4 Perceived Anonymity

A person's perceived anonymity is their perception of being unidentifiable (rather than others) [339]. Additionally, anonymity should be viewed as a continuum from total unidentifiability to total identifiability [340]. It is defined that perceived anonymity refers to the feeling that one's identity is unknown to others or that one is not identifiable [339]. To measure perceived anonymity, we adopted the instrument developed by Hite et al. [339] and made a few modifications to reflect the current context. We used this instrument to validate the influence of perceived anonymity on usage and sharing behavior. Five items are included on the scale, such as "I am confident that others do not know who I am when I use the CA (search engine/ChatGPT)" (See appendix B). A seven-point scale is used to rate each item, with the range being strongly disagreed (1) and strongly agreed (7). Each item's score is added together to determine the overall perceived anonymity score, with higher scores suggesting a stronger perception of anonymity.

#### 5.1.2.5 Perceived Trust

In IS research, measuring trust in technology is a key component. A person's trust in technology also plays a significant role in shaping their beliefs and behaviors regarding technology. Thus, to measure the trust of systems (i.e., CA vs. search engine/ChatGPT), we used Jian trust scale [341]. It is unique in that it assesses trust and distrust in two opposite directions in a single dimension instead of assessing trust only in one direction. The output may include both trust and distrust values, or it may be an all-encompassing value. The scale consists of 12 items,

and participants are asked to rate them on a scale ranging from "strongly disagree = 1" to "strongly agree = 7". Some of these items relate to distrust (5 items, Items 1-5), such as perceptions that the automation system is deceptive or that its use will lead to harmful outcomes. Further, the scale includes trust items (7 items, Items 6-12) assessing reliability, integrity, and overall trustworthiness. For both overall and positive trust scores, a higher value signifies greater trust. Conversely, for the negative trust scores, a lower value indicates stronger trust (See appendix B).

### 5.1.3 Procedure

The study was designed as a between-subjects as well as within-subjects study. This involves a group of participants using the developed CA vs. search engine and another group using the developed CA vs. ChatGPT. It combines elements of both within-subject (participants using different conditions) and between-subject (different groups assigned to different conditions) in a randomized order to minimize the effects of individual differences and potential biases (Figure 5.1).

Each session lasted between 75 and 90 minutes. The session includes three phases. First, we asked the participant to complete tasks in both conditions and respond to the cognitive walkthrough (CW) questions regarding SafeHaven system evaluation. Second, as a follow-up to completing the tasks in each system, we asked the participant to answer a post-survey detailing their experiences. Lastly, interviews with participants were conducted to discuss their views on both conditions. A sample of interview questions are: What are the advantages of using SafeHaven versus search engines or ChatGPT to assist DV survivors, respectively? What are the drawbacks of using SafeHaven and search engines/or ChatGPT to assist DV survivors,

respectively? What privacy concerns do they have about using SafeHaven versus search engines or ChatGPT to assist DV survivors? How to improve SafeHaven? (Appendix D).

At first, we explained to the participants the purpose and procedure of the study. Before conducting the user study, we obtained informed consent forms from participants. We collected information about participants' basic demographics and DV experience, or their experience working with DV survivors, and their knowledge and experience of seeking and receiving, or providing social support in relation to DV intervention. In particular, we utilized a widely used tool for screening DV called HITS, which contains four items and asks about the frequency with which the partner physically harming, insulting, threatening, and screaming at them (See Appendix C) [342]. Each participant was then assigned randomly to one condition (i.e., CA vs. search engine, or CA vs. ChatGPT) to perform multiple tasks. For example, Participant 1 was informed that they would utilize CA and a search engine, whereas Participant 2 was assigned to use CA and ChatGPT. We assigned participants randomly to different conditions, meaning that they may either use the CA first and then the search engine/ChatGPT or vice versa. Following this, the participant read a brief description of the study regarding the assigned system and then read hypothetical scenarios that were assigned randomly. They then watched a video that provides step-by-step instructions regarding performing tasks. Each participant then was asked to perform tasks using the CA and search engine/or ChatGPT.

The tasks simulated situations that may occur in the real world and require participants to interact with each system to perform them. Examples of tasks that a participant performed on the developed CA and search engine/or ChatGPT:

1. Create a DV safety plan: This task requires a participant to create a safety plan. It includes several steps like gathering supplies to take with them in an emergency, identifying a safe place to go, and creating an emergency contact list.
2. Find a nearby DV shelter: This task asks a participant to locate a DV shelter in their area, which can provide them with resources and assistance to help them through a difficult time.
3. Learn how to gather evidence of the DV incident: This task asks a participant to learn how to collect information that will help prove the occurrence of DV.
4. Learn how to protect the survivor from future cyberstalking when using her/his mobile device: This task requires a participant to learn about the different methods of protecting a survivor from cyberstalking.
5. Read a survivor story: This task asks a participant to find out a domestic violence survivor story and read it, aiming to understand the difficulties they encounter and their methods of coping.
6. The survivor feels confused and frustrated as if the situation is her fault. Try to find words of encouragement, comfort, and empathy for her: This task asks a participant to determine whether they find words that show encouragement, comfort, or empathy.

After completing each task on the developed CA, participants completed cognitive walkthrough wherein they answered questions about the task they just completed. They could then move on to the next task until they have completed all of the tasks. Once a participant has completed tasks for a condition, we asked them to complete a post-survey detailing their experiences regarding several measures. Lastly, we had short interviews with the participants to discuss their views on a condition (i.e., CA, and online searches/or ChatGPT). Through our

interview, we asked about both strengths and drawbacks in both conditions and how CA-based interventions for DV could be improved. Each interview lasted 15-20 minutes on average.

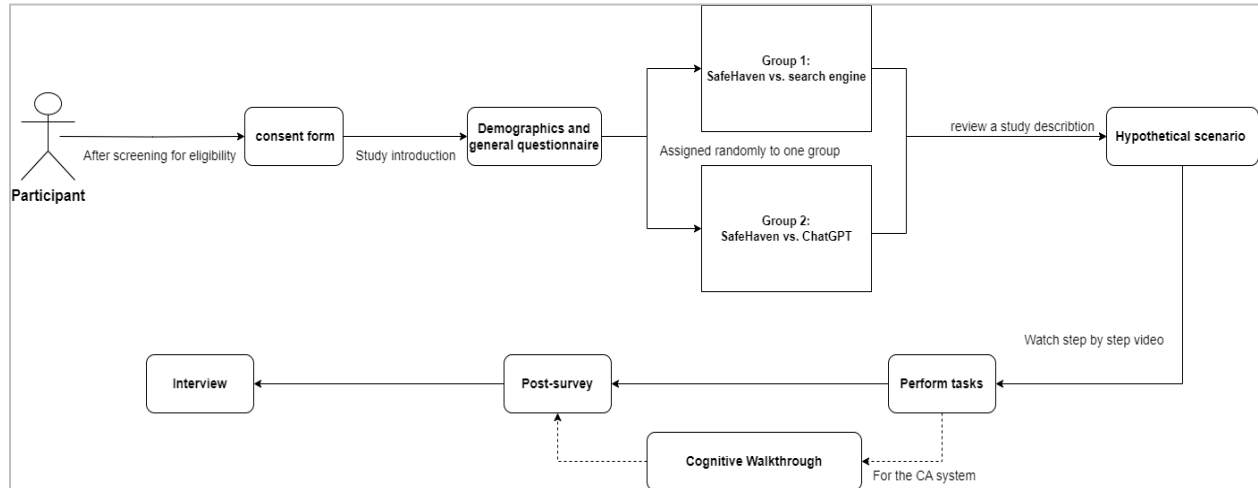


Figure 5.1: Task procedure

#### 5.1.4 Cognitive Walkthrough

In our study, we selected Cognitive Walkthrough (CW) as the research method. CW refers to a theoretically based usability evaluation method that is applicable throughout all phases of system design and development and concentrates on the ease of use for new users to complete tasks with the system [343], [344]. In other words, CW places an emphasis on understanding how new and infrequent users may be able to utilize the system. It is important to prepare a task list that contains all of the tasks that will be performed during the walkthrough (see section 5.1.3 for sample tasks), along with an action sequence that outlines the flow of the specific tasks. A participant walked through the tasks using the CA and answered questions related to the usability at each step. As part of the walkthrough, evaluators walk through the action sequences using the artifact (i.e., CA) and answered questions at each step [344]. These are four typical questions: 1) Will the user try to achieve the correct outcome? 2) Does the user recognize that the right action is available? 3) Will the user likely connect the description of the correct action to what is being

attempted? 4) Will the user be able to observe progress toward solving the task if the correct action is performed? During the predefined action sequence, the evaluator's objective is to assess the likelihood that the user will succeed at each stage. It is also important to document the reasons for this determination (See Appendix E).

#### 5.1.5 Participants Recruitment

A total of 36 participants were recruited, including 22 people who had experienced DV, 22 friends or family members of DV victims, and two professionals who had experience in the field, giving us a diverse group of individuals with different perspectives on the topic. We utilized a comprehensive recruitment strategy, including internet searches, snowball sampling, mailing lists, and social media outreach. Specifically, we considered adopting a university mailing list as a means of reaching out to a larger number of participants. Furthermore, we posted in social media communities and groups, such as "Domestic Violence Survivor's," "Domestic Violence Awareness," in Facebook, and subreddits like r/abusesurvivors and r/emotionalabuse in Reddit, to reach a large audience and gathered participants who had direct or indirect experiences with DV. In addition, we searched online to find participants from DV organizations and academic institutes in different cities of the U.S. with experience in DV support and advocacy.

A diverse range of experiences and awareness of DV are revealed in the demographic data (Table 5.1). The data indicates a notable inclination among young individuals and females to have discussions about DV, with 44.4% of respondents being between 18-25 and 75% being females. There is broad representation among the respondents in terms of racial and ethnic backgrounds, with the majority identifying themselves as White (55.26%). In terms of personal connections to DV, a significant percentage of participants identify as victims/survivors (61.11%), whereas an identical percentage identify as family or friends of victims/survivors. In



most cases, participants reported receiving support from their personal networks, mainly from friends (27.62%) and family (24.76%), highlighting the importance of maintaining close relationships while coping with DV. While the overall utilization of technology in DV interventions among participants is relatively scarce, some participants have resourcefully utilized digital tools such as online therapy sites, search engines for local services, social media for community support, anonymous legal advice, and VPN, thus suggesting a nascent but promising area of technological integration in DV support systems.

Participants' perceptions and use of CAs differ according to their familiarity with and trust in the technology (Table 5.1). The majority of participants reported using a CA a few times a week (47.22%), indicating regular engagement, while a smaller proportion utilized them less frequently, ranging from once a week to once a month. The surveyed participants have engaged with various CAs, predominantly ChatGPT (75.00%), Apple Siri (63.89%), Amazon Alexa (55.56%), and Google Assistant (41.67%), bringing a wide range of experience levels to their interactions, with 36.11% identified as beginners, 25% were intermediates, another 25% were advanced beginners, and 13.89% were proficient, ranging from nascent familiarity to adept usage, but none considered themselves experts. Only a few respondents have reported using other alternatives like Bing chatbot, Hugging Chat, and Snapchat Ai. For CA interactions, WhatsApp and Facebook Messenger were the preferred platforms among participants, chosen by 63.89% and 55.56% of participants, respectively, with Telegram, Slack, and other specified platforms like Instagram and Snapchat also preferred by some participants for CA interactions. Interestingly, participants predominantly employ CAs for textual tasks and applications (72.22%), such as querying, content creation, summarization, and writing aids, as well as for educational purposes (44.44%). As a whole, the participants appeared to take a positive stance

towards the use of CAs, with 66.66% agreeing or strongly agreeing that they found them comfortable and 72.22% indicating that they find them easy to use. In the same vein, participants exhibited a moderate level of trust in CA-provided information, with the largest segment, 38.89%, somewhat agreeing on its accuracy, while only 5.56% strongly agree. Most participants, with 55.56% strongly agreeing and 27.78% agreeing, are optimistic about CAs' future prevalence, but only 19.44% said that CAs are capable in emergency situations, reflecting a cautious attitude about their effectiveness.

Table 5.1: Descriptive statistics from demographic and general survey responses (N = 36)

Question	Value	Count (N)	Percentage (%)
Age	18-25	16	44.44%
	26-33	12	33.33%
	34-41	3	8.33%
	42-49	4	11.11%
	50-57	1	2.78%
	58-65	0	0.00%
	>65	0	0.00%
Gender	Female	27	75.00%
	Male	9	25.00%
Race/ethnicity	American Indian or Alaskan Native	0	0.00%
	Asian / Pacific Islander	6	16.67%
	Black or African American	4	11.11%
	Hispanic or Latino	6	16.67%
	White	21	58.33%
	Multiple ethnicities	1	2.78%
Connection to DV	Victim or survivor of DV	22	61.11%

	Professional (domestic violence advocate, researcher, therapist, program director)	2	5.56%
	Family or friend of a victim/or survivor of DV	22	61.11%
Type of support being received	Advocacy groups or organizations	10	27.78%
	Family Support	26	72.22%
	Policing (e.g., seeking justice and/or protection order)	13	36.11%
	Community support (School, Care service center, etc.)	9	25.00%
	Friend	29	80.56%
	Informational support (tips, advice)	16	44.44%
	Others	2	5.56%
Use of technology-related DV interventions	Yes	9	25.00%
	No	27	75.00%
CA Questionnaire			
Frequency of chatbots usage	One or several times a day	2	5.56%
	A few times a week	17	47.22%
	Once a week	7	19.44%
	Once a month	8	22.22%
	Never	2	5.56%
CA(s) being used	Apple Siri	23	63.89%
	Amazon Alexa	20	55.56%
	Google Assistant	15	41.67%
	Microsoft Cortana	4	11.11%
	Cleverbot	1	2.78%
	Mitsuku	0	0.00%
	Replika	0	0.00%
	ChatGPT	27	75.00%

	Others	6	16.67%
Experience level with the CA(s)	Beginner (Little Experience)	13	36.11%
	Intermediate (Working Knowledge)	9	25.00%
	Proficient	5	13.89%
	Expert	0	0.00%
	Advanced beginner	9	25.00%
Preferred messaging platform(s) for a CA	Facebook Messenger	20	55.56%
	WeChat	2	5.56%
	Line	0	0.00%
	WhatsApp	23	63.89%
	Slack	4	11.11%
	Telegram	6	16.67%
	Others	6	16.67%
Usage of CAs	E-commerce and retail (e.g., prices and products notifications, purchase assistance, customer support)	15	41.67%
	Travel and hospitality (e.g., make reservations, refund assistance)	9	25.00%
	Banking (e.g., personal financial information, banking services)	10	27.78%
	Healthcare (e.g., insurance and billing, booking appointments)	10	27.78%
	Media and entertainment (e.g., news, subscription management)	9	25.00%
	Education (e.g., student training and assistance, updates and announcements)	16	44.44%
	Textual tasks and applications (e.g., asking questions, summarizing/ or generating content, writing assistance)	26	72.22%
	Others	4	11.11%
Comfort when using CAs	Strongly agree	7	19.44%

	Agree	17	47.22%
	Somewhat agree	6	16.67%
	Neutral	5	13.89%
	Somewhat disagree	1	2.78%
	Disagree	0	0.00%
	Strongly disagree	0	0.00%
Easy to use CAs	Strongly agree	9	25.00%
	Agree	17	47.22%
	Somewhat agree	7	19.44%
	Neutral	2	5.56%
	Somewhat disagree	1	2.78%
	Disagree	0	0.00%
	Strongly disagree	0	0.00%
Trust accuracy of CAs' information	Strongly agree	2	5.56%
	Agree	6	16.67%
	Somewhat agree	14	38.89%
	Neutral	9	25.00%
	Somewhat disagree	2	5.56%
	Disagree	3	8.33%
	Strongly disagree	0	0.00%
Prevalence of CAs in future	Strongly agree	20	55.56%
	Agree	10	27.78%
	Somewhat agree	4	11.11%
	Neutral	1	2.78%
	Somewhat disagree	1	2.78%
	Disagree	0	0.00%

	Strongly disagree	0	0.00%
Managing emergency situations	Strongly agree	0	0.00%
	Agree	7	19.44%
	Somewhat agree	13	36.11%
	Neutral	10	27.78%
	Somewhat disagree	2	5.56%
	Disagree	2	5.56%
	Strongly disagree	2	5.56%

#### 5.1.6 Ethical Consideration

Before conducting this study, a review and approval of the protocol have been granted by the Institutional Review Board (IRB) at the University of North Carolina at Charlotte, and we adhered to the approved protocol throughout the user study (IRB-23-0836). The entire user study was conducted online via a video conferencing tool (i.e., Zoom). Participants were first prescreened and checked to see if they were eligible. Once their eligibility was confirmed, we sent them an online research consent form encompassing the research protocol and allowing them to provide voluntary consent. The consent form communicates to participants that their involvement in the study is entirely voluntary. They may choose not to participate in the study, or even if they start participating and then change their mind, they can stop it at any time. Participants were then informed that the topic is sensitive and may lead to distressing feelings or memories as they answer questions or read hypothetical scenarios. We offered the option of skipping any questions they did not wish to answer or withdrawing from the study if they wished to do so. Furthermore, drawing from insights garnered in a prior study involving interviews with domain experts in the context of DV [345], a range of measures were available to address potential crises, such as comforting them and referring them to appropriate resources. Moreover,

we assured them that their data would remain confidential and only accessible to the research team. Participants' identifiers (i.e., names and e-mail addresses) were used solely for the purposes of initial communication, scheduling, and obtaining informed consent. They were not used in the analysis or interpretation of research results.

#### 5.1.7 Data Analysis

In the quantitative analysis, we employed one-tailed paired t-tests for normally distributed data and Wilcoxon signed-rank tests for non-normal distributed data to evaluate our research question 5 and hypotheses. Further, in qualitative analysis, we conducted a thematic analysis. We used an inductive approach that focuses on reading interview transcripts and comments from the cognitive walkthrough and extracting themes and patterns that emerge. We open-coded sample scripts, and then revised the coding schema so that new categories and themes could be added. We repeated the process several times to achieve more refined and differentiated themes until saturation is achieved. As a final step, the entire data set was coded based on the coding schema we developed.

### 5.2 Quantitative Analysis Results

In this section, we investigated participants' perceptions of interacting with SafeHaven compared to using a traditional search engine or ChatGPT to gain deep insights into the relative effectiveness of SafeHaven. During our investigation, we examined multiple dimensions of the user experience, focusing on social reactions, information quality, usability, anonymity, and perceived trust. To test normality, we performed a Shapiro-Wilk test on each measure. We performed a paired samples t-test on normally distributed measures and a Wilcoxon signed-rank test on non-normally distributed measures to test our hypotheses.

We first report the results of a comparison between SafeHaven and a search engine, followed by a comparison between SafeHaven and ChatGPT. A summary of the results is presented in Table 5.2 and Table 5.3.

Table 5.2: Descriptive statistics and statistical comparisons of the effectiveness between SafeHaven and search engine

Condition	Variables	Mean (SD)		Z/t-value	P-value
		SafeHaven	Search engine		
SafeHaven vs. Search engine (N=18)	Emotional support	<b>5.48 (1.16)</b>	4.02 (1.46)	Z=-3.268	0.0005***
	Informational and instrumental support	<b>6.00 (1.05)</b>	5.25(1.23)	Z=-2.483	0.0065**
	Information quality	<b>5.91 (0.95)</b>	5.43 (1.006)	Z=-2.107	0.0175*
	Usability	74.02 (18.45)	74.02 (17.34)	t=0.00	1.00
	Perceived anonymity	<b>5.80 (0.915)</b>	5.02 (1.335)	Z=-2.123	0.017*
	Perceived trust	<b>5.81 (0.87)</b>	5.32 (0.88)	t=2.165	0.022*

Note: \*p < 0.05, \*\*p< 0.01, \*\*\*p<0.001.

Table 5.3: Descriptive statistics and statistical comparisons of the effectiveness between SafeHaven and ChatGPT

Condition	Variables	Mean (SD)		Z/t-value	P-value
		SafeHaven	ChatGPT		
SafeHaven vs. ChatGPT (N=18)	Emotional support	<b>6.016 (0.84)</b>	4.47 (1.19)	t=4.594	0.00012***
	Informational and instrumental support	<b>6.38 (0.81)</b>	5.61 (1.09)	Z=-2.591	0.005**
	Information quality	<b>6.31 (0.43)</b>	5.89 (0.63)	Z=-2.287	0.011*
	Usability	83.19 (12.77)	79.30 (13.58)	t=1.162	0.261



	Perceived anonymity	6.11 (1.18)	6.14 (0.70)	Z=0.399	0.345
	Perceived trust	<b>6.33 (0.53)</b>	6.03 (0.69)	Z=-2.445	0.007**

Note: \*p < 0.05, \*\*p < 0.01, \*\*\*p<0.001.

## 5.2.1 SafeHaven Vs. Search Engine

### 5.2.1.1 Emotional Support

A Wilcoxon Signed-Rank Test results revealed that the emotional support rating for SafeHaven (M = 5.48, SD = 1.16) was significantly higher (p = 0.0005, one-tailed) than that for the search engine (M = 4.02, SD = 1.46) (Figure 5.2). Specifically, 17 out of 18 participants rated SafeHaven higher for emotional support than the search engine, resulting in positive rankings with a mean rank of 9.44 and a total rank of 160.50. Thus, hypothesis H1, the CA-based intervention of DV provides greater emotional support than search engines, is supported.

The moderating effects of age and gender were examined. Compared to the search engine (M = 3.850, SD = 1.5831), females rated SafeHaven higher regarding emotional support (M = 5.283, SD = 1.3717) with a p-value of 0.012. In the same manner, males viewed the CA as a more effective (p = 0.028) means of receiving emotional support (M = 5.883, SD = 0.4070) than utilizing a search engine (M = 4.367, SD = 1.2675), signifying a meaningful disparity in their preferences. The findings suggested that both females and males showed a preference for SafeHaven over a search engine in terms of emotional support. Furthermore, the age-related findings were noteworthy. In the group aged 18-25, SafeHaven was more warmly received (p < 0.012) for providing emotional support (M = 6.050, SD = 0.6071) compared to a search engine (M = 4.513, SD = 1.6164). Moving to the 26-33 age cohort, there was a preference for SafeHaven (M = 5.243, SD = 1.3551, p < 0.018) to a search engine (M = 3.486, SD = 1.560). However, the trend did not hold as strongly for 42-49-year-olds. The average CA support rating

( $M = 3.900$ ,  $SD = 1.1134$ ) was slightly lower compared to search engines ( $M = 4.100$ ,  $SD = 0.8485$ ), with no significance of difference was indicated in such a comparison ( $p = 0.655$ ).

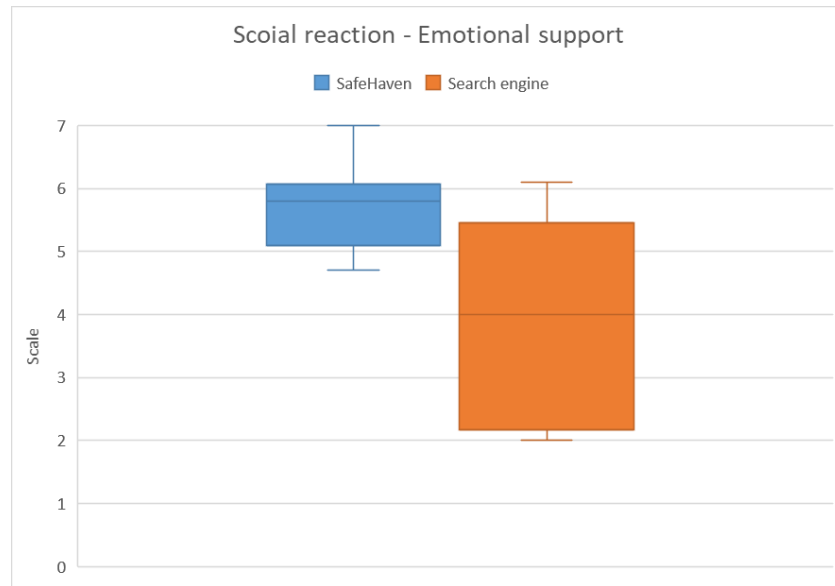


Figure 5.2: Boxplots of emotional support of SafeHaven and a search engine

#### 5.2.1.2 Informational and Instrumental Support

Based on the Wilcoxon Signed-Rank Test, a substantial difference was found in perceptions of informational and instrumental support provided by SafeHaven as compared to the search engine (Figure 5.3). Specifically, we found SafeHaven was rated significantly higher ( $M = 6.00$ ,  $SD = 1.05$ ,  $p = 0.0065$ ) than a search engine ( $M = 5.25$ ,  $SD = 1.23$ ) in providing survivors with the support that facilitates the acquisition of healthcare and shelter, sharing vital information, provides options, assists survivors in obtaining information about coping with traumatic experiences, and refers them to counseling services. SafeHaven received a maximum rating of 7.00, indicating that some participants found it to be particularly helpful. Based on the positive ranks ( $M = 9.18$ , Sum of Ranks = 128.50), 14 out of 18 participants rated SafeHaven as providing more information than the search engine. In addition, SafeHaven was rated lower in

only three instances (negative ranks), with a mean of 8.17 and a total rank of 24.50 and tied with the search engine in one instance. The results provide support for hypothesis H2 that the CA-based intervention of DV provides informational and instrumental support in a more effective manner than a search engine.

An in-depth analysis of the moderating effects of age and gender revealed insightful findings. Males rated SafeHaven significantly higher ( $M = 6.25$ ,  $SD = 0.975$ ) than a search engine ( $M = 5.25$ ,  $SD = 1.10$ ), with a  $p = 0.026$ , inferring that males prefer the CA in terms of seeking informational and instrumental support. Although females perceived SafeHaven to be more helpful for information and instrumental assistance ( $M = 5.875$ ,  $SD = 1.105$ ) than a search engine ( $M = 5.25$ ,  $SD = 1.336$ ), the difference between the two platforms was not significant ( $p = 0.118$ ). Further, when comparing the preference for informational and instrumental support from SafeHaven with a search engine across age groups, the 26-33 age group showed a statistically significant preference for the CA ( $M = 6.36$ ,  $SD = 0.79$ ,  $p = 0.018$ ). Other age groups, including 18-25 and 42-49, showed no significant differences in the way they rated SafeHaven and a search engine.

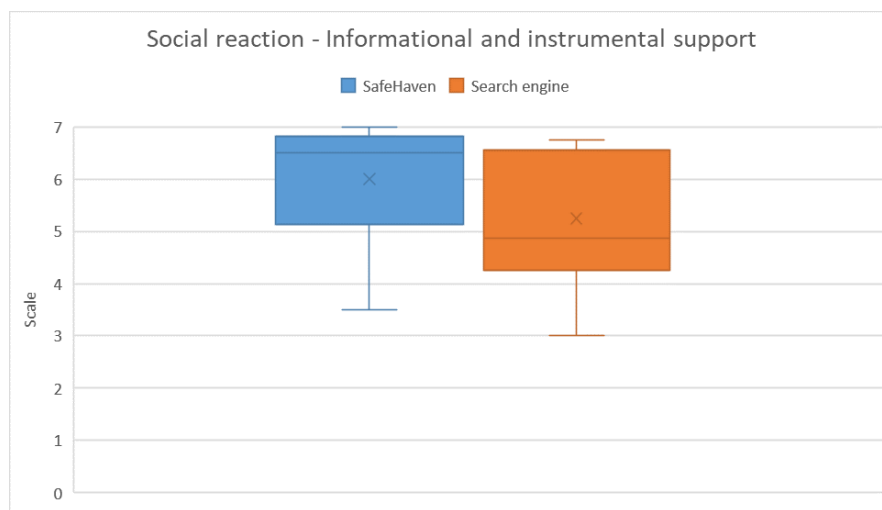


Figure 5.3: Boxplots of informational and instrumental support of SafeHaven and a search engine

### 5.2.1.3 Information Quality

We analyzed the perceived information quality between SafeHaven and a search engine across the 12 AIMQ variables. SafeHaven achieved a higher overall score ( $M = 5.91$ ,  $SD = 0.951$ ) in comparison with that of a search engine ( $M = 5.43$ ,  $SD = 1.006$ ), (one-tailed  $p = 0.0175$ ). SafeHaven also outperformed search engines across all items (Table 5.4). However, only some of the differences were statistically significant, including accessibility ( $p = 0.0320$ ), consistent information representation ( $p = 0.0490$ ), and error-free content ( $p = 0.0005$ ). Therefore, we can conclude that hypothesis H5 is partly supported.

An examination of the moderating effects of age and gender was conducted. Regarding gender, we found that female participants had mean scores of 5.8 for SafeHaven and 5.6 for a search engine, but the  $p$ -value was 0.328, suggesting that neither medium is significantly preferred. For males, however, SafeHaven's mean score of 6.1 significantly outperformed a search engine's mean of 5.0, as indicated by a  $p$ -value of 0.046. Regarding age, and according to the results, we did not observe any significant difference between SafeHaven and a search engine among participants of different ages. The average ratings were comparable for the 18-25 age group with SafeHaven scoring 6.22 compared to a search engine score of 5.78. Likewise, participants aged 26-33 rated SafeHaven at 5.95 compared to a search engine at 5.39, with both age groups showed no significant differences, as indicated by  $p$ -values of 0.160 and 0.237, respectively. Similarly, those aged 42-49 did not display any significant preference, as indicated by a  $p$ -value of 0.317. It appears that age is not a decisive factor in the preference for using SafeHaven over a search engine when it comes to information quality.

Table 5.4: Descriptive statistics and statistical comparisons of user perceptions of information quality between SafeHaven and search engine

Item	SafeHaven (Mean (SD))	Search Engine (Mean (SD))	Z	P-value (one- tailed)
Information is easily accessible	<b>6.22 (0.943)</b>	5.50 (1.249)	-1.852	0.0320*
Information is of sufficient volume for my needs	5.56 (1.338)	5.44 (1.580)	-0.196	0.4225
Information is believable	6.06 (1.162)	5.61 (1.195)	-1.469	0.0710
Information is sufficiently complete for my needs	5.61 (1.335)	5.28 (1.179)	-1.064	0.1435
Information is presented concisely	5.61 (1.577)	5.11 (1.568)	-0.952	0.1705
Information is represented in a consistent format.	<b>5.89 (1.278)</b>	5.22 (1.555)	-1.652	0.0490*
Information is easy to manipulate to meet my needs.	5.67 (1.138)	5.17 (1.618)	-1.181	0.1190
Information is correct.	<b>6.17 (0.707)</b>	5.39 (0.778)	-3.276	0.0005***
It is easy to interpret what information means.	5.89 (1.278)	5.39 (1.539)	-1.082	0.1395
Information is useful to the task at hand.	6.22 (1.003)	5.89 (0.758)	-1.303	0.0965
Information is sufficiently timely.	5.89 (1.451)	5.44 (1.097)	-1.517	0.0645
Information is easy to understand.	6.17 (1.098)	5.72 (1.364)	-1.218	0.1115
Overall (average)	<b>5.91 (0.951)</b>	5.43 (1.006)	-2.107	0.0175*

Note: \* $p < 0.05$ , \*\*\* $p < 0.001$ .

#### 5.2.1.4 Usability

The SUS scores of SafeHaven and a search engine turned out to be the same (74.028,  $p = 1.00$ ) (Figure 5.4), and both exceeded the minimum threshold of 68 [346]. Therefore, it can be inferred from these results that participants saw both SafeHaven and a search engine as practical and usable. The t-test indicated no significant difference in usability between a search engine and SafeHaven.

In a comparative analysis, the study investigated usability preferences for SafeHaven and a search engine based on gender. According to female participants, a search engine received a mean usability rating of 79.16 (SD = 17.33), and SafeHaven had a mean of 73.75 (SD = 20.04), while no statistically significant difference was observed ( $p = 0.317$ ). Males, however, found SafeHaven more usable than a search engine, with a mean score of 74.583 (SD = 16.53) versus 63.75 (SD = 13.11), though this difference was not statistically significant ( $p = 0.293$ ). As a result of paired t-tests conducted across a variety of age groups, detailed findings regarding both SafeHaven and a search engine usability have been uncovered. In the analysis, participants aged 18 to 25 showed almost equal usability ratings for SafeHaven and a search engine (73.75 and 73.125), with no distinction between them ( $p = 0.951$ ). For those aged 26 to 33, SafeHaven received an average rating of 77.5, while a search engine received an average rating of 76.071. It was found, but no significant difference ( $p=0.817$ ) existed between the two platforms. Based on the results, among 42-49-year-olds, search engines scored 57.5 compared to SafeHaven, 50, suggesting a preference for search engines, although this discrepancy did not reach statistical significance ( $p = 0.374$ ).

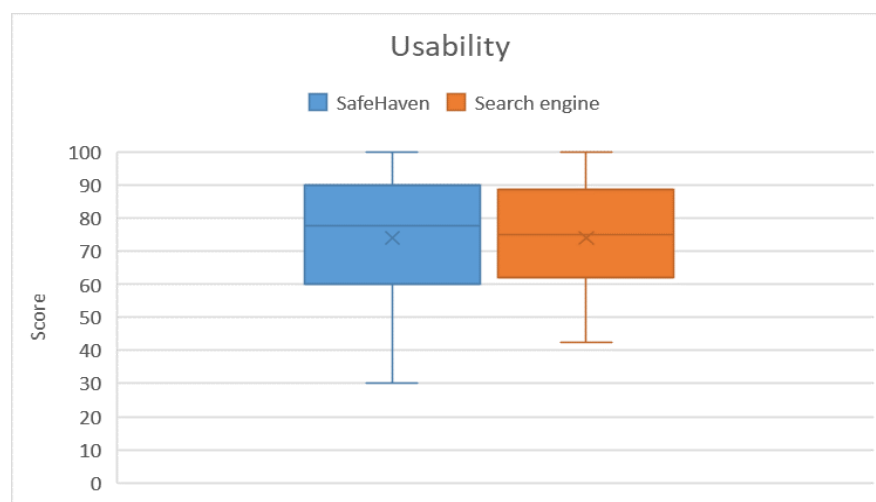


Figure 5.4: Boxplots of usability of SafeHaven and a search engine

#### 5.2.1.5 Perceived Anonymity

The descriptive statistics showed that participants felt a higher level of anonymity with SafeHaven ( $M = 5.80$ ,  $SD = 0.915$ ,  $p = 0.017$ ) compared to a search engine ( $M = 5.02$ ,  $SD = 1.335$ ) (Figure 5.5). Specifically, 11 out of 18 participants gave a higher rating to SafeHaven, five participants felt a search engine was more anonymous, and two participants reported no preferences. Based on these findings, the results support hypothesis H7.

Based on the analysis of the moderating effects of gender, we found females rated the anonymity of SafeHaven at an average of 6.00, compared to an average of 5.06 for a search engine, with a statistically significant difference based on the results of a Wilcoxon Signed Ranks Test ( $p = 0.026$ ). Similarly, male participants viewed SafeHaven as more anonymous than a search engine, scoring 5.40 on average for SafeHaven versus 4.93 for a search engine, but the difference was not statistically significant ( $p = 0.684$ ). The analysis results of perceived anonymity for SafeHaven vs. a search engine between different age groups showed that the patterns varied by age. The younger cohort (18-25) significantly ( $p = 0.046$ ) leaned toward SafeHaven ( $M = 6.150$ ) over a search engine ( $M = 4.850$ ), valuing the anonymity that SafeHaven offers. In contrast, the group of 26-33-year-olds did not exhibit a marked preference for a more anonymous medium, with both SafeHaven and a search engine scoring nearly equally ( $M = 5.686$  and  $5.543$ , respectively;  $p = 0.866$ ). In addition, although the 42-49 age group perceived SafeHaven ( $M = 5.20$ ) as more anonymous over a search engine ( $M = 4.10$ ), this preference is not statistically significant ( $p = 0.180$ ), suggesting that perceived anonymity has little influence on their preferences.

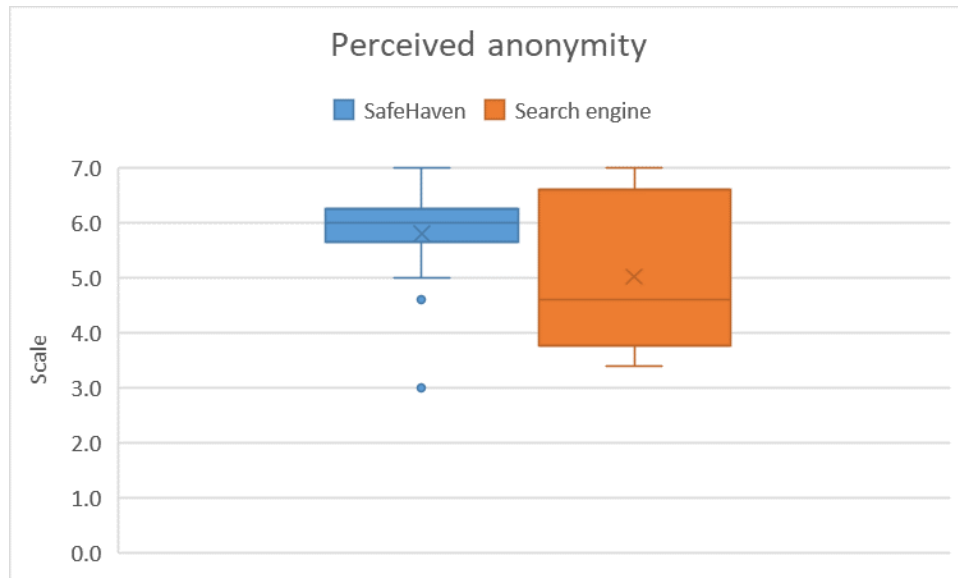


Figure 5.5: Boxplots of the perceived anonymity of the SafeHaven and a search engine

#### 5.2.1.6 Perceived Trustworthiness

The trust score was computed by averaging the positive items and reverse coding of the negative items, with higher values indicating greater trust [341]. Figure 5.6 shows the mean score for positive and negative trust items, as well as the overall trust score. Based on the paired t-test results, participants perceived SafeHaven to be more trustworthy ( $M = 5.81$ ,  $SD = 0.87$ ,  $p = 0.022$ ), in comparison to a search engine ( $M = 5.32$ ,  $SD = 0.88$ ). Thus, the results support H9 positing that the CA-based intervention of DV would be viewed as more trustworthy than a search engine counterpart.

We examined the moderating effect of gender. Based on the paired samples t-test results, both genders exhibited a higher level of trust in SafeHaven over a search engine when seeking assistance regarding DV. In particular, female participants rated SafeHaven as more trustworthy, with an average of 5.80, compared to a search engine with an average of 5.46. However, the difference was not statistically significant ( $p = 0.187$ ). In parallel, when searching for assistance



with DV, male participants were more likely to trust SafeHaven ( $M = 5.80$ ) than a search engine ( $M = 5.04$ ). However, like the female participants, there was no significant difference in the ratings of male participants ( $p = 0.167$ ). We further investigated how age influences the outcome. According to the results, the 18-25 age group showed a slightly higher degree of trust in SafeHaven with a rating of 5.94, compared to a rating of 5.44 for a search engine. The 26-33 age group demonstrated a clear preference for SafeHaven, scoring 5.80 against 5.23 for a search engine; the 42-49 age group had the smallest preference difference (5.00 versus 4.75). Despite this, the preferences for SafeHaven across different age groups did not show statistically significant results ( $p > 0.5$ ).

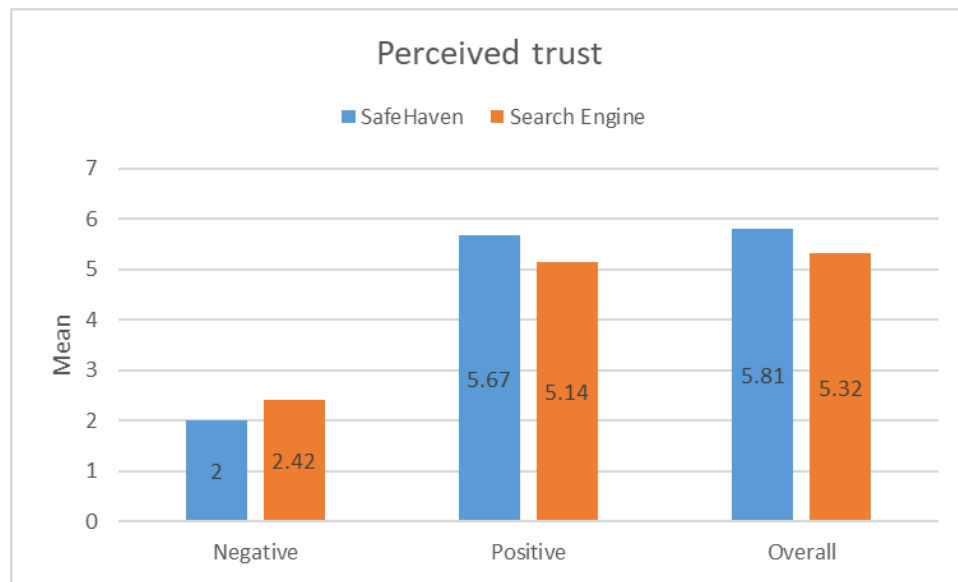


Figure 5.6: Bar charts of perceived trust of SafeHaven and a search engine

## 5.2.2 SafeHaven vs. ChatGPT

### 5.2.2.1 Emotional Support

The results of the paired t-tests reveal that SafeHaven was rated significantly higher ( $p = 0.001$  for its ability to provide emotional support ( $M = 6.016$ ,  $SD = 0.847$ ) than ChatGPT ( $M =$

4.47, SD = 1.198) (see Figure 5.7). Among the multiple dimensions of emotional support are SafeHaven's capabilities to listen attentively, its non-judgmental nature, its ability to show understanding and empathy, as well as its overall acceptance and support ( $p < 0.001$ ). Thus, the results support hypothesis H3, that the CA-based intervention of DV is perceived to be more emotionally supportive than ChatGPT.

By analyzing the moderating effects of gender, SafeHaven outperforms ChatGPT in terms of emotional support. According to the results, females reported a mean score of 5.947 for the SafeHaven compared to 4.513 for ChatGPT, which is statistically significant, with a p-value of 0.003 and a mean difference of 1.4333 for SafeHaven CA. Males exhibited an even stronger perception of offering emotional support (p-value of 0.01), scoring 6.367 for SafeHaven and 4.267 for ChatGPT. Across different age groups, participants reported a significant difference between the emotional support offered by SafeHaven and ChatGPT. Notably, the younger cohort (18-25) showed a clear preference toward SafeHaven CA over ChatGPT, scoring an average of 6.40 compared to ChatGPT's 3.76, with a p-value below 0.001. Further, those aged 34-41 preferred SafeHaven over ChatGPT, as evidenced by average scores of 5.00 and 4.35, respectively, showing a significant level of 0.049 on average. In contrast, other age groups, including both 26-33 and 42-49 groups, perceived SafeHaven as providing more emotional support than ChatGPT, yet neither was statistically significant ( $p = 0.364$ ,  $p = 0.5$ , respectively).

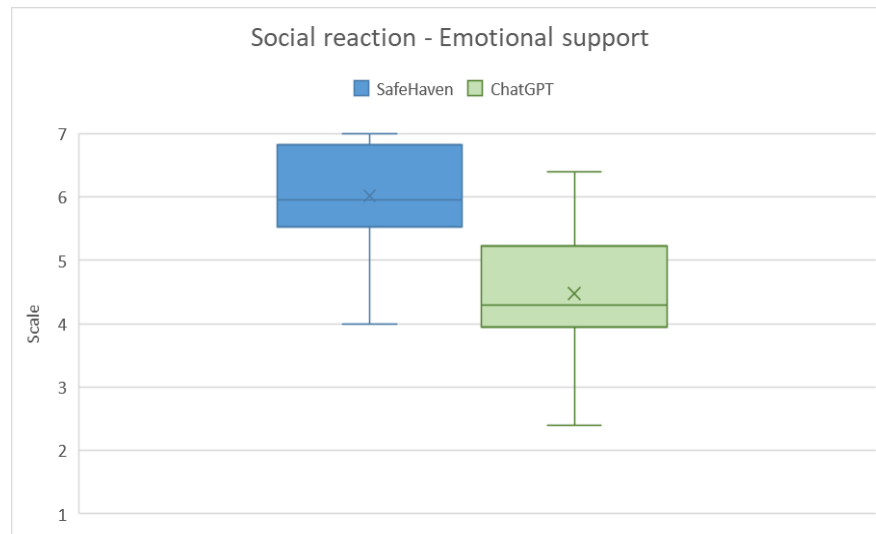


Figure 5.7: Boxplots of emotional support of SafeHaven and ChatGPT

#### 5.2.2.2 Informational and Instrumental Support

The descriptive statistics of informational and instrumental support are reported in Figure 5.8. The results of the Wilcoxon Signed-Rank Test show that SafeHaven is perceived as providing higher levels ( $p = .005$ ) of informational and instrumental support ( $M = 6.38$ ,  $SD = 0.81$ ) in comparison with ChatGPT ( $M = 5.61$ ,  $SD = 1.09$ ). Among the participants, 12 rated SafeHaven's support more favorably, 3 participants preferred ChatGPT, and the remaining three had no preference. Thus, these findings robustly support hypothesis H4, that the CA-based intervention of DV outperforms ChatGPT in providing both informational and instrumental support.

According to an analysis of how females and males perceived informational and instrumental support from SafeHaven in comparison with ChatGPT, data shows that female participants rated SafeHaven higher ( $M = 6.31$ ) than ChatGPT ( $M = 5.83$ ), showing a marginally significant  $p$ -value of 0.053. In addition, males scored SafeHaven higher than the ChatGPT, with a mean score of 6.75 compared to 4.50, although a  $p$ -value of 0.109 indicates that this finding is not statistically significant. While both genders seemed to prefer the CA, no significant difference

was observed. In the same context, based on the comparison of the perceived informational and instrumental support between SafeHaven and the ChatGPT across different age groups, SafeHaven received higher mean scores than the ChatGPT across all age groups. According to the results, participants between the ages of 18-25 scored SafeHaven at 6.46, compared to 5.56 for ChatGPT, with a p-value near-significant at 0.075. Also, in the 26-33 age group, SafeHaven was perceived to provide more information and practical assistance (mean score of 6.70 versus 5.80 for ChatGPT), but with a p-value of 0.285, indicating less significant differences. Among the 34-41 and 42-49 age groups, SafeHaven had an average of 4.87 and 6.87, while ChatGPT rated 4.37 and 6.00 respectively. Although SafeHaven was not statistically significant, there was a slight preference for it.

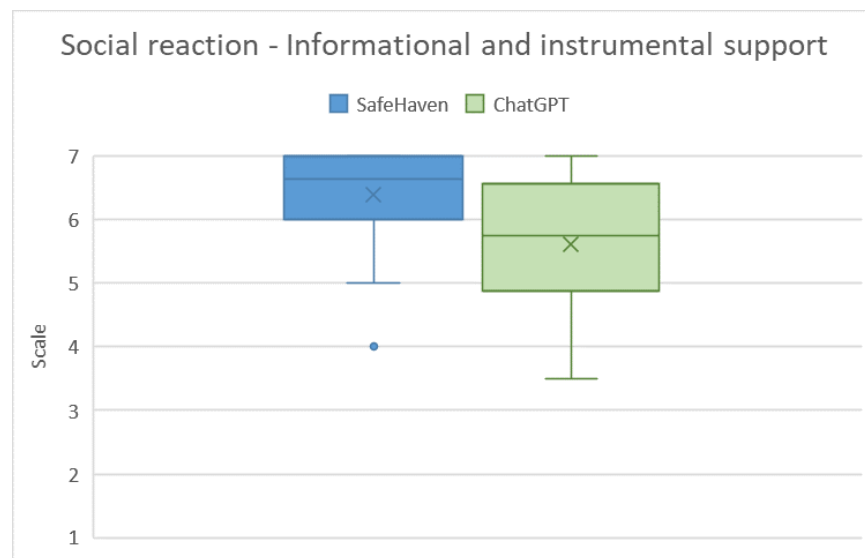


Figure 5.8: Boxplots of informational and instrumental support of SafeHaven and ChatGPT

### 5.2.2.3 Information Quality

The results of the Wilcoxon signed-rank test show that the overall information quality score for SafeHaven ( $M = 6.31$ ,  $SD = 0.431$ ) was significantly higher ( $p = 0.011$ ) than that for ChatGPT ( $M = 5.8981$ ,  $SD = 0.6313$ ). Specifically, some items of information quality showed

significant superiority of SafeHaven over ChatGPT, including the concise presentation of information ( $p = 0.0085$ ), the ease of understanding the information provided ( $p = 0.017$ ), and the sufficient volume of information ( $p = 0.0385$ ) (Table 5.5). Thus, hypothesis H6 is partly supported.

In our examination of age and gender moderating effects, we compared how different genders perceive the quality of information provided by SafeHaven and ChatGPT. According to the Wilcoxon Signed Ranks Test, females gave higher ratings to SafeHaven ( $M = 6.3167$ ) than to ChatGPT ( $M = 6.05$ ), with a  $p$ -value of 0.041. Likewise, males rated SafeHaven higher than ChatGPT ( $M = 6.27$  vs.  $M = 5.36$ ); despite this, there was no significant statistical variance ( $p = 0.28$ ) between the two mediums. In the same context, when comparing how different age groups perceived the information quality of SafeHaven versus ChatGPT, we found those between 18-25 years viewed SafeHaven to offer higher information quality with an average at 6.34, slightly higher than ChatGPT at 5.90. For the 26-33 age group, SafeHaven had an average of 6.30 against 5.93 for ChatGPT. Participants aged 34-41 rated SafeHaven at 6.33, more than ChatGPT at 5.29. Lastly, the 42-49 age group rated SafeHaven at 6.13 and ChatGPT at 5.75. Despite these consistent preferences of SafeHaven, the differences weren't statistically significant, with  $p$ -values above the significance of 0.05 across all age groups.

Table 5.5: Descriptive statistics and statistical comparisons of user perceptions of information quality between SafeHaven and ChatGPT

Item	SafeHaven (Mean, (SD))	ChatGPT (Mean, (SD))	Z	P-value (one- tailed)
Information is easily accessible	6.39 (0.698)	5.83 (1.383)	-1.502	0.067
Information is of sufficient volume for my needs	<b>6.17 (0.985)</b>	5.44 (1.723)	-1.767	0.039*
Information is believable	6.50 (0.618)	6.28 (0.669)	-1.265	0.103

Information is sufficiently complete for my needs	6.22 (0.732)	5.56 (1.617)	-1.439	0.076
Information is presented concisely	<b>6.39 (0.698)</b>	5.56 (1.381)	-2.388	0.009**
Information is represented in a consistent format	6.61 (0.502)	6.33 (0.686)	-1.518	0.065
Information is easy to manipulate to meet my needs	5.33 (1.609)	4.94 (1.434)	-1.462	0.073
Information is correct	6.39 (0.502)	6.06 (0.802)	-1.387	0.083
It is easy to interpret what information means	6.44 (0.616)	6.00 (0.907)	-1.576	0.058
Information is useful to the task at hand	6.44 (0.984)	6.06 (0.802)	-1.530	0.063
Information is sufficiently timely	6.33 (0.767)	6.56 (0.616)	-0.832	0.203
Information is easy to understand	<b>6.50 (0.618)</b>	6.17 (0.514)	-2.121	0.017*
Overall (average)	<b>6.31 (0.431)</b>	5.8981 (0.631)	-2.287	0.011

Note: \*P < 0.05, \*\*P < 0.01, indicating statistical significance.

#### 5.2.2.4 Usability

Using paired t-tests, we compared the SUS scores of SafeHaven and ChatGPT. The analysis results show that the SUS score was higher for SafeHaven ( $M = 83.194$ ,  $SD = 12.7708$ ) than ChatGPT ( $M = 79.306$ ,  $SD = 13.58$ ) (Figure 5.9), yet the difference was insignificant ( $p = 0.261$ ). Given that the SUS scores for both SafeHaven and ChatGPT were well above the usability benchmark of 68, indicating that participants found both tools usable.

We looked at how gender and age moderate the effects. Having examined the gender moderation effect, we found that females preferred SafeHaven ( $M = 81.33$ ) to ChatGPT ( $M = 79.00$ ), whereas males were more likely to prefer SafeHaven ( $M = 92.50$ ) over ChatGPT ( $M =$

80.83). Nevertheless, there were no notable variances in the usability ratings based on gender, as indicated by the p-values surpassing the significance threshold of 0.05 ( $p = 0.522$  for females,  $p = 0.349$  for males). In terms of age, perceived usability scores varied between SafeHaven and ChatGPT, but no statistically significant differences were discovered. For clarification, those aged 18-25 rated SafeHaven with an average of 83.12 versus 79.68 for ChatGPT ( $p = 0.562$ ). Similarly, those who are 26-33 rated SafeHaven with an average of 90 compared to 85 for ChatGPT ( $p = 0.326$ ). Interestingly, the age group 34-41 viewed ChatGPT to be more usable, presenting scores of 63.75 for SafeHaven and 65 for ChatGPT, with no significant difference ( $p = 0.967$ ).

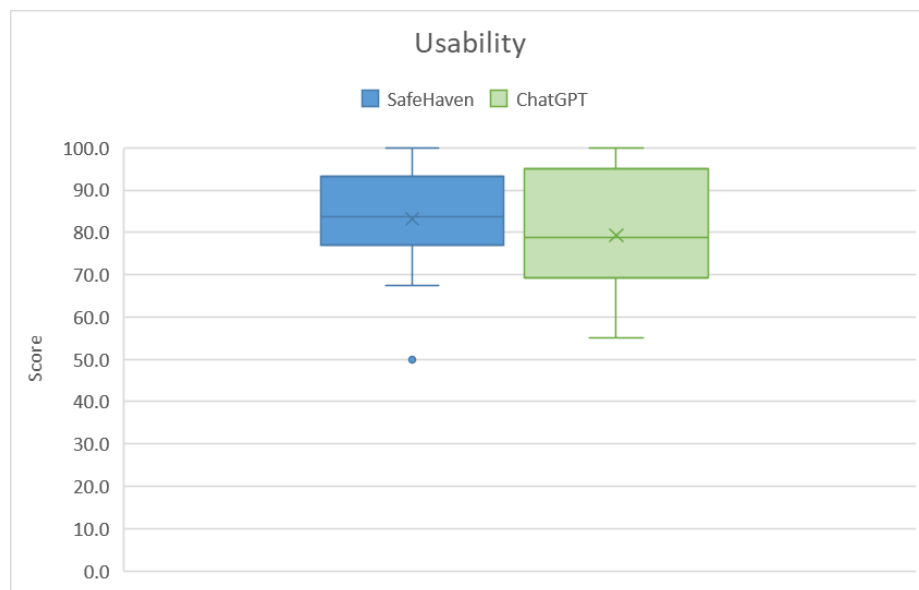


Figure 5.9: Boxplot of usability of SafeHaven and ChatGPT

#### 5.2.2.5 Perceived Anonymity

The descriptive statistics of perceived anonymity are reported in Figure 5.10, showing a slight difference between ChatGPT ( $M = 6.144$ ,  $SD = 0.7090$ ) and SafeHaven ( $M = 6.111$ ,  $SD = 1.1866$ ). However, the Wilcoxon Signed-Rank Test results did not yield any significant difference ( $p\text{-value} = 0.345$ ). Specifically, eight participants felt that SafeHaven offered more

anonymity (a mean rank of 8.38), seven favored ChatGPT (a mean rank of 7.57), and three showed no preference. As a result, H8 is not supported.

We investigated the effect of gender and age. There was gender-based difference in the perceived anonymity of SafeHaven versus ChatGPT, despite the fact that these differences were not statistically significant. SafeHaven was rated more anonymous by male participants ( $M = 6.60$ ) than ChatGPT ( $M = 5.80$ ), but no significance was noted ( $p = 0.180$ ), as shown in the Wilcoxon Signed Ranks Test. However, female participants perceived ChatGPT as slightly more anonymous ( $M = 6.21$ ) than SafeHaven ( $M = 6.01$ ), but no significant difference was found between the two mediums ( $p = 0.861$ ). In terms of age, participants' ratings of SafeHaven CA were not statistically significant compared to those of ChatGPT. We found that participants between the ages of 18-25 and 26-33 reported SafeHaven exhibited higher levels of anonymity when compared to ChatGPT, receiving average scores of 6.40 and 6.28 compared to 6.17 and 5.88, respectively. Nonetheless, the p-values suggest a lack of statistical significance. In contrast, participants in the age groups 34-41 and 42-49 rated ChatGPT as more anonymous, with mean scores of 6.50 and 6.00 against SafeHaven 4.90 and 5.50 respectively. Again, the p-value did not indicate a significant difference.



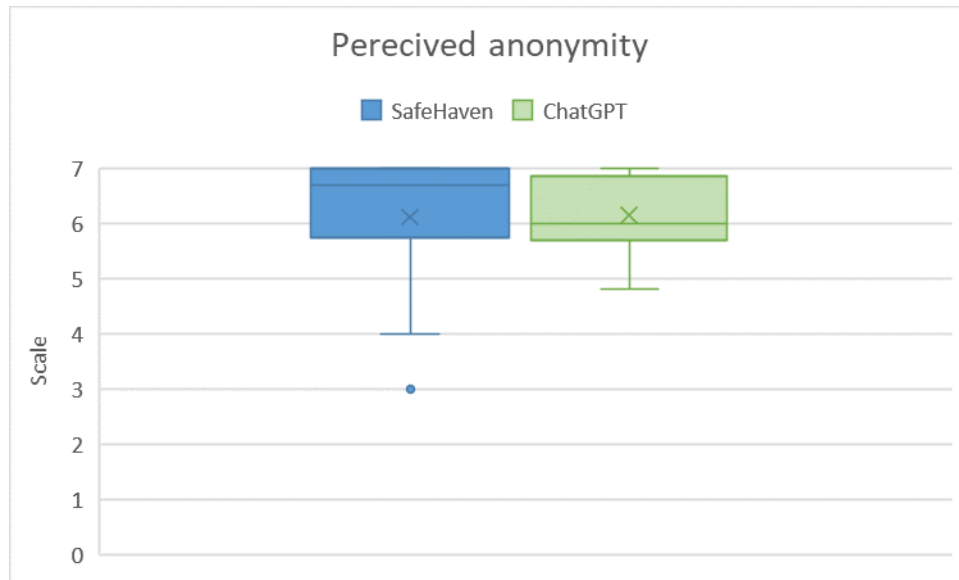


Figure 5.10: Boxplots of perceived anonymity of SafeHaven and ChatGPT

#### 5.2.2.6 Perceived Trustworthiness

The descriptive statistics of perceived trust are reported in Figure 5.11. The results of the Wilcoxon Signed-Rank Test show that the perceived trust is higher ( $p = 0.007$ ) for SafeHaven ( $M = 6.3331$ ,  $SD = 0.5346$ ) than ChatGPT ( $M = 6.0316$ ,  $SD = 0.6907$ ). This is evidenced by 13 respondents rating SafeHaven more favorably, leading to an average rank of 9.85 and an aggregate rank of 128.00 for SafeHaven. On the other hand, ChatGPT was preferred by four respondents, with an average rank of 6.25 and a cumulative rank of 25.00, and one respondent remained neutral. Based on these findings, we conclude that CA-based intervention of DV can be viewed as more trustworthy than ChatGPT. Thus, H10 is supported.

We studied the moderation effect of age and gender. For gender, we noted that females rated SafeHaven as more trustworthy ( $M = 6.32$ ) than ChatGPT ( $M = 6.09$ ), and this difference reached statistical significance ( $p = 0.047$ ). Males also evaluated SafeHaven as more trustworthy ( $M = 6.36$ ) than ChatGPT ( $M = 5.72$ ), but the difference wasn't statistically significant ( $p = 0.285$ ). As for age, a comparison of trust ratings for SafeHaven and ChatGPT across different

age groups reveals a general preference for the CA's trustworthiness, even though no statistically significant differences have been observed. SafeHaven was rated higher by participants aged 18-25 ( $M = 6.3229$ ) than ChatGPT ( $M = 6.1242$ ), with a p-value of 0.090, closest to statistical significance. Further, SafeHaven scored higher than ChatGPT in the 26-33 and 34-41 age groups, with mean scores of 6.43 and 6.37, compared to 6.10 and 5.33, respectively, but with p-values of 0.216 and 0.180, showing no significant differences. In the 42-49 age group, scores were nearly identical on both mediums ( $p = 0.655$ ).

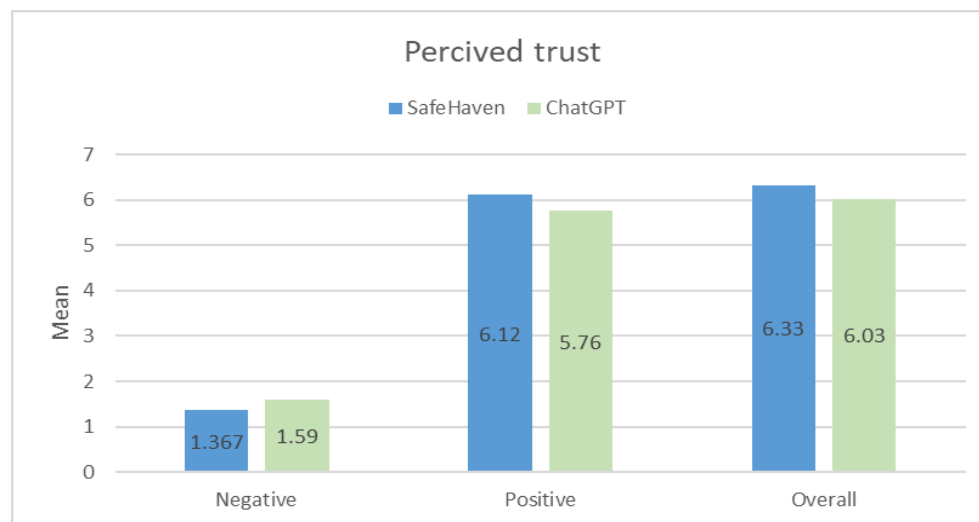


Figure 5.11: Bar charts of perceived trust of SafeHaven and ChatGPT

### 5.3 Qualitative Analysis Results

#### 5.3.1 Cognitive Walkthrough Results

The following section summarizes the cognitive walkthrough of each of the six tasks conducted by participants. Each task was structured around the action(s), and each action was evaluated based upon the expectation that participants would achieve the desired outcome, the recognition of the required action, the connection between the description of the action to what was being attempted, the observation of progress toward solving the task, and whether or not the

action was ultimately successful [344]. As part of the analysis of cognitive walkthrough results, a qualitative analysis is conducted to synthesize participants' interactions, expectations, and feedback into coherent themes that identify areas for improvement and highlight strengths [237]. Key steps include the compilation of data, the assessment of action success rates, the identification of patterns in participants' behavior, and the analysis of participants' notes for qualitative insights. We provide a detailed analysis of specific usability issues followed by a set of practical recommendations.

According to the cognitive walkthrough analysis results, participants' evaluations of SafeHaven designed for DV interventions revealed several key insights. Participants, primarily composed of DV survivors, family members of survivors, or counselors, provided feedback that highlighted SafeHaven's informative and empathetic, non-judgmental conversation style. However, they also highlighted areas needing improvement, specifically the ability to handle complex queries and address navigational issues. Table 5.6 summarizes the challenges identified across all tasks, as well as the corresponding suggestions for improvement.

An interesting observation from the cognitive walkthrough relates to the variety of navigation methods utilized by participants. Among participants, eight exclusively used buttons to navigate between different modules of SafeHaven, 17 participants navigated through text input, and the remaining 11 participants used both text input and buttons when navigating SafeHaven. Diverse navigation preferences draw attention to the need to provide multiple methods of interaction to accommodate a wide range of user preferences and needs, ensuring that the CA is both user-friendly and accessible.

Throughout this section, we will discuss key points for each of the actions within each task.

Table 5.6: Challenges, Tasks, and Recommendations for Improving User Experience with SafeHaven

Challenges	Task	Suggestion(s)
Inability to navigate to the correct module or difficulty starting	Task 1, Task 3, Task 4	<ul style="list-style-type: none"> <li>• Ensure that clear instructions are provided at the beginning of the conversation</li> <li>• Enhance the intuitiveness of the interface</li> </ul>
Unclear presentation of information	Task 1, Task 3	<ul style="list-style-type: none"> <li>• Improve question phrasing and information clarity</li> </ul>
Inadequate handling of unique or unexpected user prompts	Task 1, Task2, Task 3, Task 4	<ul style="list-style-type: none"> <li>• Training the CA to include a wider range of keywords, terms and responses.</li> <li>• Develop an algorithm for understanding semantics of user responses.</li> <li>• Implement a feedback loop that learns from every interaction.</li> </ul>
Generalized or non-specific tips or information provided by the CA	Task 3, Task 4	<ul style="list-style-type: none"> <li>• Provide more specific, actionable instructions or information that is tailored to the user's situation</li> <li>• Developing more specialized modules to address users' specific needs.</li> </ul>
Data input issues (multiple or incomplete inputs)	Task 2	<ul style="list-style-type: none"> <li>• Clarify data input instructions</li> <li>• Enhance data handling capabilities</li> </ul>
Information on external resources is not comprehensive	Task 2	<ul style="list-style-type: none"> <li>• Embedded features or an internal sophisticated database that contain all related information</li> </ul>
The overuse of emojis may affect the seriousness of the content	Task 5	<ul style="list-style-type: none"> <li>• Maintain a balance between emotional tone and emoji usage</li> </ul>
Limited emotional support capabilities of the CA	Task 6	<ul style="list-style-type: none"> <li>• Enhance language sensitivity and provide more encouragement and validation</li> </ul>

#### 5.3.1.1 Task 1: Develop a DV Safety Plan

The analysis of the cognitive walkthrough provided deep insight into the participants' perceptions of their interactions with SafeHaven when performing Task 1. In action 1 (i.e., identify where to click/write to start the plan), 86% of participants anticipated achieving the desired result, while a few did not (P1, P2, P5, P12, P30). For example, we observed that P1 navigated through several modules until she reached the safety plan module. Further, we noted that one participant faced an initial challenge of starting the plan when she entered (e.g., How do I leave my boyfriend safely?), and the CA did not recognize her input, perhaps due to the out-of-scope issue (P12). After a few attempts, the participant was able to navigate to the safety plan module successfully. Similarly, when P14 entered an unrelated phrase (e.g., Risk management),

and the participant was directed to a different module (i.e., Risk assessment). P30 also found it difficult to start the task, as shown in the action notes, *"It was a little difficult to find where to start and click on it."* This may be attributed to a lack of clear, guiding instructions or prompts on SafeHaven's home page. It is also possible that individual differences in technological proficiency or familiarity with the CA may play a role in this challenge (P30). Almost all participants connected the description of the desired action to what was being attempted, and they observed progress toward accomplishing the task. Overall, most of the participants demonstrated a thorough understanding of where the safety plan should begin, with a success rate of 97%.

Action 2 (clicking the designated button or typing a text to initiate the plan) demonstrated exemplary performance, with all participants successfully completing it (success rate: 100%). In addition to the intuitive navigability of the system, participants were able to correlate the action description with the task easily.

In the same vein, while most participants could answer the questions successfully (success rate: 94%) in action 3 (i.e., answer questions about the plan), there was a minority of participants who were not satisfied with the outcome of the presentation of the questions and answering and did not find it easy to proceed. Some participants noted that the presentation of the information or questions was unclear (Figure 5.12). The unclear presentation noted by participants was due both to the issues with question phrasing and to the inclusion of irrelevant or non-applicable questions. For instance, P13 mentioned in the action notes, *"Difficult to answer some questions if I had never thought of them before like who could pick you up from work? I said nobody, everyone else is working too, and that phrase was on the safety plan."* P13 added in the

interview, *"With the chatbot, when asked if you're alone, who would you stay with? That question was very confusing to me."*

During the review plan action (action 4), a high rate of action success was observed (97%), as participants were able to review the summarized safety plan effectively. The survivor's answers are included in the plan template, which facilitates easy reading and comprehension. As an example, P33 wrote, *"Easy to find action items."* Nonetheless, there were instances that highlighted critical limitations in SafeHaven system's handling of user input. A particular observation from P5, P11, P13, P27, P28, and P31 has brought to light the concern that when a survivor enters unique or unexpected responses, such as "nobody," "no one," or "unknown" to questions in the context of a safety plan, these responses are immediately included in the summary without providing contextual interpretation or validation (Figure 5.13). For instance, P28 explained in the action note, *"If there was 'nobody' as a response option, it did not follow up. It just added 'nobody' as an option in the safety plan. BIG issue."* Similarly, in the case of P9, an issue was raised when a survivor mentioned 'not being employed' in her plan; the CA did not provide adequate interactive options to address this response. P9 stated, *"Chatbot did not have interactive options for not being employed."* As another example, during interactions with SafeHaven, specifically after reviewing the safety plan, it was noted that the CA did not recognize certain positive acknowledgments, such as "Thank you," from P10. This could potentially lead to confusion in the course of communication.

In response to these issues, a couple of suggestions were provided to improve the flow, presentation, and interpretation of information in Task 1.

- Improving readability: P19 recommended consolidating sentences within one single paragraph rather than scattering them across separate texts. This improves readability

and minimizes potential interruptions from reading the preceding text as the screen scrolls upward with each new message.

- **Enhancing contextual interpretation:** To address the contextual interpretation issue, P11 suggested that if a survivor does not have any resource, it would be helpful to provide him/her with a pop-up or a suggestion link instead of having "no one" or "unknown" responses in the plan. This issue could also be addressed by developing an algorithm capable of understanding the semantics of user responses, ensuring that the final summary plan remains coherent and relevant to the user's safety plan intent [347].
- **Incorporating advanced interactive elements:** It is suggested that SafeHaven may introduce advanced interactive elements where survivors can review and modify their inputs if the CA flags them as potentially problematic or out of context.
- **More guidance on task initiation:** Enhancing the guidance features in this task, particularly clarifying where and how a survivor can initiate the plan, is essential (P12, P14, P30). To ensure a smooth and clear initiation of the task for all survivors, a more prominent cue or introductory guide will be necessary.
- **Improving information presentation:** Adaptive questioning techniques can be used by the CA to address the issue of unclear presentation of information and irrelevant questions (P13). Also, providing an option for users to skip questions or indicate "not applicable" can help avoid confusion and ensure the conversation flow is relevant.
- **Enhancing the CA's capabilities:** The use of advanced Natural Language Understanding (NLU) techniques can enable SafeHaven to recognize and process diverse user inputs, including affirmative responses (P10) and complex queries (P12),

thereby improving its capabilities. Expanding the CA's training data to include a wider range of responses and employing robust fallback mechanisms can further refine the CA's responsiveness and interaction quality [348], [349].



Figure 5.12: Question formatting issue (Highlighted in red rectangle)



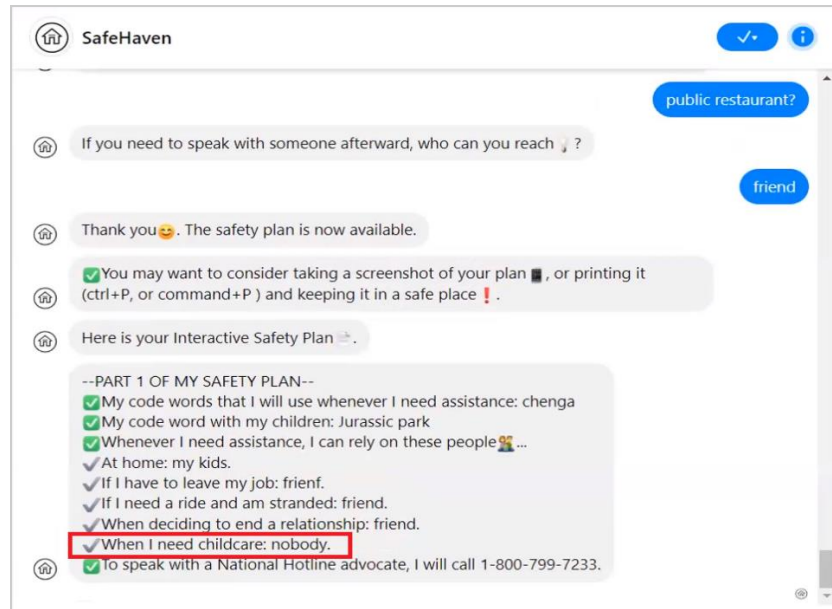


Figure 5.13: Responses display issue (Highlighted in red rectangle)

### 5.3.1.2 Task 2: Find a DV Shelter

The cognitive walkthrough analysis offered valuable insights on how participants perceived their engagement with SafeHaven while completing Task 2. In the first action, participants were tasked with identifying where to click or write to find a DV shelter. Participants acknowledged the availability of the desired action and monitored progress toward accomplishing their task by associating the action descriptions with their own objectives throughout the process. Despite some initial uncertainties, all participants were successful in identifying where to click or write to find a shelter (100%).

In action 2, the focus was shifted to the actual engagement with SafeHaven to initiate the shelter-finding process. According to the responses, participants have mixed expectations regarding the outcome, with some expecting a different outcome than what they had anticipated when they clicked or wrote to locate a shelter. Despite this, all participants were aware of the required action, correlated it to their task, and observed progress, which ultimately resulted in the

action being regarded as successful (100%). Thus, while there may be initial hesitation or skepticism about the effectiveness of the action, SafeHaven system's design, along with feedback mechanisms, are adequate for users toward successful outcomes.

During Action 3, participants were asked to provide their location, an important element of providing personalized shelter recommendations. Despite mixed expectations of participants, the data indicates that almost all of them reported success regarding this action (100%). For instance, according to P5, the process was straightforward, indicating that sharing location was not a barrier, and SafeHaven assisted well with the process. Meanwhile, some other participants (e.g., P1, P16, and P27) indicated that they did not anticipate achieving the desired outcome, indicating potential concerns or uncertainties about this step, potentially related to the clarity of instructions on sharing locations.

In the last action or the final step of the task, participants were tasked with viewing a list of nearby shelters. While most participants expected to attain the desired outcomes, few did not have high expectations for success (P1, P16, P22, and P27). Further, it was evident that all participants recognized the necessary steps, perceived a sense of progress, and ultimately marked this action as successful (100%). On the other hand, the participants also noted some issues. A specific weakness was SafeHaven's inability to handle multiple or incomplete inputs, which yielded inaccurate or no results (P8, P13). For example, P8 noted that SafeHaven was unable to handle two ZIP codes in the user input, indicating possible confusion or lack of clarity regarding multiple inputs, even though this issue can be resolved through the linked external website (Figure 5.14). P8 reported in the action notes, *"I put in two zip codes, but it provided answers for one zip code, which made sense to me. I saw that I could go in and easily fix the zip code issue and re-search."* P13 also mentioned a similar point in which missing city and state information

may result in inaccurate results (Figure 5.15). This indicates that merely mentioning a city can lead to inaccurate results because it may lack the specific context needed for the accurate retrieval of geographic information. A second weakness reported by P19 concerned the linked external website to DV shelters. It was noted that the website provided by SafeHaven did not contain shelter addresses despite showing a list of shelters, indicating that the information provided may not be sufficiently rich. P19 mentioned, *“This is external, but the provided page doesn't show the address of the shelters.”* In addition, one participant expressed concern regarding the navigation process within SafeHaven. Specifically, P22 stated that he expected to be able to conduct shelter searches directly within the CA's interface rather than having to navigate to an external website. A technical glitch was reported by P28, highlighting a disruption in conversational flow. When a button is pressed during a dialogue, SafeHaven may trigger an unintended action that disrupts the conversation and results in overlapping responses from various SafeHaven modules. P28 wrote in the action notes, *“interruptive option (another button in the middle of one action) is not well supported.”*

Based on participant feedback, several potential improvements should be considered to enhance SafeHaven's functionality and user experience.

- **Accurate input fields:** P13 suggested including a State field in addition to a city name to enrich location details in user prompts. This would enable the implementation of a more precise location-based search function.
- **Integrated user experience:** P22's expectation for the shelter location to be displayed within SafeHaven interface rather than being redirected to an external website indicates that the user desires an integrated experience. This need may be addressed by enhancing SafeHaven's interface to provide comprehensive information, perhaps

through an embedded feature or an internal sophisticated database that contains all shelter information.

- Clear instructions for ZIP code input: To address the issue of multiple ZIP codes, as reported in P8's feedback, SafeHaven should provide clearer instructions on inputting a single ZIP code. This approach will help minimize user confusion and enable the CA to offer precise assistance.

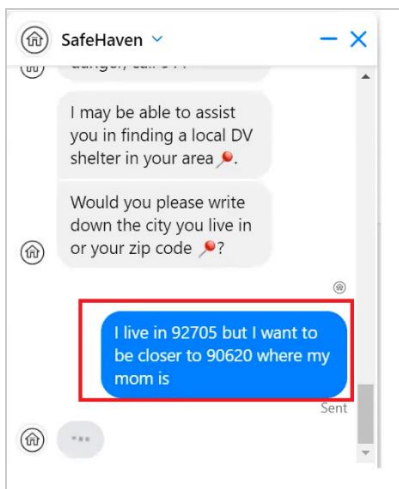


Figure 5.14: Multiple zip code inputs issue (Highlighted in red rectangle)

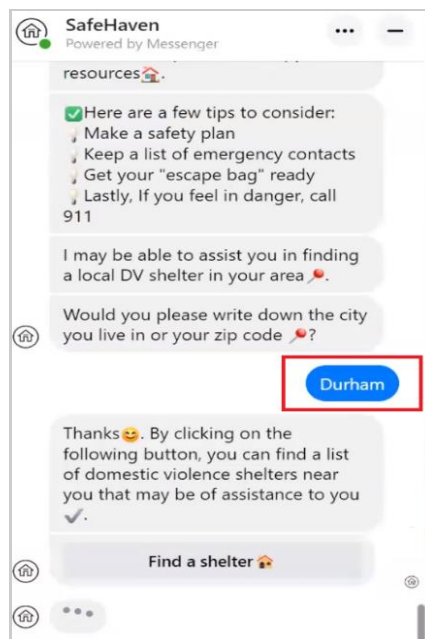


Figure 5.15: Missing city and state information issue (Highlighted in red rectangle)

### 5.3.1.3 Task 3: Gather Evidence

As part of this task, participants were expected to learn how to collect evidence using SafeHaven. In action 1, participants had to know what to click or write to start the task. It was evident that most participants were anticipating the desired outcome and were aware of the available actions, suggesting that the interface was generally intuitive and in accordance with their expectations. In addition, participants observed progress toward completing the action, which contributed to a high success rate (92%). Nevertheless, some participants encountered some navigational difficulties and experienced uncertainty about how to begin (P4, P18, P13, P23, P25, P35). As per P25, it was challenging to navigate to evidence resources, highlighting the need for a more intuitive user interface. In addition, it is noteworthy that despite P4 entering inputs multiple times, SafeHaven failed to recognize them until the word 'evidence' was included into the input. Further, P23 reported that understanding rights did not precisely align with gathering evidence, suggesting a potential mismatch between user expectations and the information available in the menu. We also observed that two participants (P13, P18) found it challenging to navigate to the designated module, 'Know my rights,' and inadvertently selected other modules before finding the correct one (Figure 5.16).

Action 2, in which participants interacted more directly with SafeHaven through clicking buttons or writing, showcased a design that achieved a high level of success (97%), recognition of available action, and observation of progress in achieving the desired outcomes. However, some participants pointed out some issues with this action. For instance, P13 discussed inaccuracies in sharing location options for legal services and irrelevant suggestions, highlighting the importance of providing accurate and contextually relevant information. Further, P19's

confusion with the provided options further indicates that while SafeHaven might be guiding users effectively, the clarity and relevance of those options can be improved to ensure they are intuitive and directly applicable to the tasks. P19 mentioned, *“I’m getting a bit confused with which options are part of the main menu and which are inside secondary menus.”*

In the final action, participants viewed tips on how to gather evidence. Although the majority of participants considered the tips helpful and believed they would help achieve the desired outcomes, three expressed uncertainty or skepticism, indicating a possible communication or perception gap (P1, P6, P22). Further, all participants recognized that the tips were available and intended to assist them in gathering evidence. All participants except for two were able to understand and relate the description of viewing tips to their objective of gathering evidence. All participants also felt that they made progress toward completing the task if the desired action was performed, and they were able to complete their intended actions within SafeHaven successfully (100%).

The participants reported numerous positive comments. For instance, P33 valued the approach of inquiring about the need for further information instead of providing an exhaustive list all at once. In addition, P34 shared a positive experience about how to perform the task, *“Very helpful. It helps me know how to successfully build evidence and a case against my abuser and to prove my defense in court.”* On the other hand, the participants also reported several issues. For example, P11, P22, and P31 noted in action notes that the tips for collecting evidence were too general, suggesting the need for more specific and actionable instructions. P22 wrote, *“The information was more general for the chatbot than the search engine. For example, the search engine told me to screenshot missed calls or threatening text messages. However, the chatbot was more general and did not really tell me what to look for specifically.”* P27 and P31 shared a

similar observation about the inclusion of specific tips within the context rather than general tips. (Figure 5.17). P31 explained in the action notes, “*Giving specific tips on how to find out the type of evidence that is wanting to be collected would be useful.*” Furthermore, P5 indicated that survivors may expect the outcome to be the provision of legal agencies rather than solely tips. P17 also noted that while interacting with SafeHaven and learning about tips for gathering evidence, it was unable to recognize some affirmative responses like “*Yes*”, “*Sure*”, and “*OK*”.

Throughout all actions, participants' notes and experiences reveal a consistent theme of seeking more specific, useful, and contextually relevant information.

- **Tailored and specialized information:** It has been suggested that SafeHaven should provide more specific and tailored information and tips, thereby empowering survivors and providing clear guidance based on their specific needs (P4, P11, P22, P27, P31). This can be accomplished by creating more specialized modules for SafeHaven, enabling it to address survivors’ queries based on the specific types or details of evidence they seek to understand. This would enable SafeHaven to utilize a more targeted set of information for its responses. Moreover, advanced machine learning techniques can be utilized in the CA to implement a feedback loop that learns from every interaction, enhancing its capability to provide relevant and useful information over time.
- **Enhancing clarity and navigation:** There has been a request for a clear distinction between the main menu items and secondary menu items in the SafeHaven interface to enhance the navigability of the CA (P19). This can be addressed by utilizing clear and intuitive icons alongside descriptive labels for all menu items and maintaining a consistent layout throughout the interface that places the main menu options in the

same location, such as the top navigation bar and the secondary items in dropdown menus or side navigation panels. Furthermore, it was recommended that the terms used in the SafeHaven system (e.g., Know my rights) be clarified to accurately reflect the actions survivors are expected to take (P23). Therefore, it is crucial to review and revise the language used by the CA to ensure that all terms and definitions are clear and precise, use plain language that is easily understandable, and avoid technical terms that might confuse survivors.

- **Enhancing CA's recognition:** Based on P4's experience, expanding SafeHaven's training dataset by incorporating a broader range of keywords and phrases related to gathering evidence could enhance its ability to recognize diverse inputs. Intent recognition algorithms can also be refined to help the CA better understand the context and relevance of user inputs. Similarly, regarding the recognition of affirmative responses (P17), enhancing SafeHaven could involve improving its natural language processing capabilities and training it to identify a broader range of affirmative responses. Providing it with a larger dataset of affirmative phrases and continuously improving its machine-learning algorithms will enhance its ability to interpret user responses more accurately.
- **Improving actionable legal guidance:** It is expected that direct links to agencies specializing in legal assistance for gathering evidence will be provided (P5). The latter is crucial to obtaining accurate information and admissible evidence, strengthening a survivor's case in court.



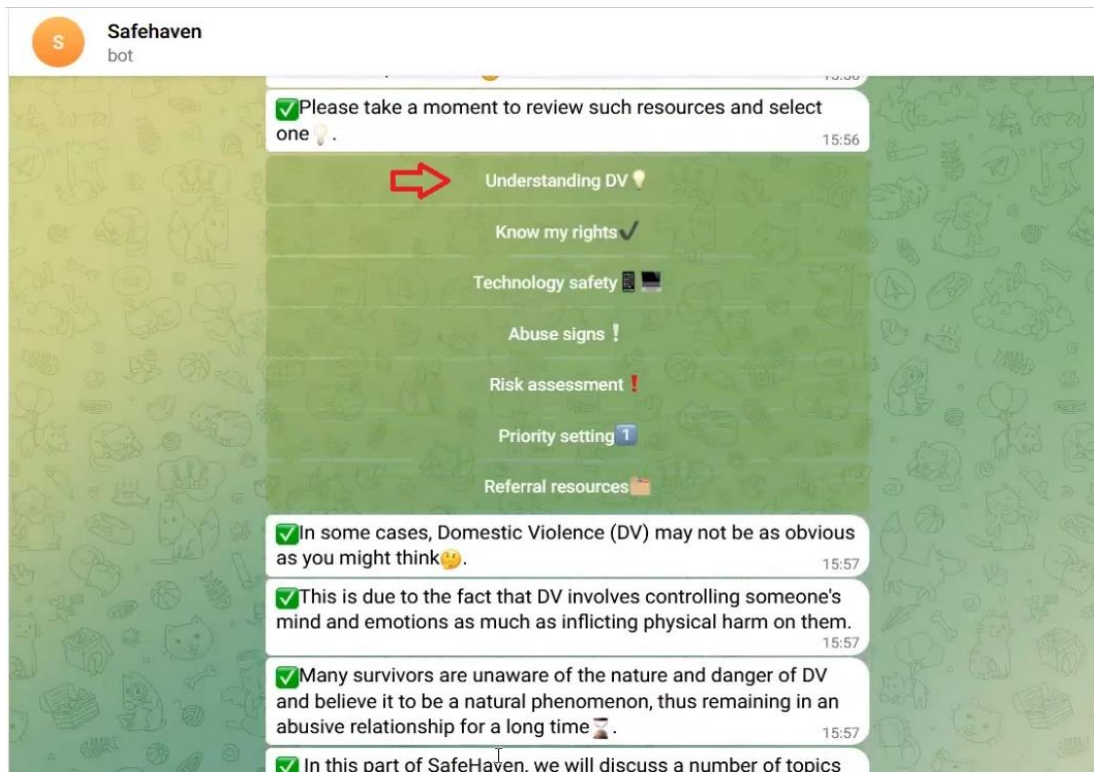


Figure 5.16: Navigation issue (Highlighted in red arrow)

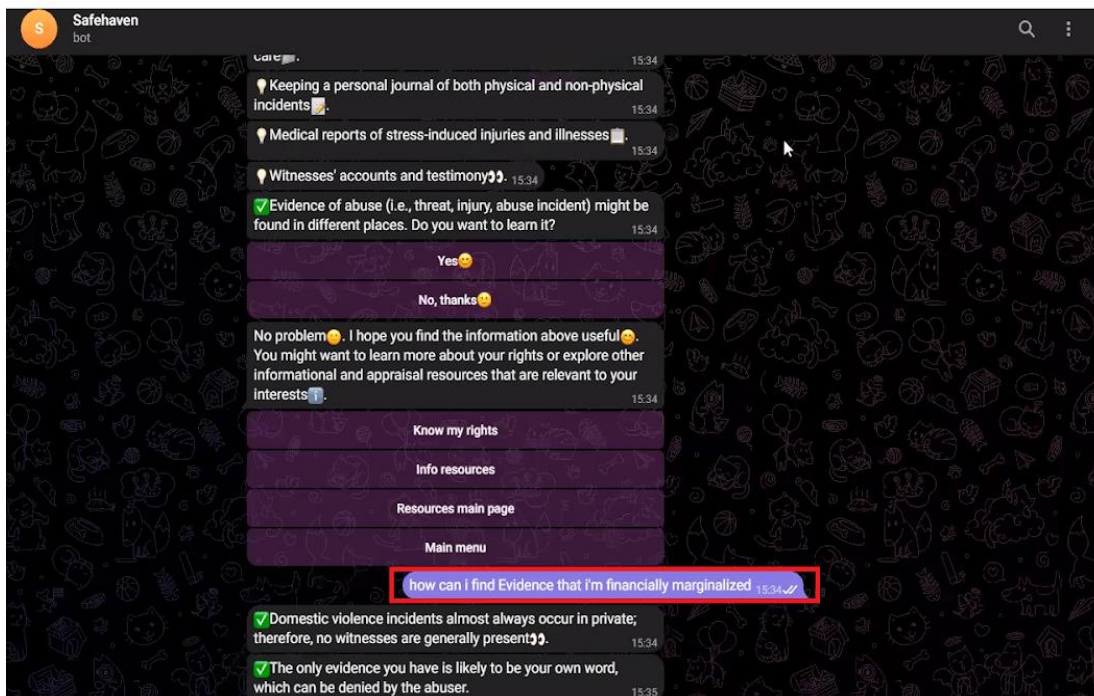


Figure 5.17: Inefficient responsiveness to specialized tips (Highlighted in red rectangle)

#### 5.3.1.4 Task 4: Learn How to Protect Survivors from Cyberstalking

In this task, participants were asked to learn how to protect survivors from cyberstalking. A cognitive walkthrough reveals several noteworthy findings. In action 1, where participants were asked to identify where to click or write to start the task, they had mixed expectations about achieving the desired outcome, suggesting issues with navigation or a partial understanding of the interface features and functionality. Nevertheless, most participants acknowledged the availability of the desired action, and all participants were able to observe the progress of the desired action as a result of their actions. This was confirmed by the high success rate of this action (94%). Despite this, one participant mentioned confusion about two menus (main and resources main), which suggests that the initial step of identifying where to click or write to learn about cyberstalking protection might pose a challenge, as well as identifying the information contained within each menu (P19).

During action 2, participants were asked to click a designated button or write a text to learn about protection. They exhibited a more positive attitude about their abilities and generally expected to succeed in the task, indicating that confidence may have been fostered by the interface design or previous experiences. The participants' recognition of the available action, along with notable progress and success rates (92%), indicate that clicking the designated button and writing a text were both successful ways to navigate to the cyberstalking protection module. Only on a few occasions, some participants, like P1 and P34, either clicked on a button redirecting them to a different module or encountered issues with the CA not recognizing their inputs accurately, such as 'cyberbullying' or 'digital harassment.'

In action 3, where participants viewed tips about cyberstalking protection, mixed expectations were once again noted, as in action 1, which indicates that participants viewed the tips from

different perspectives. Participants recognized the action, connected it to the desired outcome, and observed progress, indicating clear guidance and feedback within the interface. In addition, the action had a high success rate (94%). Both positive and negative messages were noted in the action notes. Some participants appreciated the CA's tips regarding cyberstalking protection (P13, P33, P34, P36). For instance, P13 commented, *"This one felt the most straightforward and easily accessible of the tasks."* P36 wrote in the action notes, *"I liked how they gave additional information to websites that can help protect you against cyberstalking via mobile phone and resources on how to speak with a professional about this topic."* However, some participants encountered some limitations, and their expectations were not met when viewing the cyberstalking protection tips (P5, P8, P11, P22). For example, P8 expressed concern about SafeHaven's lack of detail and repetition. When the participant asked about securing their password, it did not respond appropriately (Figure 5.18). In a similar vein, P27 inquired about antivirus software, but SafeHaven was unable to respond. This may suggest an issue in delivering detailed and adequate tips on cyberstalking protection and inefficient responsiveness in handling specialized tips. Furthermore, it was noted that some tips about cyberstalking protection are general and not applicable if the abuser owns the device (P11). One participant (P22) experienced a mismatch between expectations and outcomes when he expected specific information on preventing cyberstalking but received general Internet privacy advice instead. P22 noted, *"The chatbot gave me a different answer than what I expected. I asked the chatbot on how to prevent being cyberstalked. However, the chatbot gave me a response about general Internet privacy instead. This information did not answer my question."* We noted that when the participant inquired about preventing cyberstalking specifically on social media platforms, he was provided with identical tips that applied to averting cyberstalking on mobile devices. Such a

lack of tailor-made tips may lead to confusion and diminish the effectiveness of the guidance provided.

Participants' feedback points out the opportunities for improvement in SafeHaven's design and functionality.

- Enhanced detail and variety in responses: As a means of enriching the learning experience and maintaining user engagement, more detailed and diverse responses should be provided to SafeHaven while avoiding redundant responses (P8, P27). SafeHaven, for example, may respond to survivors' inquiries about a specific topic by providing additional information, relevant examples, or even links to external resources. This approach would enhance the CA's informational value and make the user interaction dynamic and insightful.
- Improve resource connectivity: P5's suggestion was to facilitate access to external resources, educational materials, or lectures on cyberstalking protection. SafeHaven can be paired with additional resources to extend the scope of learning and support.
- Personalization of tips: The note from P11 about considering ownership of technology accounts (e.g., in DV situations) underscores the necessity of personalizing tips to fit the survivor's specific circumstances. By determining account ownership at the beginning of the CA's session, its tips can be tailored to better suit the survivor's ability to implement them.
- Interface clarity and intuitiveness: P19's confusion regarding the navigation menus indicates the need for a more intuitive interface design. Simplifying the menu structure and clearly indicating where different types of technology safety information can be found will enhance navigation for users.

- Accuracy in response to queries: Based on our observation and in efforts to improve SafeHaven's utility, it is imperative to refine its ability to recognize and respond to diverse user inquiries (P1, P22, P34). SafeHaven should be trained with an extensive and evolving related set of keywords and terms to discern and accurately respond to varied user inquiries and link them to the appropriate module. Additionally, it would be advantageous to develop a comprehensive guide and various modules, coupled with connecting the CA to these modules with programming, which concentrates on cyberstalking prevention techniques specific to different technologies like social media, mobile devices, and the Internet.

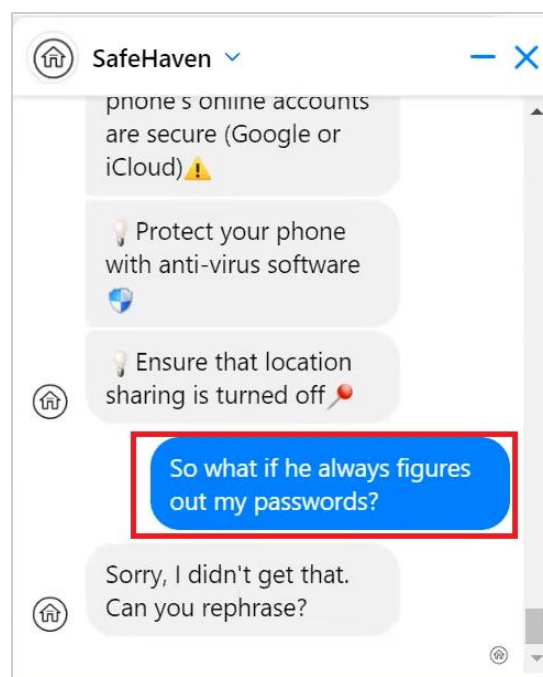


Figure 5.18: Inefficient handling of specialized tips (Highlighted in red rectangle)

#### 5.3.1.5 Task 5: Read a Survivor Story

In this task, participants were asked to find a DV survivor story and read it to gain insight into the challenges they face and their methods of coping. To evaluate the usability and effectiveness

of SafeHaven's functionality in this task, a cognitive walkthrough was conducted to assess the main actions taken by participants. The following analysis delves into action 1, which involves identifying where to click or write to read a survivor story. Based on our analysis, 91.67% of participants anticipated achieving the desired outcome, and a slightly higher percentage believed the action was available. Further, it was observed by most participants that the description of the desired action correlated with the action being attempted and that progress was being made toward completing the action. Overall, 97% of participants rated this action as successful. We can infer from the data that the interface and conversation flow were generally clear, enabling participants to identify the necessary steps to read a survivor's story. However, a few instances of failure were observed (noticed by P5 and P6), indicating that, despite its effectiveness, the CA may not be intuitive or accessible to all users.

In action 2, participants were assessed based on their ability to click the designated button or write text to start the task. Data shows that the action had a perfect success rate of 100%. In addition, approximately 91.67% of them expected to achieve the desired outcome. All participants agreed to recognize the action, connect the action description to what was being attempted, and monitor progress toward completing it. This may indicate a well-designed user interface regarding action visibility and feedback.

The third action in the CA cognitive walkthrough is to read the story of a survivor. The data shows that 94.44% of participants expected that their desired outcome would be achieved, and 100% of participants were aware that the action was available, connected the description of the action to what they were attempting, and observed progress toward solving the task. Analysis of the participants' action notes indicates a mix of expectations and experiences with this task. Specifically, P13, P34, and P36 reported positive experiences. For example, P34 noted the

importance of reading survivors' stories to provide emotional support. P34 commented, “*Other domestic violence stories helped me feel like I wasn’t alone.*” As a critical component of a tool aimed to provide support and information to a wide audience, P36 valued the inclusion of stories from diverse male and female perspectives. Further, P13 emphasized the clarity and conciseness of the task. Nevertheless, some participants raised some specific concerns. As an example, P5 was expecting longer DV stories, indicating a potential gap between what was expected and what was delivered. P5 wrote, “*The survivor story was not as expected. I thought it would've been a longer story. It was just like a brief description. I thought it would be linked to another site with many stories to read.*” Further, P11's experience suggests an issue with the natural language processing capabilities of the CA. The CA failed to correctly interpret the phrase "coping with domestic violence" when used by the participant, as it directed the participant to another module. However, it was able to correctly direct the participant to the appropriate content when the participant provided input that contained a "survivor story." In addition, P19's comment regarding the overuse of emojis suggests that while these elements can add a human touch to a message, they may also distract from its seriousness if used excessively (Figure 5.19). Furthermore, P20 found it challenging to locate specific information, such as coping strategies, within one story but found them in another, suggesting a more coherent content structure might be needed.

The cognitive walkthrough analysis reveals several areas for potential improvement in SafeHaven’s functionality and user experience.

- Long and extensive stories: There was a call for more in-depth content (P5). Rather than offering short narratives of DV survivors, SafeHaven could provide long and more comprehensive narratives or possibly link to external resources that may include

full-length testimonials, narratives, interviews, articles, or related support resources that offer extended insights into the survivors' experiences and coping strategies.

- NLP enhancement: In response to P11's feedback, SafeHaven needs to undergo additional training with various phrases and expressions relating to the stories of DV survivors, such as "coping strategies" or "methods of coping" to ensure that the CA can identify the intents of the users. This will enhance the CA's ability to comprehend and act upon a wide range of user queries.
- Emotional tone balance: After reviewing P19's critique on the overutilization of emojis, it is evident that the CA should balance the use of emotive elements while maintaining the seriousness of the subject matter.
- Incorporating coping strategies: P20's feedback suggests that SafeHaven's content be enriched by including detailed insights into the coping strategies survivors employed to cope with trauma. By doing so, all survivor stories would provide not only narratives but also valuable information on managing and overcoming adversity, ultimately inspiring users and offering practical advice.



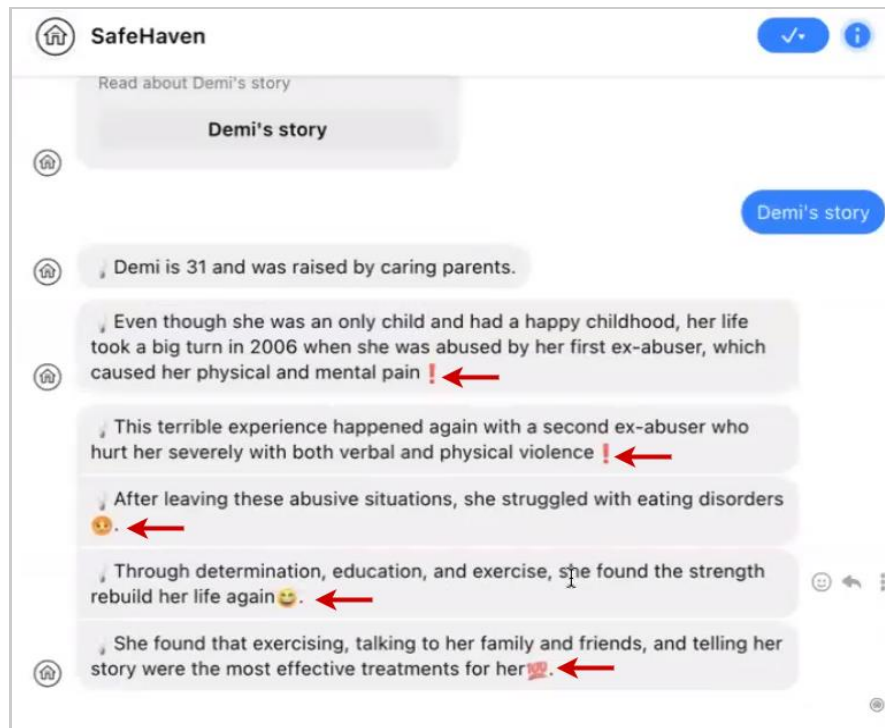


Figure 5.19: Excessive use of emojis (Highlighted in red arrows)

#### 5.3.1.6 Task 6: Find Words of Encouragement, Comfort, and Empathy for the Survivor

During this task, participants were asked if they found words of encouragement, comfort, and empathy while communicating with SafeHaven. This task concerns the delivery of emotional support, which can be a sensitive and complex interaction due to the nature of the support offered. In reviewing the participants' responses to the main questions of the cognitive walkthrough, different nuanced experiences emerge. First, we noted a low expectation rate for achieving the desired outcome (approximately 77.78%), which suggests that some participants approached the CA with skepticism about its ability to provide emotional support. Nevertheless, there was a high recognition rate of the desired action and a strong connection to the description of the action. In addition, most participants (approximately 94%) observed words of encouragement and comfort from the CA, reminding the survivor that they are not alone and that their feelings are valid as well. Furthermore, several participants showcased the CA's excellence

in offering emotional support, such as reminding them that they are not alone (P8, P22) and creating a sense of safety (P34).

Based on the analysis, some participants feel that some aspects of SafeHaven could be improved in relation to this task.

- Providing more encouragement and validation: P11's feedback highlighted the need to provide ample validation and support, especially considering that DV survivors may require additional validation to counteract experiences such as gaslighting. Additional validation words and acknowledgments indicate that the CA's role may go beyond offering empathy to acknowledging and validating the survivor's feelings and experiences.
- Language Sensitivity: P13 advocated for using survivor-centric language, avoiding generalized terms (Figure 5.20). P13 commented, *"Some language was not survivor-centered and changed terms which can be confusing like 'spouse' or 'partner' are not always interchangeable. Some validation claims can feel backhanded like 'survivors feel isolated'- that's not always true for every person so making sure the consistency of 'many survivors feel' or 'may feel.'"* In light of this recommendation, SafeHaven's language should be carefully chosen to reflect a survivor's circumstances and avoid broad statements or phrases that may not accurately represent their personal experiences.

While opinions on the CA's ability to provide emotional support varied among participants, one distinctly expressed a preference for human interaction when dealing with sensitive subjects (P1). P1 wrote, *"I understand that the chatbot was providing me with resources but with an issue like this I want to speak w/ a live person as it is a sensitive subject."* This may point to the

inherent challenge for AI, especially in certain scenarios, to fully meet the complex demands of emotional support for DV survivors.

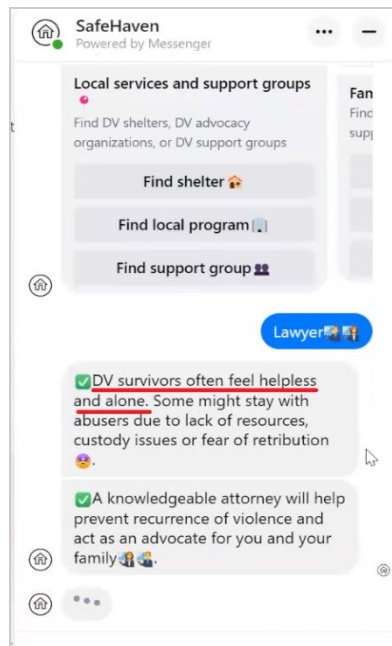


Figure 5.20: Language sensitivity issue (Underlined in red)

### 5.3.2 Interviews Results

We transcribed all the interviews after omitting all identifiable information. A thematic analysis was conducted on the interview transcripts, resulting in their thematic connections. In particular, we used an inductive method for identifying themes and patterns from interview transcripts [237]. Sample transcripts were first open-coded, followed by creating a coding schema and refining and revising it to include additional categories and themes. We repeated the above process for several rounds until the themes reached saturation. Finally, all data was coded with the finalized coding schema.

The summary of interviews is divided into five sections: advantages of SafeHaven, identified issues of SafeHaven along with recommendations for improving SafeHaven performance and

user experience, user preferences, long-term healing and recovery, and privacy and security concerns. We will discuss each of them in detail next.

### 5.3.2.1 Strengths of SafeHaven

SafeHaven was praised for its educational materials, user-friendly interface, and effective information search capabilities, positioning it as a vital resource in cases of DV. Participants appreciated its specialized features, emphasis on privacy and security, and assistance in developing custom safety plans (Figure 5.21). In this part, we will delve deeper into the benefits.

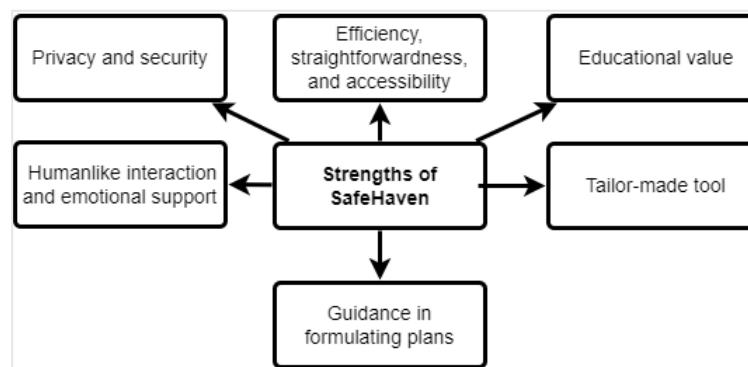


Figure 5.21: A summary of strengths of SafeHaven

- Education value:

Beyond immediate assistance, SafeHaven was viewed as a valuable educational resource, providing information related to DV, especially for those who are uncertain whether or not they are experiencing it (P6, P7, P8, P12, P15, P16, P17, P19, P25, P26, P34, P36). It was considered instrumental in raising awareness and understanding of DV, a vital first step. This CA's educational function plays a pivotal role in empowering individuals with the ability to identify abuse, enhance their awareness, and seek further assistance. P16 said, *“ChatGPT would not be able to really make them aware that they are in that*

*situation. Whereas with the chatbot, again, it goes into a lot of detail about how it's not your fault. It goes into how it just, it just goes into a lot more detail. And I think it would just really benefit any victim a lot more. Because there they will be able to read the information on there and even go into more information because you know, it keeps popping up with you know, do you want to learn more? Do you want to learn more? I think it would just all around, make a victim more aware of their situation and that they can get out of that situation.”* Unlike ChatGPT, which requires users to be certain of their needs, SafeHaven allows survivors to browse through information, offering greater convenience when they are uncertain about the specific information that they require. P15 mentioned, *“The messenger did provide the ability to look through all of the information and understand what you could ask. Whereas the ChatGPT, you did kind of have to know what you were looking for, to be able to use it.”*

- Efficiency, straightforwardness, and accessibility
  - Accessibility and availability: SafeHaven enables immediate access to vital DV resources and is available all times, a crucial feature in emergencies (P1, P2, P3, P4, P6, P10, P11, P18, P19, P29, P30, P31, P33). Since a CA is not a real human, it proves even more useful by offering information and support at any time, including late hours or when human assistance may not be available. As stated by P19, *“It's not an actual human also means that you can use it at any moment. And you do not like, you know like you could be using in the middle of the night getting answers.”* P10 also said, *“It's just as if it could be a substitute for talking with a therapist maybe if you don't have access to a therapist.”* Accessibility and availability were highlighted as one of

SafeHaven's key advantages, especially when it comes to traumatic situations where a survivor needs access to a tool that is immediately available. P11 mentioned, *"If I'm in a situation where I if it's an emergency situation or situation where I've looked like a trauma brain, and I don't have the patience and ability to just click on a bunch of different websites, I need something quick, fast, reliable, and easy to click around. chatbot did a great job at that."*

In the same vein, SafeHaven was perceived to reduce the amount of time it takes to access these resources since it does not require the completion of lengthy forms or waiting for an answer. As P6 indicated, *"the information was much more readily available, especially in a situation where you might need to look something up or get information, just within a few minutes."* The minimal typing needed facilitates rapid access to information in SafeHaven. P33 explained, *"I thought it was really quick. I liked that I didn't have to do a lot of typing and things clicked through."* P29 echoed the same point that SafeHaven offers information that is readily structured and built, making it easier and quicker to locate than a search engine, *"I think with a chatbot it like gave you the information like, already built. For the search engine, you had to like search for it more."* In addition, owing to the CA's seamless integration with various social media platforms, it is seen as an advantageous solution compared to alternative technologies, significantly streamlining accessibility. P30 said, *"Because it is already on social media, I think it will be easier to use it compared to other chatbots I mean, already you have it on your social media, so you can reach it at any time."* The integration of the CA within

widely used platforms such as Telegram and Facebook enhances its accessibility, allowing for more practical usage by survivors in need. P31 explained, *“it's on like telegram and Facebook, where a lot of people are also able to access because they have their accounts for it.”*

- Efficient information retrieval: The utilization of SafeHaven facilitated efficient and straightforward retrieval of targeted information without excessive verbosity. Relevant DV information can be retrieved easily throughout SafeHaven without needing to navigate multiple websites or access different sources of information (P3, P4, P6, P7, P12, P23, P26, P29, P35). P3 mentioned, *“The chatbot was much easier to work generally, though it is a bit fragmented with a chatbot it makes it actually makes it easier to specifically find information related to what I need to do.”* In the same context, P26 noted, *“It provided the needed information in comparison to the search engine, which was very broad, and there are a lot of the resources sometimes I couldn't be sure of that piece of information that I found online.”*
- Authenticity and relevance: As a means of raising awareness about DV, the authenticity and relevance of the information are crucial (P8, P16, P17, P28, P34). For instance, authentic survivor stories contribute to the educational value of SafeHaven by providing real, accurate, relatable, and impactful insights. This contrasts with ChatGPT's creation of fictional stories, which may lack relatability or impact as they are not based on real-life experiences. P8 stated, *“The other thing I really wanted to point out was that the chatbot used an actual survivor story. That was really brief but uplifting, and ChatGPT*

*dislike made up a story and told me that it just like made up the story, and I didn't really like that."* A similar conclusion could be drawn regarding the links and resources about DV that SafeHaven provides, which were found to be more reliable and efficient than those provided by ChatGPT. P28 explained, *"ChatGPT gave me some websites, I tried to go on the websites. But among the five websites that it showed me, only two are working. On the other hand, the chatbot actually gave me websites that most of them are working."* In the pursuit of effective support and resources for DV victims, the trustworthiness of data is critical. P28 added, *"Because I know this about ChatGPT that sometimes the links and the stuff do not work very well in there. And in the case of a domestic violence victim, this can be really bad in terms of feeling helplessness, like I can get given options, but they don't work or something. So the trustworthiness of data is definitely better in a chatbot as I think for now."* Offering authentic and relevant educational content or resource referrals empowers survivors with the knowledge and skills required for managing trauma and making informed decisions.

- Non-overwhelming source: It was noted that some participants found SafeHaven to be advantageous for its ability to convey information without overwhelming the user. This contrasts with ChatGPT or search engines, which tend to provide more extensive information and detail (P3, P8, P27, P28). As outlined by P8 in describing SafeHaven, *"It does not overwhelm you with information like ChatGPT does."* SafeHaven offered a more controlled flow of information, by allowing survivors to selectively explore options and stop



when satisfied. In contrast, ChatGPT may overwhelm users with a large amount of information. P28 said, *“The ChatGPT was, I think, a bit more in volume, about all the different suggestions. If you put one thing, one topic in ChatGPT, it will give you like a bunch of options like 7 8 9, and other options to explore. But in the chatbot, you get to choose between the options, and you can stop it like yes, I want to learn more now. So I thought that sense of being in control somewhat, was more than in the chatbot compared to the ChatGPT. The volume of information was definitely way more in ChatGPT.”* By steering clear of complex and extensive information often found in broader information sources, survivors may find the CA that provides straightforward answers and controlled experience more advantageous.

- Conciseness: SafeHaven was deemed to deliver information concisely, enabling survivors to find the necessary details about DV (P6, P25, P32). P25 noted, *“The chatbot where the information was more concisely organized, you know, versus the search engine you have to search through, you have to look through everything. Like because of the Chatbot technology, it's just more concise.”* P32 also explained, *“The first advantage is that I found the chatbot to be more concise. It gave me exactly the information that I needed to get. Right. So I asked the question, I get the exact information that I feel like I want to get. But when I did that with the ChatGPT, I asked the question, and just, it comes back at me with so much information that I didn't feel the need to actually know.”* Overall, it is vital to provide concise information via any technological tool, especially for DV survivors, as this facilitates clear, direct

advice to be provided quickly, which is vital during potentially stressful or emergency situations.

- User friendliness and understandability: Some participants found SafeHaven to be user-friendly and easy to comprehend, appreciating the simplicity of the language and the clarity of the content (P20, P24, P30). By using simple language and an intuitive CA interface, survivors could access information and support more readily. P30 pointed out, *“I will say that the Chatbot, like summarize everything that I found in the search engine. So I will say that that's an advantage. And it was, like, easier to read compared to the words that like the search engine was using.”* In the same context, P35 found that the CA's dropdown menus were user-friendly, straightforward, and less stressful, particularly for survivors who are under stress or experiencing abuse, since this method is easier to use than formulating words. P35 mentioned, *“There were there was like drop down menus, you didn't have to just type? You can, you know, it's I think it's easier for people, especially when they're under stress or experiencing abuse to be able to just kind of click through things instead of trying to form their own words.”*

- Privacy and security

- Anonymity: The CA's anonymity is crucial in situations where maintaining secrecy is vital for safety, ensuring that DV survivors can seek help without fear (P4, P15, P20, P22, P23, P32). As indicated by P20, SafeHaven allows the user to choose the name they prefer, thus enhancing their anonymity throughout the conversation, *“It did let me choose a name I wanted to go by,*

*which could be a little bit more anonymous.” The same point was also discussed by P32, “The other thing that I noticed is security. So, in the survey, I got some questions regarding security, do you feel that check if it is secure is Do you feel that your identity will be revealed? Somehow? I didn't feel that with a chatbot. Because once I'm done, everything is going to be there's going to be lost right? In the beginning, it asks me What's my name? What name do I want to be called? Right? So, I could have just given any name. I didn't need to give my correct name. But in the case of ChatGPT, I was using my I use? I mean, if I were to use it correctly, not in the in the setting, I would be using my own email, right? Yeah. So it would keep this information. And then if someone somehow gets into my email or hacks my email and then opens ChatGPT.”*

While preserving anonymity during such interactions, the technological intervention support system can foster a trust relationship with survivors, promoting their recovery and safety.

- Secure environment: Many participants perceived SafeHaven as a more secure, confidential, and private environment than search engines and ChatGPT for discussing DV-related information. The experience felt more secure than calling a hotline and explaining the situation to someone on the other end, particularly in complex or sensitive cases (e.g., sexual abuse) (P11, P18, P23, P26, P29, P32). The design of SafeHaven makes it appear less like a tool for soliciting assistance regarding DV, which is advantageous in terms of discretion. P11 said, *“That chatbot doesn't look like you're getting domestic violence assistance. So, it looks just to the eye like if my abuser were to walk*

*by me real quick, and he's just glancing over my shoulder, it does. It doesn't look like I'm doing anything that he would perceive as insidious. So that was helpful.”* There were a number of strategies that were perceived to enhance the privacy and security of survivors via SafeHaven, such as ways to exit the conversation and delete it. P23 commented, *“And also the fact that there was like in the beginning of the chatbot conversation, it gave immediate, like, exit strategies for how to exit the website, or just, you know, no longer show the website anymore. It said, it even said, it would redirect to a neutral website. ChatGPT does not have those options.”*

- Humanlike interaction and emotional support
  - Humanlike interaction: Many participants perceived SafeHaven to offer a human-like and personalized experience, thereby creating a sense of companionship and understanding (P10, P16, P20, P22, P23, P26, P31). In contrast with receiving automated, impersonal responses as with ChatGPT, many participants experienced the CA as capable of understanding and responding as if it were a human. Its ability to offer conversational support and understanding, especially in situations when immediate human assistance may be lacking, makes it a valuable tool. P6 said, *“It felt like you're talking to a person who actually understands you, especially if you don't actually have much support around you. So definitely is comforting in that way.”* Although participants recognized they were interacting with AI, the CA's engaging nature could offer a sense of companionship, as talking to a friend, which can reduce feelings of loneliness and isolation. P22 commented, *“I like that. It's*

*like I'm talking to like a friend if, because of the emojis, and that's why that's why I recommend."*

- Empathetic and comforting responses: A number of participants reported that SafeHaven provided responses that offered reassurance and validation of their experiences (P6, P8, P10, P12, P16, P19, P20, P23, P26, P27, P28, P30, P31, P33). Based on their experiences, such emotional support provided by the CA is crucial in creating an environment that is supportive and understanding. For instance, P31 mentioned, *"I think the biggest thing that I noticed the difference in was the chatbot was a lot more like, like impassionate, I guess, would be the closest thing I can describe it."* It was appreciated that SafeHaven included some reassuring messages, such as affirmations of not being alone and that the situation isn't their fault, which were absent in search engines. P33 said, *"I did like how it did say in a few points like you're not alone. It's not your fault, where the search engine did not do that."* The CA's conversational nature allows for a more personal interaction, providing emotional support in the absence of an immediate human interaction (P6, P16). Incorporating emojis into SafeHaven's interactions allows for the expression of emotions, making interactions with the CA more empathetic. P27 stated, *"So one advantage of the chatbot in comparison to the ChatGPT in my opinion, was the fact that it had more empathy towards the situation where the use of emojis and that, you know, showing expressions through those emojis kind of made it easier to discuss with the chatbot."*

- Non-judgmental demeanor: SafeHaven was also perceived to be nonjudgmental and unbiased. Due to the perception of lack of judgment, survivors feel more comfortable seeking assistance or information about their situation from the CA. P26 said, *“I don't think at this time someone would be comfortable to discuss something with someone related to them online or even talking to a person. So I think the chatbot was better than that term, because it's, you're talking to an AI understands you without judging without nothing.”*
- Tailor-made tool
  - Specialized tool: According to many participants, SafeHaven was viewed to be particularly effective in delivering information tailored to unique or rare DV scenarios, ensuring appropriate support and guidance are provided based on individual needs and experiences. Contrary to generic information sources (i.e., search engines, ChatGPT), SafeHaven was perceived to be more specialized and targeted (P16, P17, P19, P24, P26, P31, P32, P34, P35). P19 indicated, for example, that ChatGPT did not provide shelter information when requested, *“the chatbot was way more like specific, like, there were a couple of times with ChatGPT, in which I, I think it was when I asked for shelters, and another time in which it could not provide me the information. So I feel like the chatbot is was like more specialized.”* P34 also said, *“I would say the advantages of the chatbot would be that they create the information that you need.... And they're giving you exactly what you asked for.”* A capability like this is especially important for those who are experiencing atypical forms of domestic abuse or those who are in unique circumstances where general advice or information

may not be adequate or beneficial. P17 noted that *“the chatbot can always provide more accurate information and can always give advice based on or based on specific questions, and the data is much better than the search engine because the search engine can only give us some, I think it only gave us some universal responses, which are not very specific.”*

- Information and resources hub: SafeHaven was viewed by some participants as a central hub of information and resources regarding DV (P18, P31, P36). This consolidation of resources into an accessible platform streamlines the search process for survivors in need of specific DV-related information or assistance. P18 mentioned *“I think if the chatbot in the future will be similar to the one we used today, I think it's going to be helpful. And as I said before, all the resources are in one place. I don't have to search for each one individually.”* P31 referred to the same point, *“And it was all there in one place. Whereas some other things might not have all the information you need. But this kind of gave you an overview on everything you need for, like leaving to finding support to what to do after the fact.”*

- Guidance in formulating plans:

A log of participants regarded SafeHaven as a valuable tool in the development of safety plans, particularly for those with little or no experience with such plans (P3, P6, P7, P12, P14, P18, P19, P27, P32, P34). It has been highlighted as particularly valuable that SafeHaven is able to assist in constructing personalized safety plans tailored to each individual's unique situation. P18 said, *“When I used the, the chatbot. It provided me with a lot of resources and information that I didn't know about, for example, the safety plan*

*that we created. Actually, I didn't know about the safety plan and what a safety plan is, and how to create it.” P32 echoed the sentiment, highlighting the CA’s effectiveness in creating a safety plan through its use of the targeted questions and tailored responses. He noted, “But for chatbot, it just gave me exactly the information I needed. It asked me questions. It was always asking the right questions. And based on my answers, it just gave me the safety plan that I needed. And it was much, much better than the one that I was able to create through ChatGPT.”*

#### 5.3.2.2 Issues and Suggestions for Improvement

Despite SafeHaven's perceived effectiveness in providing DV support, many participants encountered challenges in forming user queries on the basis of DV terminology, which makes keyword identification and response functions challenging. Furthermore, concerns regarding slow reaction times, difficulty navigating, and inadequate coverage of topics indicate that improvements must be made in the speed, comprehensiveness of information, and usability of the user interface (Figure 5.22). In this section, we will discuss drawbacks of SafeHaven, along with potential areas for further development.



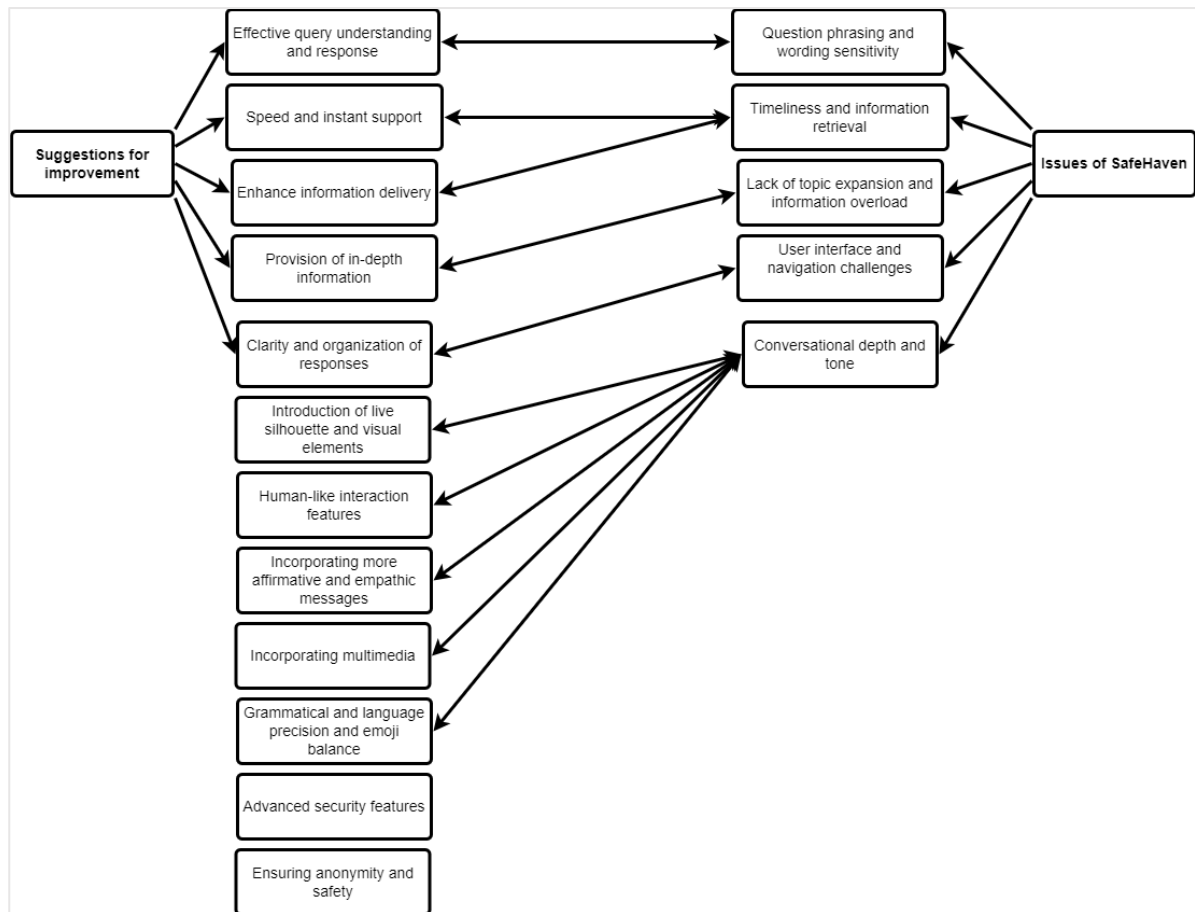


Figure 5.22: A summary of SafeHaven issues with suggestions for improvement

- Question formulation and wording sensitivity
  - Limitations
    - Out-of-vocabulary phrases in DV scenarios: Several participants noted challenges in completing some tasks with SafeHaven, primarily due to its limited capability to respond comprehensively to specific complex inquiries related to DV. In spite of the CA being integrated with Google Dialogflow and having been trained to recognize a variety of keywords, several participants felt it lacked certain response capabilities, especially for terms and scenarios related to DV (P1, P4, P7, P11, P12, P21, P22, P23, P30). This was rooted in its reliance on specific keywords to function

properly. When these keywords or phrases are not used or recognized, the CA might be unable to respond comprehensively to those inquiries. In P4's response, she mentioned, *"Just trying to figure out what to put, you know, as far as my question, or what I was looking for, I didn't always understand. You know, what I was asking, so I had to retype it and try to figure out how to say it."* It is noted that SafeHaven was not able to handle open-ended questions in some cases, which can be a hindrance to those who prefer direct communication, whereas ChatGPT performed well in this regard. P12 discusses this point, *"I feel like with the chatbot, you couldn't really like free like, open-ended questions. Just a question in that I think ChatGPT does really well."* Some participants expressed that they sometimes may be required to rephrase their queries in order to obtain satisfactory answers, and the process can be time-consuming and potentially frustrating. As stated by P22, *"One of the disadvantages there was, it gave me an answer I did not expect. So I was thrown off. In a search engine, if one of the things was not what I wanted, I could just go to the next website, or go to the next website. But so maybe that's one thing that I'm like, is a disadvantage."*

- Common expressions recognition limitation: Some participants referred to SafeHaven's inability to correctly interpret or respond to common expressions. For example, it was noted by some participants that basic acknowledgments, such as "thank you," were not recognized and appropriately responded to, impacting the natural flow of the conversation

(P8, P17). P8 mentioned, *“There were a few times where I just said like, thank you or just like basic, like phrases. And the chatbot didn't seem to understand like thank you, or thanks in. And it didn't understand.”* This Limitation can be attributed to its dependence on predetermined responses and keywords, along with the framework of models that are yet to be trained to recognize such common expressions.

- Recommendation

- Effective query understanding and response: Some participants called to enhance SafeHaven’s capabilities to understand and respond to user queries effectively. This is crucial for enhancing user experience, as highlighted by participants P4, P8, P13 and P17. For instance, P4 and P13 noted the importance of improving the language style of SafeHaven so that it correctly interprets user queries and takes into account both their literal meanings and context of user input. P4 mentioned, *“I think just like I said, communicating, asking the questions the right way. Maybe a little more user friendly, kind of stumbling and guessing, how to ask what, I need to get that information. Because, like, if you use your phone, you know, you do, you know, Google voice assist, or whatever they call that really, you know, go on there and say, I don't know, nearest restaurants, you know, near me or whatever. It'll pull up that way. But if you don't ask it properly, you might get something that you don't even want.”* It may be possible to address this issue by providing more instruction at the beginning of conversation regarding how to formulate a query, while

simultaneously training the CA with related keywords and utilizing advanced NLP techniques to grasp the nuances of various questions. In the same context, P8 and P17 emphasize the importance of SafeHaven being able to recognize common phrases and simple affirmations such as "yes," "no," "Thank you," or "OK." This understanding helps reduce user frustration and maintain a natural conversation flow. P17 explained, "*The only recommendation is that I recommended that maybe improve the chatbots recognition like some common words like okay, yes no.*" It is possible to reduce survivors' frustration and maintain a natural conversation flow by training the CA to recognize commonly used phrases and simple acknowledgment. One approach is to use machine learning algorithms to analyze a large dataset of conversations and identify patterns in how survivors respond to them.

- Timeliness and information retrieval
  - Limitations
    - Slower response: Although some participants preferred to slow the pace at which survivors received information (P5, P13), others have pointed out SafeHaven's slower response time as a shortcoming compared to search engines/ or ChatGPT (P2, P3, P9, P18, P23). This may pose a challenge in urgent situations that require the quick retrieval of information, such as locating DV shelters, in order to ensure that survivors are safe and are able to meet their immediate needs. For instance, when we asked about the shortcomings of SafeHaven, P3 answered, "*it could be a bit too slow, but*

*that's just me, because I can make I can lead fairly quickly, per se. I mean, at least to me, it was a bit slow.”*

- Lengthy responses: A notable limitation of SafeHaven identified by some participants is its tendency to provide lengthy responses to questions regarding DV (P17, P18). This may confuse survivors seeking direct information and instant assistance. P17 explained, *“I think the only disadvantage is that sometimes the chatbot it gives some very lengthy responses, but with the search engines always give us some very short and straightforward responses.”*
- Recommendations
  - Speed and instant support: Several participants (P2, P3, P22, P18 and P29) recognized that quick responses are an essential feature of SafeHaven. It is especially advantageous for survivors in an emergency situation to receive prompt responses from the CA, so that they could obtain information and support immediately, regardless of whether they are searching for shelter, making safety plans or connecting with emergency services. P2 said, *“I think the chatbot would just have just be quicker, just because it was a lot of steps just to find one answer rather than a search engine was so fast. You just have to Google it.”* In the same context, P3 suggested a speed option for users, allowing them to select the pace of interactivity based on their personal preferences, *“Make the speed a bit faster, at least for me, or, maybe like, show them not gage, but have the user like, select a speed*

*option.*” This helps address concerns regarding slow interactions and provide more efficient support.

- Enhance information delivery: This theme explores both the content and presentation of DV-related information (i.e., how much and how it's delivered). There is a suggestion for reducing text length (P18). This refers to making individual responses shorter and more concise. Lengthy responses can be difficult to digest, especially when they are being viewed on mobile devices or in cases where survivors are seeking immediate assistance. P18 clarified, *“I think it's just a little bit lengthy so just reduce the amount of text.”*
- Lack of topic expansion and information overload
  - Limitations
    - Limited breadth: Some participants reported that SafeHaven sometimes lacked the ability to expand comprehensively on certain topics (P3, P27, P28 P31, P35). For instance, SafeHaven was perceived to provide less detailed and less informative responses compared to ChatGPT, as highlighted by P35, *“It was less detailed, less information available.”* A few participants also encountered the initial difficulty of navigating SafeHaven, particularly when entering additional queries beyond the scope of the tasks. This limitation may hinder the CA’s effectiveness in responding to more complex or specific inquiries. P27 explained this when she tried to search for antivirus software in Task 4, *“I wish that if I could ask them further, more about something they would expand, like, I was*

*talking about the antivirus software. And I asked him to, like, expand a little more on to that. And it, I don't think he understood my message. So I am kind of hoping that the chatbot would maybe soon be able to have additional information about everything it provides to us in a way.*” In the same vein, two participants noted that even though SafeHaven offered a customized plan, it lacked refinement in its responses, as it seems it was not able to address situations such as having nobody to care for children at risk (P28, P33). P28 mentioned, *“And in the chatbot as I mentioned, like they're giving me it's giving me a plan, customized plan, but I thought when it asked me who can look up, look after your kids, while you're not home, or you're in danger, I typed nobody. And it gave me a plan that has nobody in the plan. And that is like a huge thing that I would say. Because if I don't have an don't have anyone that should be addressed in there.”* Therefore, it is still necessary to improve the review and finalization of plans, delving comprehensively into more complex or specific user needs and scenarios. Overall, this limitation in exploring topics in greater breadth based on participants' requests is due to the CA's development and implementation, which may not encompass a broad range of topics and keywords to deal with complex and specific inquiries. They suggest that improvements can be made by incorporating a broader spectrum of information into the CA and training it with a more extensive array of keywords to address the issue effectively.

- Perceived limited resources: There was a perception among some participants that SafeHaven's range of information and resources on DV was limited or incomplete, implying that it may not include all aspects or details (P25, P26, P27). Even though SafeHaven is specifically tailored for DV matters, this perceived constraint in its resource pool might impede its ability to offer thorough support to individuals in need of assistance or information related to DV. P25 clarified, *"We mentioned the idea of having limited resources, because if using chatbot would actually provide with the needed, like minimum needed information. And that's it. However, the search engines would open new gate, like there are a lot of multiple resources. A lot of information probably leads here to different like, topics."*
- Information overload: One participant pointed out that SafeHaven displayed an extensive amount of information at once, which may result in the user feeling overwhelmed (P16). Nevertheless, this participant also indicated that SafeHaven gives the user the option to choose whether to receive this additional information, thereby helping alleviate the issue. P16 said, *"The only thing I could really say is that it is just like, you'll ask a question, and it will give the answer, but then it will keep giving more information. So sometimes, it can just be a little overwhelming. It can just be a little too much on the screen. But at the same time, it gives you the option of whether you want more information. So, I guess in that way, you can choose not to, and then you avoid that problem."*



- Recommendations

- Provision of in-depth information: There are suggestions to improve the SafeHaven's content by not only providing summary responses to survivors but also offering them comprehensive DV-related resources and more in-depth and breadth information to really help them understand and cope with their situations (P3, P5, P8, P21, P24, P27, P31). As an example, P3 indicated the importance of offering expanded safety plans or additional development options based on the information provided at the beginning. P3 clarified, *"If you want to make a bigger safety plan, like you know, expand it, expand it after you get the starting points down there. Maybe ask if you want to continue developing it or if you want to show it."* P5 and P21 also pointed to improving SafeHaven's capabilities by offering comprehensive information and providing links to additional national DV related resources. Similarly, P31 stressed the importance of improving SafeHaven's ability to expand on its answers and provide detailed and specific information to survivors who may not be aware of what to look for in specific situations. P31 stated, *"I guess just like specification of stuff, because like I mentioned earlier, how, when I was talk, or whatever, I asked a chatbot about, like, what evidence I should get. And it came back with the one about finances, and I asked it, well, how am I supposed to know? And it just returned the same answers again. I think just because some people like I genuinely didn't know, like, what I was supposed to look for. So I thought that'd be a good thing to ask it. So I think that being*

*able to provide specific information, or at the very least, like examples, that would be definitely good.”*

Along the same line, a suggestion from P24 was to include more questions that delve into the specific circumstances of the survivors. SafeHaven could thus gain a better understanding of the survivor's needs, and then offer them responses tailored to their specific situation. Furthermore, P8 and P27 advocated for the addition of more branches and options for elaboration, enhancing SafeHaven's capability to provide in-depth and specific information. In this way, the CA would be able to provide survivors with advice tailored to their scenarios, without having to use the internet to search for more information. For example, P27 mentioned, *“If there were more branches into the question, like if some information is provided, and if someone asks about that specific piece of information, I feel like it'd be kind of cool, if there could have been some more options that they could elaborate on. I think that would be really cool because then it would be used more in depth where the I think the victim can get more knowledge from the chatbot without having to move away to the internet and search more things by themselves.”* Considering these suggestions collectively emphasizes the importance of SafeHaven providing comprehensive answers, which potentially increases its effectiveness as a tool for providing survivors with detailed and tailored guidance.

- User interface and navigation challenges

- Limitations

- Learning curve for new users: For survivors, especially the elderly, who are not well-versed in the CA technology, understanding and using the CA interface may present a challenge (P10, P15, P19, P30). P30 explained, “*It was new for me to use a chatbot, like in Messenger or even in Telegram. And I think it was kind of difficult for me. I mean, after some tasks, I got used to it, but maybe for someone like, old, like an older person, it can be difficult.*” This complexity is likely to discourage those unfamiliar with new technology from fully benefiting from its assistance. In the same regard, due to the novelty and unfamiliarity of the CA, concerns have been raised regarding its reliability. P10 said, “*It's unfamiliar and, new technology. So you don't know how reliable it is?*”
- Navigation issues: Some participants encounter navigational issues namely moving through the various modules of SafeHaven, hindering their interaction and experience (P15, P19, P23, P30). It is exacerbated by the difficulty in identifying different modules and sections of SafeHaven, which might be due to a lack of sufficient clarity, thereby impacting its usability and effectiveness. P19 commented, “*the chatbot was a little bit confusing at times. Like, it wasn't intuitive for me, like, I think it was main resources and Main Menu, like I wouldn't intuitively know the difference, or what I can find under which of them I feel like, you know, like the structure was confusing somehow, like, which information would be under which topic gets set up.*”

- Excessive processes: SafeHaven, in some instances, necessitates navigating through multiple steps to yield results, a process that participants found to be less efficient (P2, P23). In situations where quick information retrieval is crucial (e.g., DV), the multi-step procedure may impede survivors who are accustomed to the immediate results provided by conventional search engines. P23 commented, *“You had to click on things multiple times, or, like, there were like, subsections of things you had to like click to in order to be able to find certain things.”*
- Comprehension issues: One observation noted a drawback that survivors could potentially face in grasping the full range of functionalities offered by SafeHaven when seeking information about DV (P6). A survivor may initially find it difficult to comprehend what the CA can do, which aspects of DV it can provide information on, and how to interact with it effectively. P6 discussed this point in her response, *“I didn't realize at first that it had so much information in so many ways that you could input different things to get information from the bot. But once you realize you can type in anything, it'll actually give you real results back then that's when it started to become very easy to use.”*
- Trust issue: A further shortcoming was pointed out by P36, who explained that since the CA is not a human, it may not be able to understand a survivor's particular situation, even if it appears to do so, thereby negatively affecting trust. P36 mentioned, *“I guess a disadvantage could be that it's a bot and not an actual person, because even though it can show*

*that it understands you, you might not necessarily feel completely understood about your specific situation.”*

- Recommendations

- Clarity and organization of responses: As far as the need to enhance user experience is concerned, there was a focus on improving the clarity and organization of SafeHaven's responses and features (P9, P12, P16, P19, P23, P24, P28, P30, P33). These suggestions covered the layout, format, navigation, features and the quality of communication of SafeHaven. For example, P12 discussed the importance of clarity on SafeHaven's features such as the safety plan, to ensure that survivors understand the purpose of the feature before engaging with it. P12 stated, *“I guess that like, I didn't really understand what a safety plan was explained to after I clicked it, but my only concern is would someone not click on that option if they didn't know what it was?”* P28 also suggested refining the review process before finalizing a safety plan, ensuring the inputs are meaningful and providing points of contact with backup options.

There was an emphasis on the need to improve the layout or format of a conversation, particularly by Improving the information presentation in menus to enhance clarity and organization while retaining those elements that contribute to user comfort (P16). In the same context, P9 discussed the point of including explicit prompts informing survivors that they may either select a button or type freely, thus providing a mix of guided options and open-ended input selection. P9 said, *“there were a couple of parts*

*where I wish it had reminded me that I could type something. It was just said, like, you know, pick a button or it could have said, like, pick a button or tell me what's on your mind. What do you need help with? Here are some options or type anything you want?"* P30 referred to a similar recommendation, suggesting that clicking buttons and writing queries should both be available to accommodate different survivors' preferences.

One suggestion for improving the navigability of SafeHaven is to implement a single, non-embedded menu instead of multiple menus and submenus, meaning for a more intuitive user experience (P19, P35). Further, incorporating a visual map or guide could also further assist in navigating SafeHaven's modules (P19). P19 noted, *"Maybe not have them, like the menus embedded into one another? Just like one single menu, maybe that would be easier. Or, I don't know, maybe providing some sort of map."* There was also a discussion concerning the importance of having a consistent presentation across all menus (P35). P35 clarified, *"First, that first question when it says, like, how are you feeling? Why does it give you a different like button, as opposed to the rest of the questions? Like if that could be, if that could give you the same menu as the rest of the questions."* Another suggestion was made by P23 that an interactive sidebar would be a valuable complement to SafeHaven. This sidebar would allow survivors to seamlessly switch between different conversations and topics (e.g., safety plan, finding shelters), even while

the CA is generating responses. This would enable them to multitask within the chat, accessing various information sources without losing track of their current conversation. P23 said, *“maybe like a sidebar of not like a collapsible one, but like a non-collapsible sidebar of like the main menu where you could flip in between topics, even when the chatbot is generating response, I felt like that would have been easier to use than waiting for it to generate response and then give you the option to go back to the main menu to look at something else.”* Similarly, P25 recommended improving user experience of SafeHaven by making the top menu more visible and easier to navigate. In summary, improving user experience, particularly for first-time users, involves enhancing clarity, organization, and navigational ease by improving layout, explicit prompting, and intuitive features such as sidebars and single menus.

- Conversational depth and tone
  - Limitations
    - Lack of personalization: While many participants praised SafeHaven’s personalized approach, some felt that its responses lacked the depth and nuance associated with human interactions (P7, P14). Despite tailored responses, these participants felt that the CA's responses lacked detailed engagement that resembled real-life conversations. P7 said, *“So it's not really personalized yet. I can see I mean, personalized in the future, though more, the more that, you know, they work on it.”*

- Conversational tone: Although some participants praised the incorporation of emojis throughout the conversation as they made communication feel more personal and approachable, others indicated that emojis were used more often and made the CA's conversational tone appear informal (P9, P12, P16, P19, P23, P33). It was felt that the incorporation of emotive symbols affected the naturalness of the interaction. P33 mentioned, *“I thought I thought there were a lot of emojis everywhere. And to me, that just kind of feels fake, I guess.”*
- Recommendations
  - Introduction of live silhouette and visual elements: Some participants have expressed the need for a more personalized CA experience (P1, P7). One suggestion discussed the use of a live silhouette to create an illusion of speaking with a real person, which can foster a deeper sense of understanding and connection (P1). P1 noted, *“How to be a more personalized experience. Be able to have like a live person on there, like, are maybe like, a live silhouette of a person so that you think that you're talking to a live person.”* Another suggestion by P7 was to use recognizable and familiar faces, perhaps using emojis or some other visual elements, as a way of enhancing engagement and personalized interactions. P7 explained, *“Something to recognize. In the chatbot. Like, one familiar face. Lady or man or, maybe like a familiar face or something like that. That personalized, like with the emojis. I love that. I do love that. But like maybe some of them.”* The essence of these



recommendations revolves around not only delivering practical support in DV scenarios but also offering a more human touch in their interactions. By incorporating live silhouettes and familiar visual cues to enhance the anthropomorphic cues, the CA can promote a sense of empathy toward DV survivors [350].

- Human-like interaction features: Providing an experience akin to speaking with a friend or family member was also suggested to improve SafeHaven's human-like features (P26, P28). An illustration of this concept would involve adjusting the tone of SafeHaven according to the context of the conversation and to maintain a conversational style like that of a human. P26 said, *“I think they can be enhanced by making it more human like imagine like, because the feeling that you have talking to a friend or a family member or your mom like the person you're talking to has a soul and that soul like brings more feelings and sensation when you talk to someone.”*
- Incorporating more affirmative and empathic messages: There was a suggestion that more affirmative and empathic messages are needed in SafeHaven (P14, P19, P34). This type of message can help survivors feel assured and validated, which may contribute to a more humane and empathetic interaction. The CA might respond, for example, with "It is not your fault" or "You are not alone." These statements may allow survivors to feel heard and validated, allowing them to speak freely and seek support. P34 stated, *“I would say maybe after they give the information,*

*like, um, helped me devise a safety plan, the chatbox gives like the safety plan details and then validates their situation like again, I'm sorry for what happened to you."* In the same context, immediate safety messages should be accessible to survivors, serving as a reminder to prioritize their safety. P19 mentioned, *"Maybe more emphasis, because I did see that. It's like sometimes, but more emphasis on if you are in danger right now. Hold on one. You know, Chatbot. And like, don't feel guilty for calling 911."* The CA, with its thoughtful and empathic composed messages, offers validation and affirmation to DV survivors [29].

- Incorporating multimedia: Expanding on the theme of personalization features, there has been a suggestion of integrating multimedia features into SafeHaven (P14, P28). In particular, it was recommended to incorporate elements that may improve the survivor's mood throughout the CA interaction, such as music, voice or video features, to increase its ability to demonstrate empathy. For example, P28 explained in her response the potential of incorporating music or video features into SafeHaven, *"Maybe suggesting something nice, just to make them feel good. Maybe your music, maybe a video like, Okay, you're done with your safety plan. Now try to relax and enjoy this funny thing, or nice music like just to distract people, because the people who are asking for a safety plan, they're obviously distressed, and you're trying to introduce the concept of empathy, and sympathy and humanistic words of affirmation, that might be a thing, just hey, I helped you with this. Now, just let me help*

*you lift your mood. That could be one thing, because I realize you're focusing on the humanistic characteristics and a chatbot. So that could be one thing you could think about integrating.*” Multimedia features can enhance the social interaction experience in CAs [351].

- Grammatical and language precision and emoji balance: There was an emphasis on enhancing SafeHaven’s interactions by using clear, simple language and grammar, along with a balanced use of emojis (P6, P9, P16, P22, P23, P28, P33, P35). The call for grammatical precision, as highlighted by P6, P9, and P33, underscores the importance of adhering to proper grammatical standards in SafeHaven's responses. For example, P9 mentioned, *“There were a couple things where I think there were some grammatical errors in the way that it had in the way that it was written.”* Attention to linguistic precision transcends mere accuracy; it is essential to enhance the professional demeanor and perceived credibility of the information conveyed. This, in turn, promotes a sense of trust and assurance among survivors seeking dependable and accurate assistance from the CA in matters of DV. In the same context, P22 discussed the point of simplifying the language to facilitate easier understanding. Concurrently, there was strong advocacy, noted by a number of participants (P6, P9, P16, P23, P28, P33, P35), for a balanced approach to the use of emojis and address the tone of SafeHaven. For instance, P28 noted, *“Maybe go easy on using the emojis, it might distract some people. It distracts me, because I'm not a big fan of emojis. But I also understand*

*that's a standard practice.*” Emojis may add personalization and relatability to the CA conversation, making the CA more approachable and user-friendly. However, there's a delicate line to tread, especially in contexts like DV, where the overuse of emojis might undermine the seriousness of the subject matter and potentially discourage survivors. In summary, maintaining proper grammar and using emojis in a balanced manner are vital for creating engaged CA interactions. The CA must be acutely aware of survivors' emotional and informational needs and efficiently communicating information.

- Security and privacy measures
  - Proactive enhancement
    - Despite not being explicitly identified as a limitation, the importance of strengthening security and privacy measures in SafeHaven cannot be overstated, especially given the sensitive nature of DV interactions.
  - Recommendation
    - Advanced security features: Incorporating advanced security features into SafeHaven, as recommended by P11, is of paramount importance when it comes to sensitive interactions, particularly those involving DV. One key recommendation by P11 is implementing a robust verification process (e.g., multi-factor authentication) to enhance the CA's security framework and ensure secure login. This goes beyond common security questions since abusers are often familiar with survivors' personal information. P11 explained, *“I think maybe beyond that a to verification. To get into the*

*chatbot, like it being more than just an email because again, if he hacks that email, and if he has access to my phone number, like and it's like, you know, share point like, I need there to be some kind of verification process where that would be something my abuser would not be able to get into. And not the whole, what's your mother's maiden name, stuff like that? Because again, abusers know their victims really well, usually. So, keeping that information secure and safe from somebody that knows you as well, as you know yourself, is very, very difficult to do. So, I would say something along those lines, like multiple steps of verification.”* In addition to a password, survivors may be asked to enter a code that is sent to their email address, answer a secret question, or scan their fingerprints as part of multifactor authentication. In the same vein, P11 gave another recommendation by introducing the concept of utilizing unique codes for accessing different sections within the chat interface to compartmentalize and safeguard sensitive information. This strategy prevents unauthorized access, while at the same time minimizing the risk in the event that certain parts of the CA are compromised. P11 continued in her response, *“And then also, I think, maybe a little code like ChatGPT, there's the individual like little chat titles on the side on the left, maybe a code to get into each one of those. So when you click on that to even access it, you need to have a code. That would be helpful, because that would mean even if he did access the entire like site, he couldn't get into each of those little chats.”* Overall, implementing more robust and advanced security measures is

crucial for creating a safe and secure environment when communicating with SafeHaven. Survivors feel more comfortable sharing their personal and confidential stories with the CA when they know their information is safe from any unauthorized access [352].

- Ensuring anonymity and safety: Throughout our interviews, there was a call for more anonymity and safety when communicating with SafeHaven (P25). P25 proposed to begin interactions with upfront inquiries about the survivor's current safety and desire for anonymity. P25 explained, “*Just the thing with the security and the safety, like anonymity saying, you know, do you want to remain anonymous? Are you safe now before you start the chat?*” Such a feature would allow survivors to express their need for discretion, which may enable the CA to adjust its interaction protocol accordingly. As an example, if a survivor indicates his/her desire for anonymity, the CA may not store personal data or conversation logs, or may provide the option to exit the chat immediately or delete the chat history. In this way, survivors are offered control over their information and interaction as well as a proactive stance used to ensure their safety.

### 5.3.2.3 User Preferences, Long-Term Recovery, and Privacy Concerns

#### 5.3.2.3.1 User Preferences: SafeHaven Vs. Search Engine/ ChatGPT for DV Assistance

- SafeHaven preference: It emerged that 21 out of 36 participants favored using SafeHaven over other tools like search engines or ChatGPT, mainly due to its specialized approach as well as personalized interactions tailored to DV issues. These participants expressed

appreciation for the CA's ability to deliver pertinent and specific information regarding DV, such as local support services and informational tips and advice, in a way that fostered companionship and comprehension among the chatter's users.

- Search engine preference: Among the participants, six participants indicated that they preferred search engines over SafeHaven for acquiring information or seeking assistance regarding DV. Search engines provide a wide range of information regarding DV with efficiency, which explains this preference. The ease of use and familiarity also play a significant role in this preference. Search engines are a familiar tool, and participants have been using them for a long time, making them more accessible. When dealing with sensitive issues such as DV, this familiarity with search engines fosters comfort and reliability. In addition, their user-friendly and simple interfaces and the large amount of search results make search engines a preferred source for some participants.
- ChatGPT preference: Only two participants indicated a preference for ChatGPT compared to SafeHaven. These participants perceived ChatGPT to be less restrictive and capable of processing complete questions more efficiently. Its direct answers and ability to handle complex questions made ChatGPT particularly appealing for providing more detailed information about the DV topic or providing technical advice regarding safety measures (e.g., mobile safety, online safety, cyberstalking prevention). They found that ChatGPT provided more relevant answers addressing the entire context of the query, rather than focusing on a limited set of keywords. It was observed that the ability of ChatGPT to consider the entire statement, preventing dominant terms such as "domestic abuse" from overshadowing other essential aspects of the query, was especially advantageous.

- Mixed preferences based on situation and need: It was observed that six participants had mixed preferences. The choice between SafeHaven, search engines, or ChatGPT, depended on the specific situation, the type of information sought, the emotional status, or the case urgency. Among the factors influencing their decisions was the nature of the DV inquiry, the level of detail that was required, and the context of the information search (e.g., providing assistance to a victim versus gathering information for a friend or family member, the preference of detailed information vs. general information). Participants acknowledged that each platform offers unique advantages, but their platform preference may depend upon their immediate needs and emotional distress.
- Preference for direct communication with experts: As compared with all other participants, one participant expressed a preference for direct communication with knowledgeable professionals, instead of utilizing CAs or other technological tools. The preference highlights the importance placed on the depth of knowledge and personal touch that arises from communicating with experts in the field. Such communication offers nuanced insights and tailored advice and may reflect a level of expertise and empathy that may not be fully captured by technology-based solutions.

#### 5.3.2.3.2 Conversational Agents: Facilitating Long-term Recovery for DV Survivors

The participants exhibited diverse perspectives on how CAs may contribute to fostering long-term recovery for DV survivors. Some participants believed that CAs can provide survivors with essential resources and referrals. Others felt that CAs could provide emotional support and act as a source of comfort and guidance. Finally, some thought CAs could simplify the recovery



process and offer guidance, especially in the initial stages. More details will be discussed in this section.

- **Resource provision and referrals:** Several participants underscored the potential of CAs in providing essential resources and referrals, a key component in the long-term recovery process for DV survivors (P4, P5, P6, P7, P11, P12, P17, P18, P20, P22, P25, P29). The participants perceive CAs as a potential resource hub that enables a survivor to access a variety of information and resources. A few examples are given in P4, P5, P22, and P23, which emphasize the importance of chatbots in guiding survivors to vital services such as hotlines, shelters, and counseling services. P4 mentioned, *"I think just having different resources available and you know you had the 800 hotline and shelters and, you know, people to talk to every everything that you might need to help you recover? I think it was very thorough. I don't think they left anything out. They don't."* Further elaborating on this theme, P6, P7, P11, P12, and P18 emphasize the CA's role in facilitating connections with support groups. This element facilitates the connection between survivors who have undergone similar traumas, fostering a sense of community. P11 said, *"I guess chatbot can link you, you can like use chatbot to find maybe I could chat and I can like to type in their online support groups. And I'm sure chatbot would like bring up some online support groups that I can click that I can just click directly to and then boom, I have access."* P15, P17, P25, P27, P28, and P34 further suggested that CAs might share stories of past survivors, thus providing a sense of hope and connection. P15 commented, *"By sharing the success stories of other people who are in the situation, it helps to it helps to show them that they're not that they're not just the one person that this has happened to the there is a chance for*

healing, and there is a chance to get things so there's a chance that things will get better.” Another insight is that CAs can enhance their contribution by providing links to resources such as uplifting videos, music, and inspirational stories. This approach aims to deliver a more humanized and engaging experience, extending beyond traditional text-based interactions and multimedia. It will also include a supportive aspect that creates a deeper emotional resonance with DV survivors. P28 commented, *“So in terms of providing resources, I think if the chatbot would give me some videos, or link to some YouTube videos, or free access tools, like music gonna give me like, Hey, if you're feeling kind of sad, this, check out this music, it might lift you up, or check out this inspiring story, not just sending me text all the time, but maybe funny, or kind of uplifting video or music, or something that involves seeing another human being instead of just seeing texts all the time, because no matter what it doesn't do, it will always be bought, there's not a person. So sometimes seeing a person in there, it's kind of helpful.”* Counseling referrals are another area where participants specifically highlight the role of CAs in providing information about them and facilitating access to such services (P4, P5, P6, P12, P17, P20, P29). Participants assert that CAs have the potential to refer survivors to DV professionals or psychological services, which is crucial for long-term recovery. P12 explained, *“I know has it on there, but definitely counseling. Maybe different, like, communities of support where they have like meetings. I feel like the main one is counseling and therapy, affordable counseling and therapy.”* CAs can compassionately connect DV survivors with treatment for their ongoing needs by offering timely referrals and valuable information about counseling options.

- Emotional support and affirmation: Many participants discussed a CA's ability to provide emotional support and affirmation as a means of facilitating long-term healing for DV survivors (P8, P9, P10, P15, P22, P23, P24, P26, P27, P30, P34, P35). Empathetic components of CAs assist DV survivors in understanding their circumstances, emphasizing that they are not at fault, and affirming their victim status. By providing immediate reassurance and information regarding available resources, this approach significantly contributes to healing during its early stages. For instance, P8 demonstrated the importance of continuous affirmations and reassurances, especially for survivors that require ongoing emotional support. P8 stated, "*The whole message of it's not your fault. It's never deserved. No one deserves to go through that. That's something that a DV survivor like, even after years of not being abused, like still needs to hear going on. So it's not like just like, when you're leaving the situation, do you need help, like, you need help for years, you know, so having like emotional support, like online that can remind you of like the good things and remind you that it's not your fault is really positive.*" It was suggested by P9, P24 and P26 that CAs in general may be enhanced to function as companions by providing features such as journaling, check-ins, memory recall, and daily affirmations. P9 said, "*Maybe if you had like a messaging or a journaling feature, if there was a memory for different users, and you could check in, and my chatbot could message you and ask you how you're feeling.*" It pointed out the importance of empathic components in CAs, which can be demonstrated by understanding survivors' circumstances and periodically inquiring about their well-being, even if a CA is believed to be an AI program. In the same vein, it is suggested that if a survivor shares his/her situation that may require

more empathic language, a CA can intelligently understand and respond with increased empathy and emotional understanding. P27 said, *“if you could have a conversation with it, like not specifically asking for stuff, but like kind of telling them your situation, would it kind of still continue to show like more empathy and emotions?”*

- Simplification and guidance in early stages: Participants like P29, P31, and P32 pointed out the importance of CAs in simplifying the recovery process, especially in the initial stages. It was emphasized that while CAs are currently focused on addressing immediate needs, they have the potential to offer comprehensive resources for long-term recovery, including basic information and general safety plans. Participants viewed the potential benefits of CAs in providing plans and guidance during the post-crisis period. For instance, a CA can provide advice on what to do next, what safety precautions to take in future, and how to maintain privacy. P31 explained some examples, *“there was the information on how to like prevent cyberstalking in the future with like, keeping up with apps on your phone regularly and keeping sure are making sure all your passwords are kept secure, and things like that. But I don't know, maybe there's like, there could be another like section or a subsection in that main menu to talk about, like, what to do after the fact. I think that would definitely be helpful. Because for trying to find the Preventing cyber stalking, I had to, like manually type it in which I like, as soon as I did it, it came up, obviously. But I do think that having like another section subsection for feet, like how to account for what might happen in the future would be useful.”* P32 discussed a similar idea that CAs may keep survivors' records and offer a recovery plan that may span a long time

such as a couple of years. P32 said, *“if it can keep records for that specific person, or have that been going through and then like, gives a plan for instead of like a safety plan for what you need to do now, or first, the plan that you need to use whenever you need, they're not something for the future. Okay? So like a 10 steps, process, right? You do this and this and this. And then when you get to the 10th step, that would be like two or three years in the future.”*

- Limitations and future potential: Some participants expressed skepticism about the effectiveness of current CAs in promoting long-term healing for DV survivors, citing limitations in their present applications and capabilities (P1, P3, P13, P33, P36). In their opinion, CAs could address immediate needs. However, to maintain long-term recovery, human advocates and face-to-face interactions such as therapy and support groups would be considered essential. When asked about how chatbots can promote long-term healing for DV survivors, P33 responded, *“I don't really think they can. I don't think there's a replacement for human therapy.”* P13 also added in the matter, *“In terms of like longer term support, like I couldn't, I don't think like I would use a chatbot to then be like, Hey, this is what my story is, how can you help me because it wouldn't recognize or understand that. And like with some of the emotional validation, like, there are definitely some key points that you don't see in search engines. But I also think like, those key points can be misconstrued where it feels very robotic, and like not genuine to be like, you're not the only one who is experiencing this or, like, during tough times. But I also think like domestic violence is very nuanced. And there's no great way to offer like, empathy. In a like, general statement like that, like, that's what makes domestic violence advocates so important.”* P36 was also skeptical

about the ability of CAs to contribute to long-term healing for DV survivors, “*I feel like people would have to go through lots of therapy in order to actually address the issues and to work through them. So no, I don't see that.*” Overall, while human advocates are irreplaceable due to their nuanced understanding and emotional validation in DV situations, CAs could complement their efforts by facilitating ongoing recovery and offering additional support.

#### 5.3.2.3.3 Privacy and Security Concerns of SafeHaven, Search Engines, and ChatGPT

Our interviews revealed several concerns about privacy and data security concerning SafeHaven, search engines, and ChatGPT.

- Privacy and security concerns of SafeHaven:

On one hand, some participants (P4, P5, P16, P26, P29, P31, P34) reported that they had no privacy or security concerns using SafeHaven, which may reflect their confidence in the technology, their choice of platforms, or perhaps their lack of understanding of potential privacy risks. On the other hand, despite the implementation of various safety and privacy measures intended to reduce risks for survivors, concerns about the use of SafeHaven still persist.

- Search history and information being discovered
  - Risk of abusers discovering information: A common concern was that abusers could discover survivors' chat histories or search information and track their activities, a scenario that could escalate the danger for survivors (P1, P2, P8, P11, P14, P15, P17, P24, P25, P33, P35, P36). This

apprehension was particularly heightened when the chatbot was integrated with personal social media accounts. There was a fear that such integration could allow abusers to track a survivor's online activities and see all past conversations. P11 explained, *"It would be the only the only concern I would have is that it would be linked to a Facebook or linked to an account that if my abuser ever discovered that he might have access to all of that information, and that is very scary."* P1 referred to the same concern, *"I just want to make sure that my attacker is going to come after me and like find out like that I did this his back. I just don't want him to go on my computer and find out information. I think it is probably my number one concern I see."* In the same context, one participant expressed concern about survivors inadvertently disclosing their identities if they are not explicitly instructed to stay anonymous or use fake names, especially if they are asked for a name before assessing safety (P25). Engaging in such discourse may heighten the risk of abusers discovering their victims. P25 described her concern, *"I would say, when it asks, what you want to be called, have maybe put a little bit further instruction that says use a, you know, use a name that no one can identify, use a name that is, you know, anonymous or something like that, just because it when you first, you know, say let's get started with the chatbot. It doesn't say like, it asks you after your name, if you're safe, or if you feel safe. So maybe have it ask those questions first. And if you say no, then say, Okay, what code name? Would you like? Something like that? Because not, not everybody could*

*understand that that's where it is going. And they might end up, you know, revealing themselves.*” Survivors expressed concerns about potential discovery by their abusers, underscoring the need for heightened security measures in such sensitive applications.

- Concerns about deleting chat histories: Some participants were concerned about the ability to delete chat histories, as traumatized individuals may not remember how to delete the chat history (P13, P15, P18). This concern was particularly pronounced in platforms like Facebook Messenger, where there's a perceived risk of past conversations being accessed by others. In her response, P15 questioned about the capability of the CA to delete a chat history, *“Does the chat delete once you close it because if not, then the abuser if they're or controlling the person's social media can look and see that the person was looking for help, which could put the person in immediate danger.”* In light of this, while SafeHaven offers an option to delete chats and provide instructions for survivors, it may be more crucial to consistently remind them of this feature and improve the methods for facilitating chat deletion.
- Platform-specific privacy concerns: Participants expressed diverse opinions regarding the security of different messaging platforms used in conjunction with SafeHaven (P3, P10, P11, P13, P14, P19, P22, P30, P33). A notable point of discussion was the perceived security of Telegram (P3, P19, P22). The platform's robust encryption protocols were regarded as more secure by the participants, which provided better protection of their communications and



personal information. On the other hand, there were significant concerns regarding the use of Facebook Messenger among the other participants (P10, P13, P14, P19, P22, P30, P33) because it could lead to data breaches or inadvertent sharing of sensitive information. Most of such concerns were associated with Facebook's data handling practice. Given the platform's extensive reach and its historical privacy issues, participants were worried about how their data might be used or potentially exposed. P22 said, *“Their issues might be that they steal data. So I wasn't really thinking of that when using the chatbot. Because it's, it's like a message. It's one on one. So, I'm not thinking of privacy. But in the back of my mind, knowing that it was Facebook, I guess I was thinking that maybe my information will be stored.”*

- The risks of hacking, data breaches, and surveillance
  - Risk of hacking and data breaches: Due to the sensitive nature of information typically shared in DV situations, some participants expressed concerns over security breaches and hacking (P14, P20, P21, P22, P27). Specifically, they were concerned about unauthorized individuals (i.e., abusers) who could gain access to their personal conversations by hacking or stealing their usernames, emails or passwords, and other sensitive data, leading to further victimization or security compromise. P14 commented, *“Public concern was you know how people can hack into your Facebook account, for example? Or wondering if they could, by any chance? Since you did send touch for use messenger with a chatbot? Where if somebody ends up, you know, packet into your Facebook account itself, will they be*

*able to see what you have in the past? And then that gives that information against you on your social media?”* Similarly, P27 thought it safe to use the CA but had some worries about the possible leakage of data via emails. P27 said, *“I think overall, I do think it's very, like safe to use. I was just asking about the email just in case like if someone's worried about their information getting leaked, but I think that's pretty good overall.”* To address such concerns, a robust approach to security, featuring two-factor authentication and end-to-end encryption, as well as secure data storage, must be implemented on these online platforms.

- **Location tracking and real-time monitoring:** A further concern raised by participants P13, P32 and P36 revolves around the possibility of tracking location and monitoring real-time internet connections. It is possible that location-based services, whether activated intentionally or inadvertently, could reveal the location of the survivor, potentially placing them at risk. P13 noted, *“it asked me to share my location. Because if your computer or your cell phone doesn't have location services enabled, and then like it pops that up, it could accidentally open it for like, all of the services, again, are like all of your apps again. I also didn't like that it took you to a new window to share your location.”* In addition, real-time monitoring of an internet connection poses a significant threat to the safety and privacy of a survivor. Unauthorized individuals (e.g., abusers) may be able to intercept and read communication as it occurs, potentially exposing survivors to increased risk and undermining the confidentiality of such

scenarios. P32 commented, *“I’m not sure if this is this would be the case or something if someone was pulling on my connection, while I’m using the chatbot, and would get information that I’m sending immediately to the chatbot.”*

- Privacy and security concerns of search engines

The use of search engines has raised concerns about the privacy of participants seeking information or assistance of DV. According to the responses from participants, we identified a broad spectrum of concerns, particularly those related to search history, disclosure of personal information, compromised websites, targeted ads, IP tracking, and cookies. The following are major concerns highlighted by participants.

- Exposure of personal information and search history: The apprehension regarding the exposure of personal information and search history was a prevalent concern among a significant number of participants (P1, P2, P5, P6, P9, P14, P17, P18, P25, P29, P30, P33), highlighting the potential risks associated with the use of search engines. They were concerned that their search engine activities, including all links clicked and search terms entered, could easily be discovered by their abusers, thereby complicating the matter and putting them at greater risk. For instance, P6 stated, *“I feel like there would be a lot more information about me, given doing web searches, as well, as like I said, it’s easy to see what somebody has been searching on the internet, like as far as websites from your service provider.”* P9 also mentioned, *“it would just see every link that you clicked on, and every search term that you put on there would be accessible to someone if they wanted to*

*see it.*” The fear goes beyond just entering search terms; it extends to the associated data, such as a survivor's location and email account, which may also be compromised. P14 said, *“I will say when you try and find like a shelter, so they'll probably ask for your location if I allow it or not. Or if I say if I do go search. And I'm still connected to my email.”* In spite of the prevailing fear expressed by several participants regarding the discovery of their search histories, one participant did not perceive search engines as a threat to privacy. There was a sense of control expressed by P13 regarding search histories, especially when using features such as incognito tabs, suggesting that some users believe they are capable of managing their privacy effectively while searching online. Overall, despite the variety of attitudes toward search engine privacy, ranging from confidence in personal management strategies to deep-rooted fears of exposure, there is a prevailing fear regarding the need for greater safeguards when pursuing sensitive personal information, such as DV assistance, online. Online tracking and privacy concerns are compounding this underlying fear.

- Cookies, IP tracking, and shared device risks: The concerns surrounding cookies, IP tracking, and the use of shared devices, as highlighted by some participants. P2, P9, P10, P14, P17, P25 and P29, point to a deep-rooted anxiety about the vulnerability of personal data in the digital realm. These participants expressed unease over how search engines and websites use cookies and IP tracking, which not only track their online activities but also potentially expose their search preferences and history across multiple devices

connected to the same Wi-Fi network. P2 clarified, *“Also having cookies or if you search something on your computer, then whatever you search for might come up on a different device just because it's you're using the same Wi-Fi so that would be a concern.”* Some participants expressed a similar concern that search results might be visible to others when using a search engine on a public computer if the browsing history has not been cleared (P17, P29). P17 mentioned, *“Sometimes if we use public computers, then the sudden our search history might be available to other people, if they opens a search engine, and as you can see, it has a list of the items that we have searched.”* This kind of surveillance may pose a privacy concern due to the possibility of unintentionally sharing sensitive information via the same device or network. An additional concern is the possibility of being tracked by one's IP address. Individuals seeking information about sensitive topics such as DV should be particularly concerned about this matter (P10). It was brought to P14's attention that cookies are frequently mandated on certain websites; although they may seem innocuous, they have the ability to not only track and record user activity but also introduce a risk of malicious software or viruses being downloaded onto computers. To summarize, these concerns mirror a prevalent unease regarding the technological mechanisms associated with digital monitoring and surveillance, particularly when using search engines, emphasizing the need for greater privacy protection.

- Online security threats and compromised websites: Fear surrounding online security threats was a significant issue raised by participants P5, P21, and P33,

illustrating a complex web of digital vulnerabilities. A range of potential online risks were articulated, starting with the concern that personal information could be compromised. In addition to the fear of the exposure of search data, such breaches may compromise more sensitive personal information of a survivor (P5). There is also the risk of encountering compromised or fake websites that can be used by malicious actors to steal information such as credentials or personal information (P21). Moreover, P33 highlighted the issue of privacy breaches and cyberstalking, especially when logged into personal accounts. In her response to the concerns using search engines for DV assistance, P33 said, *“I would say if you're like signed into, like your Google account, and then just if things get like hacked, as well, that information can become known. And then the search history is still available unless you go in and like learn how to clear all of that. And with the cyberstalking aspect of it, too.”* As observed, this concern is twofold: one is the fear that personal accounts may be hacked, leading to further exposure of sensitive information, and the other is the risk of being cyberstalked based on search activities. Due to the participants' concerns, increased awareness and strong security measures are needed regarding online security threats and compromised websites, particularly for those seeking sensitive information like DV.

- Targeted advertising and misuse of data: A growing concern about the invasive nature of digital tracking, and its implications for privacy and security, is evident in participants P22, P25, and P26 concerns related to targeted advertising and possible misuse of data. In particular, P22 and P25 drew

attention to the uncomfortable reality that one is being tracked through browser advertisements and trackers, where search histories are not just kept, but actively used for advertising purposes. P22 mentioned, *“For the search engine, the concerns are mainly just having information like so into ads or so it's similar problem, I guess, to the chatbot. But I was more concerned about the search engine than the chatbot because the search engine kept giving results that were ads.”* P25 also refer to the same point in her response, *“the ads on browsers that can track your searching and stuff, the things that will track you because if you are using, like, you know, if a violent person got on your device and saw like the history, the search history, like you know, ads suggested for you and stuff like that, and then it could cause issues.”* Despite its seemingly harmless nature, this practice may take on a more sinister tone for a survivor discussing sensitive topics, like assistance with DV. Another viewpoint is offered by P26, who raises concerns regarding the social implications of disclosing search data. There is a fear that personal stories and searches may be exploited, judged, or misinterpreted by others. P26 explained, *“I believe that sometimes it was a using search engines are a little bit more exploit. I feel like sharing information there. might bring some judgment. Besides that, the information that I read there, it's, I don't trust them.”* DV survivors may suffer unintended and possibly harmful consequences as a result of data tracking and targeted advertising, as indicated by these voices. Therefore, digital privacy practices and measures should be adopted.

- Privacy and security concerns of ChatGPT

The participants hold a diverse range of opinions on ChatGPT's privacy and security in the context of DV. Most of the concerns centered around potential risks associated with linking email addresses to user accounts, the lack of transparency regarding the handling of user data, and inadequate safety measures. Conversely, many participants voiced no privacy concerns, attributing their trust to ChatGPT's anonymity, as it does not typically require any personal information from its users. It is possible that their confidence is based on their trust in the platform's security measures, their lack of awareness of potential risks, or their belief that communications on ChatGPT do not involve the exchange of sensitive information. Based on the range of perceptions, the privacy concerns of users and perceived security measures of digital platforms such as ChatGPT are intricately intertwined. Below, we will discuss key concerns that the participants raised.

- Association of email and account identity: ChatGPT has raised privacy concerns among some participants, particularly when it comes to the association of email addresses (P3, P11, P23, P24, P27, P32, P35, P36). For example, P3, P27 and P36 pointed out the risk of personal identification, noting that logging into ChatGPT with an email address could associate the account with a specific individual. For instance, P36 commented about her concerns with ChatGPT, *“just making sure that my identity and my family's identity would be secure while using it. Not be tracked.”* This poses a serious risk in DV situations in which anonymity is paramount. Even though deleting search history may mitigate this risk, personal identification information remains a concern (P27). Moreover, P11, P32, and P35 highlighted the threat of email hacking. A possibility has been raised that an abuser could hack into



an email account and obtain sensitive information, such as a safety plan detailed in a chat message. P11 stated, *“I think it would just be like, I had to log into that right with using. I think for me, it would be if my abuser, whatever, or whenever he eventually will hack an email account, would he be able to access the ChatGPT account, and then once you have all of that information, especially with this safety plan, he'll be able to read my safety plan and know a play by play what I'm going to be doing and where I'm going to be in that safety plan. And that as a domestic violence survivor, safety plans are not easily changed.”* Similarly, P32 noted that ChatGPT records all interactions associated with the user's email account, so sensitive conversations (i.e., DV related conversations) could be recorded. These records may be accessible if someone were to hack this account unless the user deletes their chat history proactively. In the same context, concerns have been raised about the traceability of search history (P12, P23, P24, P27). Even with precautions, the possibility that one's search history, especially regarding sensitive topics like DV, could be traced or accessed remains a significant concern. P27 explained, *“It's just the fact that using an email, when searching that it can like show up on your search history, but I guess you can delete it.”* They exhibit a significant fear among survivors regarding privacy when using ChatGPT, suggesting that it is essential to implement robust privacy measures like end-to-end encryption to secure conversations, recommending a survivor to delete conversations and use anonymous account, implementing strong user authentication methods, and regularly updating and patching security software.

- Lack of clarity and transparency in data usage: Some participants in the study voiced worries about the lack of clarity and transparency regarding how ChatGPT collects, stores, and uses data (P8, P16). In particular, P8 expressed concern regarding the unclear purposes for storing personal information and was concerned about whether personal information might be misused or sold to third parties, highlighting a general mistrust and uncertainty regarding data handling. P16 also perceived a lack of transparency and clarity in ChatGPT's privacy and security policies. Such ambiguity may severely undermine the trust of DV survivors in the platform, a sentiment that is likely heightened when the platform is used to discuss sensitive topics like DV. P16 said, *“So I wasn't really thinking about the privacy concerns when I was using ChatGPT. But after using the chatbot, and seeing how upfront they were about privacy concerns and everything, it makes me wonder how, how safe or you know, private conversations are with ChatGPT just because they are not as upfront about it. If I was currently a victim going through something, I would probably not trust ChatGPT as much just because they're not what's the word? You know, they're not relieving my fears right off the bat, whereas the chatbot does.”* The overall trend of these concerns indicates that platforms like ChatGPT need to communicate clearly and explicitly regarding their data usage policies and practice, so that a survivor feels confident in the privacy and security of their data.
- Lack of safety measures: ChatGPT's design presents a critical limitation with regard to the safety of DV survivors, which was flagged by one participant. A

specific reference was made to the absence of safety measures during a conversation with ChatGPT. According to P23, the lack of immediate exit strategies or the ability to rapidly redirect users to neutral websites compromises survivors' safety and security. As part of her response to ChatGPT's privacy concerns, P23 stated, *“The fact that there was like in the beginning of the chatbot conversation, it gave immediate, like, exit strategies for how to exit the website, or just, you know, no longer show the website anymore. It said, it even said it, would redirect to a neutral website. ChatGPT does not have those options. So in that, in that way, I feel a little less secure and less tailored for victims and survivors of domestic abuse.”* To address the concern, more thoughtful, user-centered design is necessary for such a platform to ensure the immediate safety and peace of mind of DV survivors.

## 6 DISCUSSION AND CONCLUSION

The primary objective of this dissertation was to conceptualize, develop, and evaluate a CA, specifically designed to provide comprehensive support to DV survivors. It demonstrated the potential of technology-driven interventions to assist survivors in addressing their diverse needs. SafeHaven serves as a promising tool for DV intervention by providing informational guidance, emotional comfort, and practical assistance. The tool has proven its ability to function as an invaluable resource for DV survivors, offering assistance and empowerment through features such as personalized feedback, empathetic conversational interactions, safety planning modules, and resource referrals.

In this chapter, we delve into the critical examination of the findings, elucidating the implications and significance of SafeHaven in providing support to DV survivors. It examines the nuanced interactions between the CA and empathy, the CA-based support system, anonymity and trustworthiness. We also provide a discussion of some of the technological and societal implications of our study, as well as its limitations and future research.

### 6.1 Main Findings

In this dissertation, the primary objective was to investigate the role and effectiveness of technology-based interventions in supporting DV survivors, with a specific focus on CAs. The growing prevalence of technology and advances in mobile and web-based applications present a unique opportunity to provide essential support for survivors of DV. As detailed in the research, technological interventions are becoming more accepted due to their accessibility, confidentiality, and ability to provide real-time, personalized assistance (RQ1). For educational and therapeutic purposes, a variety of digital platforms have been developed, such as mobile

apps, web-based tools, and CAs. Different features have been incorporated into such technologies as safety planning, risk assessment, emotional support, and research has proved their effectiveness in the well-being of survivors.

In this dissertation, the key findings are centered on meta-requirements, design principles, and evaluation of SafeHaven (RQ2-RQ4). Expert consultations led to the identification of essential design requirements, including informational, emotional and instrumental support, active listening, safety advice, resource referrals and building trust (RQ2). As part of the research, opportunities were identified that could be provided by CAs, such as continuous availability, impartiality, and anonymity, whereas challenges that include concerns about privacy, usability, and provision of adequate emotional support should be taken into consideration. In light of the expert consultations, SafeHaven has been designed with specific features to address these needs, including modules for greeting and safety assurance, abuse identification, safety planning, sharing survivor stories, and providing referrals to available resources (RQ3, RQ4).

Based on an evaluation of SafeHaven, its enhanced functionality was demonstrated over conventional search engines and ChatGPT (RQ5). Compared to traditional search engines and ChatGPT, the CA system demonstrated superior performance as a means of providing emotional, informational, and instrumental support. SafeHaven has also demonstrated a strong ability to maintain user anonymity and build trust. It was found that SafeHaven's evaluation, though positive in many aspects, also revealed several areas in which it needs to be improved, including difficulties in formulating queries, delays in responding, and navigational issues. Overall, the results of this dissertation confirm the efficacy of CA as a tailored support tool for survivors of DV, thereby establishing a benchmark for future CA-based interventions in this domain.

### 6.1.1 Emotional Support

Designed specifically to meet the needs of DV survivors, SafeHaven serves as a solution and source of emotional support, exemplifying an effective use of technology with empathy. Participants felt SafeHaven as having a profound understanding of the emotional turmoil and psychological challenges they faced, unlike conventional search engines or CAs (e.g., ChatGPT). Participants perceived the interactive, compassionate, and empathetic interface of SafeHaven as a close resemblance to human interaction. Upon further analysis, participants perceived a higher level of emotional support from SafeHaven than from search engines or ChatGPT regardless of their gender. In addition, younger users were significantly more appreciative of SafeHaven's emotional support than older users. A younger generation may be more open to new forms of technology and innovative ways to receive support, such as AI-driven emotional support, than an older generation, who may prefer conventional communication methods.

A few participants in our study acknowledged that SafeHaven could not completely replace human empathy. Considering DV's traumatic experience, SafeHaven's design must strike a balance between automated empathy and human empathy while offering emotional support. DV survivors often require emotional support in addition to help-seeking [280]. Thus, when SafeHaven is unable to offer adequate comfort, it should transfer the responsibility of providing emotional support to human professionals.

Further, none of the participants experienced emotional distress during interactions with SafeHaven's emotional support messages. Some participants said that more affirmation was needed during the conversation (e.g., at the beginning, middle, and end of the conversation). Thus, even if the CA's interactions are predominately informational, its capacity for emotional support should not be overshadowed. One participant suggested that after providing information

(e.g., a safety plan), the CA could validate the survivor's situation through empathetic acknowledgments, such as expressing sympathy for their experiences.

### 6.1.2 Usability

SafeHaven as well as search engines and ChatGPT scored above the benchmark for usability. Compared to traditional search engines, SafeHaven's usability was identical. Compared to ChatGPT, SafeHaven's usability was slightly higher despite no significant difference. In addition, only males show a marginal preference for SafeHaven over search engines and ChatGPT. However, SafeHaven did not significantly differ from the other tools being evaluated when analyzed across age groups.

Following a thorough analysis of the usability scale, we found that SafeHaven was favored for its ease of use, better integration of functions, and consistency. In contrast, a search engine was perceived to require less initial learning and instill more confidence in users. In the same vein, in the usability comparison between SafeHaven and ChatGPT, SUS scores indicate that both platforms exhibit distinct strengths. Specifically, SafeHaven was favored due to its lower complexity, higher level of integration of functions, and consistency, all of which may suggest an easier and more intuitive user experience. Participants also felt more confident and needed less technical support. As far as ease of use, and the learning curve for ChatGPT, they were comparable to those of SafeHaven.

We provide several explanations for the above findings. First, SafeHaven meets modern user expectations by offering a familiar and intuitive experience. Secondly, SafeHaven follows integrated design principles by combining the straightforwardness and efficiency of a search engine with ChatGPT's conversational design. Those characteristics enhance the usability of the CA by enabling easy information access as well as a more usable experience. Further, as users

become more familiar with various digital platforms and more tech-savvy, they are more capable of navigating these platforms easily.

### 6.1.3 Anonymity and Trustworthiness

Anonymity and trustworthiness are foundational to SafeHaven's design principles, which directly address the privacy and safety concerns frequently cited by DV survivors. This study provides strong empirical evidence for the strengths of SafeHaven in terms of anonymity and trust. The design of anonymity in SafeHaven greatly lessened the participants' concerns about being judged or stigmatized. Moreover, the participants rated the level of anonymity higher for SafeHaven than for a search engine. This could be attributed to the lack of mechanisms provided by search engines to ensure anonymity and security [353]. Search engines often track user activity, store search histories, and personalize results, which compromise the anonymity of users, particularly DV survivors. Furthermore, ChatGPT was perceived to be comparable to SafeHaven in terms of anonymity. The similarity could be attributed to the common nature of these platforms. In both instances, participants engage with a CA, an environment where physical identity is less prominent, leading participants to feel disassociated from their real-world identity.

Similarly, SafeHaven was perceived as significantly more trustworthy than a search engine and ChatGPT. This finding can be attributed to several design features of SafeHaven. First, it focuses specifically on the context and concern regarding DV, participants may perceive it as safer, more private, and more empathic. The findings suggest that technologies designed with a focus on the specific needs of their target audiences are more likely to be trusted. Second, SafeHaven provides enhanced privacy and security features that contribute to its higher trust ratings, especially when compared to generic platforms such as ChatGPT and search engines.



For instance, SafeHaven builds in privacy features (such as exiting a chat room, deleting or hiding a chat, and storing data securely) and anonymity options, and even the overall design of SafeHaven has the goal of protecting user privacy and providing a safe forum for discussing sensitive issues. Third, SafeHaven's ability to tailor conversations and responses to meet the needs of the people involved could also affect user trust. SafeHaven helped individuals in creating personalized safety plans suited to their specific situations. Moreover, SafeHaven can provide personalized assistance based on the user's specific location, needs, and situation.

A critical insight emerged regarding the capabilities of CAs during an emergency. According to the survey, less than 20% of participants believe that the current CAs can respond to emergencies efficiently, indicating significant reservations about the reliability, decision-making, and rapid response capabilities of these systems under high-pressure conditions. There is a gap in existing CA systems, which indicates the need for robust CA technologies that can respond effectively in urgent situations, especially for critical applications such as DV interventions in which timely response is vital.

## 6.2 Research Contributions

This dissertation research contributes to the IS, HCI and DV literature, which can be categorized in three folds.

First, we identify the requirements and principles for designing a CA for DV survivors. Moreover, we identify the opportunities for CAs to provide instrumental, informational, and emotional support tailored to the needs of DV survivors, determine the major challenges of their utilizing CAs and offer suggestions to overcome these challenges. These requirements and principles include a broad range of capabilities and features, providing appropriate support to DV survivors. In particular, the CA can be enhanced with design features in terms of conversational

elements, language, support, and trust. CA has the potential to offer informational, emotional, and instrumental support, serve as a good starting point for survivors, provide an anonymous space, and be always available and accessible. Additional design implications are provided concerning the development of CA-based intervention of DV.

Second, we detail the CA design process with the goal of inspiring researchers and designers to develop systems to assist DV survivors. In particular, we outline the flow of the proposed CA-based system and delineate each phase a user may go through during their conversation with the CA. Furthermore, we identify essential features and functions of an application aimed at aiding DV survivors in addressing their needs. Additionally, we operationalize the design guidelines with a prototype system, detailing its major modules and sub-modules, including assuring safety, screening and identifying abuse, developing a safety plan, and providing resources. These design and development processes of the CA can guide practitioners and researchers in creating such systems to address the needs of DV survivors.

Third, this is the first research that evaluates the effectiveness of CAs in assisting individuals affected by DV. The evaluation focuses on whether CA can help DV survivors manage the negative impacts of DV through various forms of support, including emotional, informational, instrumental support in a convenient and efficient manner within the context of individuals' lives. Additionally, we conducted a comparison between the CA and both search engines and ChatGPT to determine the most suitable platform for seeking assistance, along with the strengths and weaknesses of each technology. Our study revealed that SafeHaven provided nuanced support customized to the unique needs of DV survivors, surpassing the effectiveness of conventional search engines and emerging ChatGPT. This analysis not only highlights the specialized functions of SafeHaven but also sheds light on the strengths and weaknesses of

different technological approaches in facilitating DV intervention. Such insights will be instrumental in guiding the development of future CA-driven support systems for DV survivors.

### 6.3 Implications

The findings of this dissertation have implications for a wide range of stakeholders, including CA developers, academic researchers, DV organizations and support groups. For CA developers, this dissertation holds considerable significance for the conceptualization and implementation of CA-based interventions of DV, thereby contributing to the advancement of CA design and implementation in a specialized area. Incorporating our derived DPs into the development of CAs helps improve the engagement and support of DV survivors. Considering factors such as language preferences, emotional states, and individual needs can enhance the relevance and effectiveness of CA interventions. Our DPs can help guide the conception, development, and implementation of CAs that are empathetic and responsive, capable of responding to crisis situations and providing appropriate support. The tone and style of the conversation should be adjusted as needed to deliver tailored interventions to survivors. In addition to mimicking typical in-person conversations in DV interventions, CAs can also provide innovative interventions, such as crisis interventions or real-time detection of suicidal thoughts, which may not be achieved through traditional interaction with the same human experts or workers. For instance, it is worthwhile for CA developers to incorporate a feature of detecting changes in mood, language, and behavior that indicate the urgent need for immediate assistance by extending state-of-the-art machine learning techniques. CA developers might also leverage large language models (LLMs) such as GPT that enable more natural and dynamic conversations to better suit the users' needs.

For DV organizations and support groups, the findings of this dissertation can be utilized to offer survivors additional sources of information, empathy, anonymity, and trustworthiness as

forms of support. Building on the insights gained from the research, organizations and groups may integrate CAs to augment traditional intervention methods. By adopting proactive and technologically advanced approaches, support services can become more inclusive and accessible to survivors who might otherwise feel marginalized, fearful, or unable to access conventional services [22].

The dissertation holds great social significance, especially in raising public awareness about and enhancing community understanding of DV. CA can serve as a valuable initial resource for survivors seeking to enhance their understanding of their circumstances. As an information repository, CA can provide access to educational resources and support groups for survivors as well as those who suspect they are survivors to assist them in understanding abuse dynamics and identifying signs of unhealthy relationships. In addition, by facilitating connections with support groups and services, CAs like SafeHaven have the potential to enhance social connectivity and support networks [354]. This, in turn, enables survivors to cultivate stronger social support networks. This is especially valuable for individuals who face barriers to accessing related support, such as language barriers, cultural stigma, and resource shortages.

In addition to DV intervention, our study may have implications for mental health and other sensitive issues. The derived principles, combined with the evaluation results of SafeHaven, suggest that the principles of anonymity, accessibility, and empathetic interaction can be applied to other relevant contexts (e.g., mental health and substance abuse.). In view that stigma, privacy concerns, and accessibility pose common obstacles across these domains [355], the design of CA has the potential to overcome these obstacles, thereby bringing about positive changes to the lives of the most vulnerable. For instance, individuals with depression or anxiety need a safe, stigma-free environment to address their mental health issues. Similarly, the nonjudgmental

support and resource guidance provided by CAs can empower individuals who experience sexual assault or substance abuse to access the assistance that they need under those situations [29], [356].

#### 6.4 Limitations and Future Work

This dissertation shows several limitations that merit consideration for future research. First, we recruited DV domain experts in our study that focused on identifying the meta requirements and design principles of the CA for survivors. Future research may consider engaging DV survivors directly by adopting survivor-centered, trauma-informed approaches [357]. Our sample size was relatively small but comparable to other interview studies [229], [252], and was considered sufficient given that the primary goal of this study is to identify design principles and CA capabilities.

Second, SafeHaven was evaluated about its short-term impact and single-session use, which may not fully represent all significant outcomes or shed light on long-term changes in behavior or continued improvements in the user experience and well-being. Therefore, conducting long-term studies on the impact of CA-based interventions over a longer period of time can provide a more comprehensive understanding of their effectiveness.

Third, in view of several technical limitations of SafeHaven, including navigation, timeliness of responses, topic expansion, conversation depth and language processing issues, as participants pointed out in this study, future improvements could include refining the interface of SafeHaven to facilitate easier navigation and improving timeliness of responses and information retrieval. It is suggested that incorporating advanced machine learning techniques, such as transformer models (e.g., BERT) and LLMs, would enhance the responsiveness and output accuracy of SafeHaven. With these sophisticated techniques, the CA can navigate between modules or

submodules effectively and generate more accurate outputs, particularly in critical functions such as screening and abuse identification.

Fourth, future research could involve conducting a more comprehensive comparison between our system and other CA-based interventions designed to assist DV survivors. While we have compared our system with general tools like search engines and ChatGPT, a direct comparison with similar CA-based interventions could provide more in-depth insights. However, it is important to note that existing systems remain scarce. Among the few systems that were available, they either had very limited functionality, were still in the design stage and not yet implemented, or have been implemented in contexts outside the U.S., all of which could impact the outcomes of such comparisons.

Fifth, despite the fact that some measurement tools have proven effective in assessing certain aspects of technology use, they may not keep up with the rapid pace of technological advancement. For example, even though there are valuable insights that can be gained from the Jian trust scale [341] in terms of trust and distrust in polar opposite directions on a single dimension, a measurement tool such as this may not fully reflect technology's rapid advancement. As technology evolves, factors that influence user trust and distrust also change. Considering this, future research should take this into account by utilizing or adapting the latest measurement tools that are more capable of responding to technological advances. In the same context, other measures, (e.g., Technology Acceptance Model), might also be helpful in assessing the perceptions of participants about the technology that they use in a more comprehensive way.

## 6.5 Conclusion

This dissertation explored the potential of CAs in supporting DV survivors, offering insights into the requirements, design principles, implementation, and evaluation of such type of technology. Following the design requirements and principles, the development and deployment of SafeHaven underscores the feasibility of adopting advanced technological solutions to offer emotional, informational, and instrumental support to DV survivors. In addition, the evaluation results from the user study demonstrate that CA can serve as a non-judgmental, anonymous, and confidential platform for survivors. They also suggest the need to further enhance language processing techniques and the depth of interaction and navigation. The findings of this dissertation provide a theoretical framework and practical guide for CA researchers and developers to advance support systems for DV survivors.

## REFERENCES

- [1] M. C. Black *et al.*, *The National Intimate Partner and Sexual Violence Survey (NISVS): 2010 Summary Report*. The National Center for Injury Prevention and Control of the Centers for Disease Control and Prevention., 2011. [Online]. Available: [https://www.cdc.gov/violenceprevention/pdf/nisvs\\_report2010-a.pdf](https://www.cdc.gov/violenceprevention/pdf/nisvs_report2010-a.pdf)
- [2] M. D. Pagelow and L. W. Pagelow, *Family Violence*. Praeger, 1984.
- [3] I. H. Frieze and A. Browne, "Violence in Marriage," *Crime Justice*, vol. 11, pp. 163–218, Jan. 1989, doi: 10.1086/449154.
- [4] S. D. HERZBERGER, *VIOLENCE WITHIN THE FAMILY: social psychological perspectives*. S.l.: ROUTLEDGE, 2019.
- [5] S. E. Ullman and L. Peter-Hagene, "SOCIAL REACTIONS TO SEXUAL ASSAULT DISCLOSURE, COPING, PERCEIVED CONTROL, AND PTSD SYMPTOMS IN SEXUAL ASSAULT VICTIMS: Social Reactions and PTSD," *J. Community Psychol.*, vol. 42, no. 4, pp. 495–508, May 2014, doi: 10.1002/jcop.21624.
- [6] S. E. Ullman, S. M. Townsend, H. H. Filipas, and L. L. Starzynski, "Structural Models of the Relations of Assault Severity, Social Support, Avoidance Coping, Self-Blame, and PTSD Among Sexual Assault Survivors," *Psychol. Women Q.*, vol. 31, no. 1, pp. 23–37, Mar. 2007, doi: 10.1111/j.1471-6402.2007.00328.x.
- [7] M. R. Sable, F. Danis, D. L. Mauzy, and S. K. Gallagher, "Barriers to reporting sexual assault for women and men: perspectives of college students," *J. Am. Coll. Health J ACH*, vol. 55, no. 3, pp. 157–162, Dec. 2006, doi: 10.3200/JACH.55.3.157-162.
- [8] N. Montalvo-Liendo, "Cross-cultural factors in disclosure of intimate partner violence: an integrated review," *J. Adv. Nurs.*, vol. 65, no. 1, pp. 20–34, Jan. 2009, doi: 10.1111/j.1365-2648.2008.04850.x.
- [9] M. Lutenbacher, A. Cohen, and J. Mitzel, "Do We Really Help? Perspectives of Abused Women," *Public Health Nurs.*, vol. 20, no. 1, pp. 56–64, Jan. 2003, doi: 10.1046/j.1525-1446.2003.20108.x.
- [10] N. El-Bassel, L. Gilbert, V. Rajah, A. Foleno, and V. Frye, "Social support among women in methadone treatment who experience partner violence: Isolation and male controlling behavior," *Violence Women*, vol. 7, no. 3, pp. 246–274, 2001.
- [11] J. A. Forte, D. D. Franks, J. Forte, and D. Rigsby, "Asymmetrical role-taking: Comparing battered and nonbattered women," *Soc. Work*, vol. 41, no. 1, pp. 59–73, 1996.
- [12] R. E. Mitchell and C. A. Hodson, "Coping with domestic violence: Social support and psychological health among battered women," *Am. J. Community Psychol.*, vol. 11, no. 6, pp. 629–654, 1983.
- [13] E. Hilberman and K. Munson, "Sixty battered women.," *Victimology*, 1977.
- [14] H.-F. Lin, J. L. Postmus, H. Hu, and A. M. Stylianou, "IPV Experiences and Financial Strain Over Time: Insights from the Blinder-Oaxaca Decomposition Analysis," *J. Fam. Econ. Issues*, May 2022, doi: 10.1007/s10834-022-09847-y.
- [15] E. Stark, *Coercive control: The entrapment of women in personal life*. Oxford University Press, 2009.
- [16] C. K. Baker, S. L. Cook, and F. H. Norris, "Domestic violence and housing problems: A contextual analysis of women's help-seeking, received informal support, and formal system response," *Violence Women*, vol. 9, no. 7, pp. 754–783, 2003.



- [17] A. M. Alquaiz, M. Almuneef, A. Kazi, and A. Almeneessier, "Social determinants of domestic violence among Saudi married women in Riyadh, Kingdom of Saudi Arabia," *J. Interpers. Violence*, vol. 36, no. 3–4, pp. NP1561–1585NP, 2021.
- [18] J. Lee, E. C. Pomeroy, and T. M. Bohman, "Intimate partner violence and psychological health in a sample of Asian and Caucasian women: The roles of social support and coping," *J. Fam. Violence*, vol. 22, no. 8, pp. 709–720, 2007.
- [19] S. R. Robinson, K. Ravi, and R. J. Voth Schrag, "A Systematic Review of Barriers to Formal Help Seeking for Adult Survivors of IPV in the United States, 2005–2019," *Trauma Violence Abuse*, vol. 22, no. 5, pp. 1279–1295, Dec. 2021, doi: 10.1177/1524838020916254.
- [20] P. A. Crane and R. E. Constantino, "Use of the Interpersonal Support Evaluation List (ISEL) to Guide Intervention Development with Women Experiencing Abuse," *Issues Ment. Health Nurs.*, vol. 24, no. 5, pp. 523–541, Jan. 2003, doi: 10.1080/01612840305286.
- [21] F. Esposito, "Self-help groups with women victims of intimate partner violence".
- [22] C. Emezue, "Digital or Digitally Delivered Responses to Domestic and Intimate Partner Violence During COVID-19," *JMIR Public Health Surveill.*, vol. 6, no. 3, p. e19831, Jul. 2020, doi: 10.2196/19831.
- [23] J. Alhusen, T. Bloom, A. Clough, and N. Glass, "Development of the MyPlan Safety Decision App with Friends of College Women in Abusive Dating Relationships," *J. Technol. Hum. Serv.*, vol. 33, no. 3, pp. 263–282, Jul. 2015, doi: 10.1080/15228835.2015.1037414.
- [24] K. J. Debnam and T. Kumodzi, "Adolescent Perceptions of an Interactive Mobile Application to Respond to Teen Dating Violence," *J. Interpers. Violence*, vol. 36, no. 13–14, pp. 6821–6837, Jul. 2021, doi: 10.1177/0886260518821455.
- [25] J. Koziol-McLain *et al.*, "Efficacy of a Web-Based Safety Decision Aid for Women Experiencing Intimate Partner Violence: Randomized Controlled Trial," *J. Med. Internet Res.*, vol. 19, no. 12, p. e426, Jan. 2018, doi: 10.2196/jmir.8617.
- [26] N. E. Glass *et al.*, "The Longitudinal Impact of an Internet Safety Decision Aid for Abused Women," *Am. J. Prev. Med.*, vol. 52, no. 5, pp. 606–615, May 2017, doi: 10.1016/j.amepre.2016.12.014.
- [27] J. Suler, "The Online Disinhibition Effect," *Cyberpsychol. Behav.*, vol. 7, no. 3, pp. 321–326, Jun. 2004, doi: 10.1089/1094931041291295.
- [28] N. Andalibi, O. L. Haimson, M. De Choudhury, and A. Forte, "Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose California USA: ACM, May 2016, pp. 3906–3918. doi: 10.1145/2858036.2858096.
- [29] H. Park and J. Lee, "Designing a Conversational Agent for Sexual Assault Survivors: Defining Burden of Self-Disclosure and Envisioning Survivor-Centered Solutions," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, May 2021, pp. 1–17. doi: 10.1145/3411764.3445133.
- [30] T. Bauer, E. Devrim, M. Glazunov, W. L. Jaramillo, B. Mohan, and G. Spanakis, "#MeTooMaastricht: Building a chatbot to assist survivors of sexual harassment," *ArXiv190902809 Cs*, Sep. 2019, Accessed: Sep. 06, 2021. [Online]. Available: <http://arxiv.org/abs/1909.02809>
- [31] S. Rennick-Egglestone, S. Knowles, G. Toms, P. Bee, K. Lovell, and P. Bower, "Health Technologies 'In the Wild': Experiences of Engagement with Computerised CBT," in

- Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose California USA: ACM, May 2016, pp. 2124–2135. doi: 10.1145/2858036.2858128.
- [32] G. M. Lucas *et al.*, “Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers,” *Front. Robot. AI*, vol. 4, p. 51, Oct. 2017, doi: 10.3389/frobt.2017.00051.
  - [33] H. Rhee, J. Allen, J. Mammen, and M. Swift, “Mobile phone-based asthma self-management aid for adolescents (mASMAA): a feasibility study,” *Patient Prefer. Adherence*, p. 63, Jan. 2014, doi: 10.2147/PPA.S53504.
  - [34] G. M. Lucas, J. Gratch, A. King, and L.-P. Morency, “It’s only a computer: Virtual humans increase willingness to disclose,” *Comput. Hum. Behav.*, vol. 37, pp. 94–100, Aug. 2014, doi: 10.1016/j.chb.2014.04.043.
  - [35] A. N. Joinson and C. B. Paine, *Self-disclosure, Privacy and the Internet*, vol. 1. Oxford University Press, 2012. doi: 10.1093/oxfordhb/9780199561803.013.0016.
  - [36] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, “Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial,” *JMIR Ment. Health*, vol. 4, no. 2, p. e19, Jun. 2017, doi: 10.2196/mental.7785.
  - [37] A. Chandrashekar, “Companion Chat.” 2019. Accessed: Sep. 09, 2021. [Online]. Available: <http://www.anjalic.com/companion-chat>.
  - [38] W. Maeng and J. Lee, “Designing and Evaluating a Chatbot for Survivors of Image-Based Sexual Abuse,” in *CHI Conference on Human Factors in Computing Systems*, New Orleans LA USA: ACM, Apr. 2022, pp. 1–21. doi: 10.1145/3491102.3517629.
  - [39] “Hello Cass Chatbot.” [Online]. Available: <https://hellocass.com.au/>
  - [40] AI for Good, “rAInbow.” 2018. Accessed: Sep. 08, 2021. [Online]. Available: <https://www.hirainbow.org/>
  - [41] Md. E. Hossain, A. U. Najib, and Md. Z. Islam, “Combating Domestic Violence during COVID-19 Pandemic in Bangladesh: Using a Mobile Application integrated with an Effective Solution,” in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, DHAKA, Bangladesh: IEEE, Dec. 2020, pp. 1–6. doi: 10.1109/ICCIT51783.2020.9392691.
  - [42] J. Ren, T. Bickmore, M. Hempstead, and B. Jack, “Birth Control, Drug Abuse, or Domestic Violence? What Health Risk Topics Are Women Willing to Discuss with a Virtual Agent?,” in *Intelligent Virtual Agents*, vol. 8637, T. Bickmore, S. Marsella, and C. Sidner, Eds., in Lecture Notes in Computer Science, vol. 8637. , Cham: Springer International Publishing, 2014, pp. 350–359. doi: 10.1007/978-3-319-09767-1\_46.
  - [43] Kona Connect, “Sophia – The Chatbot.” 2022. Accessed: Oct. 06, 2021. [Online]. Available: <https://kona-connect.org/sophia/>
  - [44] “JAEL.AI.” [Online]. Available: <http://jael.ai/>
  - [45] “Position statement on domestic violence.” Australian Medical Association, 1998.
  - [46] L. F. Lowenstein, “Domestic Violence Recent Research (2004–2005) Part I: Incidence, Types and Causes,” *Police J. Theory Pract. Princ.*, vol. 78, no. 2, pp. 147–157, Jun. 2005, doi: 10.1350/pojo.2005.78.2.147.
  - [47] M. J. Breiding, J. Chen, and M. C. Black, *Intimate partner violence in the United States–2010*. The National Center for Injury Prevention and Control of the Centers for Disease Control and Prevention., 2014.

- [48] M. Breiding, K. C. Basile, S. G. Smith, M. C. Black, and R. R. Mahendra, "Intimate partner violence surveillance: uniform definitions and recommended data elements. Version 2.0," 2015, [Online]. Available: <https://stacks.cdc.gov/view/cdc/31292>
- [49] J. C. Campbell, "Women's responses to sexual abuse in intimate relationships," *Health Care Women Int.*, vol. 10, no. 4, pp. 335–346, 1989.
- [50] C. R. Hanneke, N. M. Shields, and G. J. McCall, "Assessing the prevalence of marital rape," *J. Interpers. Violence*, vol. 1, no. 3, pp. 350–362, 1986.
- [51] R. K. Bergen, *Wife rape*, vol. 2. Sage, 1996.
- [52] C. Adams, "I just raped my wife! What are you going to do about it, pastor," *Transform. Rape Cult.*, pp. 57–86, 1993.
- [53] I. H. Frieze, "Investigating the causes and consequences of marital rape," *Signs J. Women Cult. Soc.*, vol. 8, no. 3, pp. 532–553, 1983.
- [54] A. E. Adams, C. M. Sullivan, D. Bybee, and M. R. Greeson, "Development of the Scale of Economic Abuse," *Violence Women*, vol. 14, no. 5, pp. 563–588, May 2008, doi: 10.1177/1077801208315529.
- [55] B. H. Spitzberg and W. R. Cupach, "The state of the art of stalking: Taking stock of the emerging literature," *Aggress. Violent Behav.*, vol. 12, no. 1, pp. 64–86, Jan. 2007, doi: 10.1016/j.avb.2006.05.001.
- [56] S. G. Smith *et al.*, "The national intimate partner and sexual violence survey: 2015 data brief—updated release," 2018.
- [57] W. E. Goodlin and C. S. Dunn, "Three Patterns of Domestic Violence in Households: Single Victimization, Repeat Victimization, and Co-occurring Victimization," *J. Fam. Violence*, vol. 25, no. 2, pp. 107–122, Feb. 2010, doi: 10.1007/s10896-009-9274-z.
- [58] M. Mele, "VICTIM/OFFENDER CONTACT AND REPEAT DOMESTIC VIOLENCE VICTIMIZATION," *J. Crime Justice*, vol. 29, no. 2, pp. 51–68, Jan. 2006, doi: 10.1080/0735648X.2006.9721648.
- [59] M. P. Thompson *et al.*, "Partner Violence, Social Support, and Distress Among Inner-City African American Women," *Am. J. Community Psychol.*, vol. 28, no. 1, pp. 127–143, Feb. 2000, doi: 10.1023/A:1005198514704.
- [60] S. Cobb, "Social Support as a Moderator of Life Stress:," *Psychosom. Med.*, vol. 38, no. 5, pp. 300–314, Sep. 1976, doi: 10.1097/00006842-197609000-00003.
- [61] J. C. Chang *et al.*, "Health care interventions for intimate partner violence: What women want," *Womens Health Issues*, vol. 15, no. 1, pp. 21–30, Jan. 2005, doi: 10.1016/j.whi.2004.08.007.
- [62] K. Greene, V. J. Derlega, and A. Mathews, "Self-Disclosure in Personal Relationships," in *The Cambridge Handbook of Personal Relationships*, A. L. Vangelisti and D. Perlman, Eds., Cambridge: Cambridge University Press, 2006, pp. 409–428. doi: 10.1017/CBO9780511606632.023.
- [63] S. R. Chaudoir and J. D. Fisher, "The disclosure processes model: Understanding disclosure decision making and postdisclosure outcomes among people living with a concealable stigmatized identity.," *Psychol. Bull.*, vol. 136, no. 2, pp. 236–256, 2010, doi: 10.1037/a0018193.
- [64] N. Andalibi, O. L. Haimson, M. D. Choudhury, and A. Forte, "Social Support, Reciprocity, and Anonymity in Responses to Sexual Abuse Disclosures on Social Media," *ACM Trans. Comput.-Hum. Interact.*, vol. 25, no. 5, pp. 1–35, Oct. 2018, doi: 10.1145/3234942.

- [65] Y. Mou and K. Xu, "The media inequality: Comparing the initial human-human and human-AI social interactions," *Comput. Hum. Behav.*, vol. 72, pp. 432–440, Jul. 2017, doi: 10.1016/j.chb.2017.02.067.
- [66] O'Campo, P., Ahmed, F., and Cyriac, A., "The role of health care professionals in preventing and intervening with intimate partner violence," in *J. Keeling & T. Mason (Eds.), Domestic Violence. A multi-professional approach for healthcare practitioners*, McGraw Hill, pp. 107–115.
- [67] M. Rodríguez, J. M. Valentine, J. B. Son, and M. Muhammad, "Intimate Partner Violence and Barriers To Mental Health Care for Ethnically Diverse Populations of Women," *Trauma Violence Abuse*, vol. 10, no. 4, pp. 358–374, Oct. 2009, doi: 10.1177/1524838009339756.
- [68] S. B. Plichta and M. Falik, "Prevalence of violence and its implications for women's health," *Womens Health Issues*, vol. 11, no. 3, pp. 244–258, May 2001, doi: 10.1016/S1049-3867(01)00085-8.
- [69] M. A. Rodríguez, E. McLoughlin, G. Nah, and J. C. Campbell, "Mandatory Reporting of Domestic Violence Injuries to the Police: What Do Emergency Department Patients Think?," *JAMA*, vol. 286, no. 5, p. 580, Aug. 2001, doi: 10.1001/jama.286.5.580.
- [70] K. E. Edin and U. Högberg, "Violence against pregnant women will remain hidden as long as no direct questions are asked," *Midwifery*, vol. 18, no. 4, pp. 268–278, Dec. 2002, doi: 10.1054/midw.2002.0318.
- [71] K. M. Edwards, C. M. Dardis, and C. A. Gidycz, "Women's disclosure of dating violence: A mixed methodological study," *Fem. Psychol.*, vol. 22, no. 4, pp. 507–517, Nov. 2012, doi: 10.1177/0959353511422280.
- [72] R. L. Heron and M. C. Eisma, "Barriers and facilitators of disclosing domestic violence to the healthcare service: A systematic review of qualitative research," *Health Soc. Care Community*, vol. 29, no. 3, pp. 612–630, May 2021, doi: 10.1111/hsc.13282.
- [73] L. Bates, L. Hancock, and D. Peterkin, "'A little encouragement': health services and domestic violence," *Int. J. Health Care Qual. Assur.*, vol. 14, no. 2, pp. 49–56, Apr. 2001, doi: 10.1108/09526860110386474.
- [74] K. Cann, S. Withnell, J. Shakespeare, H. Doll, and J. Thomas, "Intimate partner violence: A comparative survey of levels of detection, knowledge, and attitudes in healthcare workers," *Public Health*, vol. 115, no. 2, pp. 89–95, Mar. 2001, doi: 10.1038/sj.ph.1900749.
- [75] J. M. Spangaro, A. B. Zwi, and R. G. Poulos, "'Persist. persist.': A qualitative study of women's decisions to disclose and their perceptions of the impact of routine screening for intimate partner violence," *Psychol. Violence*, vol. 1, no. 2, pp. 150–162, 2011, doi: 10.1037/a0023136.
- [76] T. A. Battaglia, E. Finley, and J. M. Liebschutz, "Survivors of intimate partner violence speak out: Trust in the patient-provider relationship," *J. Gen. Intern. Med.*, vol. 18, no. 8, pp. 617–623, Aug. 2003, doi: 10.1046/j.1525-1497.2003.21013.x.
- [77] C. Kaukinen, "The Help-Seeking Strategies of Female Violent-Crime Victims: The Direct and Conditional Effects of Race and the Victim-Offender Relationship," *J. Interpers. Violence*, vol. 19, no. 9, pp. 967–990, Sep. 2004, doi: 10.1177/0886260504268000.
- [78] K. M. Sylaska and K. M. Edwards, "Disclosure of Intimate Partner Violence to Informal Social Support Network Members: A Review of the Literature," *Trauma Violence Abuse*, vol. 15, no. 1, pp. 3–21, Jan. 2014, doi: 10.1177/1524838013496335.

- [79] D. Houry *et al.*, “Differences in Female and Male Victims and Perpetrators of Partner Violence With Respect to WEB Scores,” *J. Interpers. Violence*, vol. 23, no. 8, pp. 1041–1055, Aug. 2008, doi: 10.1177/0886260507313969.
- [80] B. J. Barrett and M. St. Pierre, “Variations in Women’s Help Seeking in Response to Intimate Partner Violence: Findings From a Canadian Population-Based Study,” *Violence Women*, vol. 17, no. 1, pp. 47–70, Jan. 2011, doi: 10.1177/1077801210394273.
- [81] J. Liebschutz, T. Battaglia, E. Finley, and T. Averbuch, “Disclosing intimate partner violence to health care clinicians - What a difference the setting makes: A qualitative study,” *BMC Public Health*, vol. 8, no. 1, p. 229, Dec. 2008, doi: 10.1186/1471-2458-8-229.
- [82] P. Niolon *et al.*, *Preventing Intimate Partner Violence Across the Lifespan: A Technical Package of Programs, Policies, and Practices*. Division of Violence Prevention, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, 2017.
- [83] D. L. Ansara and M. J. Hindin, “Formal and informal help-seeking associated with women’s and men’s experiences of intimate partner violence in Canada,” *Soc. Sci. Med.*, vol. 70, no. 7, pp. 1011–1018, Apr. 2010, doi: 10.1016/j.socscimed.2009.12.009.
- [84] R. Moors and R. Webber, “The dance of disclosure: Online self-disclosure of sexual assault,” *Qual. Soc. Work*, vol. 12, no. 6, pp. 799–815, Nov. 2013, doi: 10.1177/1473325012464383.
- [85] K. Hegarty *et al.*, “Protocol for a randomised controlled trial of a web-based healthy relationship tool and safety decision aid for women experiencing domestic violence (I-DECIDE),” *BMC Public Health*, vol. 15, no. 1, pp. 1–9, 2015.
- [86] N. Hussain, S. Sprague, K. Madden, F. N. Hussain, B. Pindiprolu, and M. Bhandari, “A Comparison of the Types of Screening Tool Administration Methods Used for the Detection of Intimate Partner Violence: A Systematic Review and Meta-Analysis,” *Trauma Violence Abuse*, vol. 16, no. 1, pp. 60–69, Jan. 2015, doi: 10.1177/1524838013515759.
- [87] L. Tarzia, D. Iyer, E. Thrower, and K. Hegarty, “‘Technology doesn’t judge you’: young Australian women’s views on using the internet and smartphones to address intimate partner violence,” *J. Technol. Hum. Serv.*, vol. 35, no. 3, pp. 199–218, 2017.
- [88] R. E. Constantino *et al.*, “Comparing Online with Face-to-Face HELPP Intervention in Women Experiencing Intimate Partner Violence,” *Issues Ment. Health Nurs.*, vol. 36, no. 6, pp. 430–438, Jun. 2015, doi: 10.3109/01612840.2014.991049.
- [89] Y. Albalawi and J. Sixsmith, “Agenda Setting for Health Promotion: Exploring an Adapted Model for the Social Media Era,” *JMIR Public Health Surveill.*, vol. 1, no. 2, p. e21, Nov. 2015, doi: 10.2196/publichealth.5014.
- [90] B. M. Leiner *et al.*, “A brief history of the internet,” *ACM SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 5, pp. 22–31, Oct. 2009, doi: 10.1145/1629607.1629613.
- [91] B. Lichtenstein and I. M. Johnson, “Older African American Women and Barriers to Reporting Domestic Violence to Law Enforcement in the Rural Deep South,” *Women Crim. Justice*, vol. 19, no. 4, pp. 286–305, Oct. 2009, doi: 10.1080/08974450903224329.
- [92] R. B. Felson, S. F. Messner, A. W. Hoskin, and G. Deane, “REASONS FOR REPORTING AND NOT REPORTING DOMESTIC VIOLENCE TO THE POLICE\*,” *Criminology*, vol. 40, no. 3, pp. 617–648, Aug. 2002, doi: 10.1111/j.1745-9125.2002.tb00968.x.
- [93] M. N. Vranda, C. N. Kumar, D. Muralidhar, N. Janardhana, and P. T. Sivakumar, “Barriers to Disclosure of Intimate Partner Violence among Female Patients Availing

- Services at Tertiary Care Psychiatric Hospitals: A Qualitative Study,” *J. Neurosci. Rural Pract.*, vol. 09, no. 03, pp. 326–330, Jul. 2018, doi: 10.4103/jnrp.jnrp\_14\_18.
- [94] J. D. Cravens, J. B. Whiting, and R. O. Aamar, “Why I Stayed/Left: An Analysis of Voices of Intimate Partner Violence on Social Media,” *Contemp. Fam. Ther.*, vol. 37, no. 4, pp. 372–385, Dec. 2015, doi: 10.1007/s10591-015-9360-8.
- [95] H. L. McCauley, A. E. Bonomi, M. K. Maas, K. W. Bogen, and T. L. O’Malley, “#MaybeHeDoesntHitYou: Social Media Underscore the Realities of Intimate Partner Violence,” *J. Womens Health*, vol. 27, no. 7, pp. 885–891, Jul. 2018, doi: 10.1089/jwh.2017.6560.
- [96] S. Subramani, H. Wang, M. R. Islam, A. Ulhaq, and M. O’Connor, “Child Abuse and Domestic Abuse: Content and Feature Analysis from Social Media Disclosures,” in *Databases Theory and Applications*, vol. 10837, J. Wang, G. Cong, J. Chen, and J. Qi, Eds., in Lecture Notes in Computer Science, vol. 10837, Cham: Springer International Publishing, 2018, pp. 174–185. doi: 10.1007/978-3-319-92013-9\_14.
- [97] S. Subramani, H. Wang, H. Q. Vu, and G. Li, “Domestic Violence Crisis Identification From Facebook Posts Based on Deep Learning,” *IEEE Access*, vol. 6, pp. 54075–54085, 2018, doi: 10.1109/ACCESS.2018.2871446.
- [98] K. E. Carlyle, J. P. D. Guidry, S. A. Dougherty, and C. W. Burton, “Intimate Partner Violence on Instagram: Visualizing a Public Health Approach to Prevention,” *Health Educ. Behav.*, vol. 46, no. 2\_suppl, pp. 90S–96S, Dec. 2019, doi: 10.1177/1090198119873917.
- [99] S. Subramani, S. Michalska, H. Wang, J. Du, Y. Zhang, and H. Shakeel, “Deep Learning for Multi-Class Identification From Domestic Violence Online Posts,” *IEEE Access*, vol. 7, pp. 46210–46224, 2019, doi: 10.1109/ACCESS.2019.2908827.
- [100] S. Subramani, H. Q. Vu, and H. Wang, “Intent Classification using Feature Sets for Domestic Violence Discourse on Social Media,” *ArXiv180403497 Cs*, Feb. 2018, Accessed: Sep. 29, 2021. [Online]. Available: <http://arxiv.org/abs/1804.03497>
- [101] R. Alaggia and S. Wang, “‘I never told anyone until the #metoo movement’: What can we learn from sexual abuse and sexual assault disclosures made through social media?,” *Child Abuse Negl.*, vol. 103, p. 104312, May 2020, doi: 10.1016/j.chiabu.2019.104312.
- [102] K. W. Bogen, K. Bleiweiss, and L. M. Orchowski, “Sexual violence is #NotOkay: Social reactions to disclosures of sexual victimization on twitter.,” *Psychol. Violence*, vol. 9, no. 1, pp. 127–137, Jan. 2019, doi: 10.1037/vio0000192.
- [103] A. Bandura, “Social cognitive theory: An agentic perspective,” *Annu. Rev. Psychol.*, vol. 52, no. 1, pp. 1–26, 2001.
- [104] J. Lawson, “Sociological Theories of Intimate Partner Violence,” *J. Hum. Behav. Soc. Environ.*, vol. 22, no. 5, pp. 572–590, Jun. 2012, doi: 10.1080/10911359.2011.598748.
- [105] H. Al-Alosi, “Fighting fire with fire: Exploring the potential of technology to help victims combat intimate partner violence,” *Aggress. Violent Behav.*, vol. 52, p. 101376, May 2020, doi: 10.1016/j.avb.2020.101376.
- [106] C. Emezue, J. D. Chase, T. Udmuangpia, and T. L. Bloom, “Technology-based and digital interventions for intimate partner violence: A systematic review and meta-analysis,” *Campbell Syst. Rev.*, vol. 18, no. 3, Sep. 2022, doi: 10.1002/cl2.1271.
- [107] C. Hassija and M. J. Gray, “The Effectiveness and Feasibility of Videoconferencing Technology to Provide Evidence-Based Treatment to Rural Domestic Violence and Sexual Assault Populations,” *Telemed. E-Health*, vol. 17, no. 4, pp. 309–315, May 2011, doi: 10.1089/tmj.2010.0147.

- [108] P. Klasnja and W. Pratt, "Healthcare in the pocket: Mapping the space of mobile-phone health interventions," *J. Biomed. Inform.*, vol. 45, no. 1, pp. 184–198, Feb. 2012, doi: 10.1016/j.jbi.2011.08.017.
- [109] N. Glass, K. B. Eden, T. Bloom, and N. Perrin, "Computerized Aid Improves Safety Decision Process for Survivors of Intimate Partner Violence," *J. Interpers. Violence*, vol. 25, no. 11, pp. 1947–1964, Nov. 2010, doi: 10.1177/0886260509354508.
- [110] K. B. Eden *et al.*, "Use of Online Safety Decision Aid by Abused Women," *Am. J. Prev. Med.*, vol. 48, no. 4, pp. 372–383, Apr. 2015, doi: 10.1016/j.amepre.2014.09.027.
- [111] A. M. Akinsulure-Smith, T. Chu, E. Keatley, and A. Rasmussen, "Intimate Partner Violence among West African Immigrants," *J. Aggress. Maltreatment Trauma*, vol. 22, no. 2, pp. 109–126, Feb. 2013, doi: 10.1080/10926771.2013.719592.
- [112] H. Bui and M. Morash, "Immigration, Masculinity, and Intimate Partner Violence From the Standpoint of Domestic Violence Service Providers and Vietnamese-Origin Women," *Fem. Criminol.*, vol. 3, no. 3, pp. 191–215, Jul. 2008, doi: 10.1177/1557085108321500.
- [113] D. Freed, J. Palmer, D. E. Minchala, K. Levy, T. Ristenpart, and N. Dell, "Digital Technologies and Intimate Partner Violence: A Qualitative Analysis with Multiple Stakeholders," *Proc. ACM Hum.-Comput. Interact.*, vol. 1, no. CSCW, pp. 1–22, Dec. 2017, doi: 10.1145/3134681.
- [114] J. Finn and M. Banach, "Victimization Online: The Downside of Seeking Human Services for Women on the Internet," *Cyberpsychol. Behav.*, vol. 3, no. 5, pp. 785–796, Oct. 2000, doi: 10.1089/10949310050191764.
- [115] Statista, "Number of smartphone users in the United States from 2018 to 2025." [Online]. Available: <https://www.statista.com/statistics/201182/forecast-of-smartphone-users-in-the-us/>.
- [116] Statista, "Number of internet users in the United States from 2010 to 2025." [Online]. Available: <https://www.statista.com/statistics/325645/usa-number-of-internet-users/>
- [117] E. K. Choo, C. Zlotnick, D. R. Strong, D. D. Squires, C. Tapé, and M. J. Mello, "BSAFER: A Web-based intervention for drug use and intimate partner violence demonstrates feasibility and acceptability among women in the emergency department," *Subst. Abuse*, vol. 37, no. 3, pp. 441–449, Jul. 2016, doi: 10.1080/08897077.2015.1134755.
- [118] B. Palanisamy, S. Sensenig, J. Joshi, and R. Constantino, "LEAF: A privacy-conscious social network-based intervention tool for IPV survivors," in *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014)*, Redwood City, CA, USA: IEEE, Aug. 2014, pp. 138–146. doi: 10.1109/IRI.2014.7051883.
- [119] M. Ford-Gilboe *et al.*, "Longitudinal impacts of an online safety and health intervention for women experiencing intimate partner violence: randomized controlled trial," *BMC Public Health*, vol. 20, no. 1, p. 260, Dec. 2020, doi: 10.1186/s12889-020-8152-8.
- [120] M. I. Ragavan, V. Ferre, and M. Bair-Merritt, "Thrive: A Novel Health Education Mobile Application for Mothers Who Have Experienced Intimate Partner Violence," *Health Promot. Pract.*, vol. 21, no. 2, pp. 160–164, Mar. 2020, doi: 10.1177/1524839919890870.
- [121] K. Hegarty *et al.*, "An online healthy relationship tool and safety decision aid for women experiencing intimate partner violence (I-DECIDE): a randomised controlled trial," *Lancet Public Health*, vol. 4, no. 6, pp. e301–e310, Jun. 2019, doi: 10.1016/S2468-2667(19)30079-9.
- [122] Pew Research Center. (2021), "Mobile Fact Sheet." Apr. 2021. [Online]. Available: <https://www.pewresearch.org/internet/fact-sheet/mobile/>

- [123] R. Miller, C. Z.-W. Qiang, M. Yamamichi, and V. Hausman, "Mobile applications for the health sector," 2012.
- [124] A. Masoumzadeh, L. Jin, J. Joshi, and R. Constantino, "HELPP Zone: Towards protecting college students from dating violence," 2013.
- [125] M. Lindsay *et al.*, "Survivor Feedback on a Safety Decision Aid Smartphone Application for College-Age Women in Abusive Relationships," *J. Technol. Hum. Serv.*, vol. 31, no. 4, pp. 368–388, Oct. 2013, doi: 10.1080/15228835.2013.861784.
- [126] "Nursing in Tele-Mental Health," *J. Psychosoc. Nurs. Ment. Health Serv.*, vol. 44, no. 5, pp. 20–27, May 2006, doi: 10.3928/02793695-20060501-05.
- [127] R. Constantino, P. A. Crane, B. S. Noll, W. M. Doswell, and B. Braxter, "Exploring the feasibility of email-mediated interaction in survivors of abuse," *J. Psychiatr. Ment. Health Nurs.*, vol. 14, no. 3, pp. 291–301, May 2007, doi: 10.1111/j.1365-2850.2007.01080.x.
- [128] P. Crane and R. Constantino, "Is email interaction feasible for intervention with women and children exposed to domestic violence? A literature review," *Adv. Pract. Nurs.*, vol. 3, pp. 1–16, 2003.
- [129] A. Afdal *et al.*, "Increasing Life Satisfaction of Domestic Violence Victims through the Role of Supporting Group Therapy on Social Media," presented at the the International Conference on Education Technology (ICoET 2019), in *Advances in Social Science, Education and Humanities Research*, vol. Proceedings of the International Conference on Education Technology (ICoET 2019). Atlantis press, 2019. [Online]. Available: <https://www.atlantis-press.com/proceedings/icoet-19/125925070>
- [130] T. O'Neill, "'Today I Speak': Exploring How Victim-Survivors Use Reddit," *Int. J. Crime Justice Soc. Democr.*, vol. 7, no. 1, pp. 44–59, Mar. 2018, doi: 10.5204/ijcjsd.v7i1.402.
- [131] M. Wood, E. Rose, and C. Thompson, "Viral justice? Online justice-seeking, intimate partner violence and affective contagion," *Theor. Criminol.*, vol. 23, no. 3, pp. 375–393, Aug. 2019, doi: 10.1177/1362480617750507.
- [132] A. Powell, "Seeking rape justice: Formal and informal responses to sexual violence through technosocial counter-publics," *Theor. Criminol.*, vol. 19, no. 4, pp. 571–588, Nov. 2015, doi: 10.1177/1362480615576271.
- [133] "A Best Practice: Using Social Media for Sexual Violence Prevention." Prevention and Education Committee of the Task Force., 2014.
- [134] T. H. Chu, Y. Su, H. Kong, J. Shi, and X. Wang, "Online Social Support for Intimate Partner Violence Victims in China: Quantitative and Automatic Content Analysis," *Violence Women*, vol. 27, no. 3–4, pp. 339–358, Mar. 2021, doi: 10.1177/1077801220911452.
- [135] B. N. Davies, "Online Dialogue in Response to Disclosures of Intimate Partner Violence," Brigham Young University, 2020.
- [136] J. Whiting, R. Dansby Olufowote, J. Cravens-Pickens, and A. Banford Witting, "Online Blaming and Intimate Partner Violence: A Content Analysis of Social Media Comments," *Qual. Rep.*, Jan. 2019, doi: 10.46743/2160-3715/2019.3486.
- [137] M. Sivagurunathan, D. M. Walton, T. Packham, R. Booth, and J. MacDermid, "'Punched in the Balls': Male Intimate Partner Violence Disclosures and Replies on Reddit," *Am. J. Mens Health*, vol. 15, no. 4, p. 155798832110396, Jul. 2021, doi: 10.1177/15579883211039666.
- [138] O. L. Haimson, J. R. Brubaker, L. Dombrowski, and G. R. Hayes, "Disclosure, Stress, and Support During Gender Transition on Facebook," in *Proceedings of the 18th ACM*



- Conference on Computer Supported Cooperative Work & Social Computing*, Vancouver BC Canada: ACM, Feb. 2015, pp. 1176–1190. doi: 10.1145/2675133.2675152.
- [139] T. Matthews *et al.*, “Stories from Survivors: Privacy & Security Practices when Coping with Intimate Partner Abuse,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver Colorado USA: ACM, May 2017, pp. 2189–2201. doi: 10.1145/3025453.3025875.
- [140] S. Park *et al.*, “Designing a Chatbot for a Brief Motivational Interview on Stress Management: Qualitative Case Study,” *J. Med. Internet Res.*, vol. 21, no. 4, p. e12231, Apr. 2019, doi: 10.2196/12231.
- [141] J. Weizenbaum, “ELIZA—a computer program for the study of natural language communication between man and machine,” *Commun. ACM*, vol. 9, no. 1, pp. 36–45, Jan. 1966, doi: 10.1145/365153.365168.
- [142] K. Del Valle, “Conversational commerce: A new opportunity for card payments. MasterCard.” Jan. 2018. [Online]. Available: <https://newsroom.mastercard.com/documents/conversational-commerce-a-new-opportunity-for-card-payments>
- [143] “Chatbot.” Oxford Living Dictionaries., 1990. [Online]. Available: <https://en.oxforddictionaries.com/definition/chatbot>
- [144] M. McTear, Z. Callejas, and D. Griol, *The Conversational Interface: Talking to Smart Devices*, 1st ed. 2016. Cham: Springer International Publishing: Imprint: Springer, 2016. doi: 10.1007/978-3-319-32967-3.
- [145] B. KINSELLA, “U.S. Smart Speaker Users Rise to 57 Million.” Voicebot.ai, Oct. 2018. [Online]. Available: <https://voicebot.ai/2018/10/30/u-s-smart-speaker-users-rise-to-57-million/>
- [146] “Gartner Says Worldwide Spending on VPA-Enabled Wireless Speakers Will Top 3.5 Billion dollar by 2021.” Gartner, Aug. 2017. [Online]. Available: <https://gtnr.it/2AfUFOx>
- [147] E. Luger and A. Sellen, “‘Like Having a Really Bad PA’: The Gulf between User Expectation and Experience of Conversational Agents,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose California USA: ACM, May 2016, pp. 5286–5297. doi: 10.1145/2858036.2858288.
- [148] S. Hwang, “Would satisfaction with smart speakers transfer into loyalty towards the smart speaker provider?,” 2018.
- [149] “Smart Speakers 2018: World Market Forecast to 2023 - Key Vendors Covered include Alphabet, Amazon, Harman Intl, Alibaba and Sonos.” Cision PR Newswire, 2018. [Online]. Available: <https://www.prnewswire.com/news-releases/>
- [150] E. Maggio, “Apple says that 500 million customers use Siri.” Business Insider., Jan. 2018. [Online]. Available: <http://www.businessinsider.com/apple-says-siri-has-500-million-users-2018-1>
- [151] B. Inkster, S. Sarda, and V. Subramanian, “An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study,” *JMIR MHealth UHealth*, vol. 6, no. 11, p. e12106, Nov. 2018, doi: 10.2196/12106.
- [152] M. Chatzimina, L. Koumakis, K. Marias, and M. Tsiknakis, “Employing Conversational Agents in Palliative Care: A Feasibility Study and Preliminary Assessment,” in *2019 IEEE*

- 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, Athens, Greece: IEEE, Oct. 2019, pp. 489–496. doi: 10.1109/BIBE.2019.00095.
- [153] A. Cheng, V. Raghavaraju, J. Kanugo, Y. P. Handrianto, and Y. Shang, “Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes,” in *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, Las Vegas, NV: IEEE, Jan. 2018, pp. 1–5. doi: 10.1109/CCNC.2018.8319283.
- [154] A. van Heerden, X. Ntinga, and K. Vilakazi, “The potential of conversational agents to provide a rapid HIV counseling and testing services,” in *2017 International Conference on the Frontiers and Advances in Data Science (FADS)*, Xi’an: IEEE, Oct. 2017, pp. 80–85. doi: 10.1109/FADS.2017.8253198.
- [155] B. Liu and S. S. Sundar, “Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot,” *Cyberpsychology Behav. Soc. Netw.*, vol. 21, no. 10, pp. 625–636, Oct. 2018, doi: 10.1089/cyber.2018.0110.
- [156] T. Giorgino *et al.*, “Automated spoken dialogue system for hypertensive patient home management,” *Int. J. Med. Inf.*, vol. 74, no. 2–4, pp. 159–167, Mar. 2005, doi: 10.1016/j.ijmedinf.2004.04.026.
- [157] H. Wang, Q. Zhang, M. Ip, and J. T. Fai Lau, “Social Media-based Conversational Agents for Health Management and Interventions,” *Computer*, vol. 51, no. 8, pp. 26–33, Aug. 2018, doi: 10.1109/MC.2018.3191249.
- [158] D. Ireland *et al.*, “Hello Harlie: Enabling Speech Monitoring Through Chat-Bot Conversations,” *Stud. Health Technol. Inform.*, vol. 227, pp. 55–60, 2016.
- [159] C.-Y. Huang, M.-C. Yang, C.-Y. Huang, Y.-J. Chen, M.-L. Wu, and K.-W. Chen, “A Chatbot-supported Smart Wireless Interactive Healthcare System for Weight Control and Health Promotion,” in *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Bangkok: IEEE, Dec. 2018, pp. 1791–1795. doi: 10.1109/IEEM.2018.8607399.
- [160] M. Souza, T. Miyagawa, P. Melo, and F. Maciel, “Wellness Programs: Wearable Technologies Supporting Healthy Habits and Corporate Costs Reduction,” in *HCI International 2017 – Posters’ Extended Abstracts*, vol. 714, C. Stephanidis, Ed., in Communications in Computer and Information Science, vol. 714, Cham: Springer International Publishing, 2017, pp. 293–300. doi: 10.1007/978-3-319-58753-0\_44.
- [161] D. Jurafsky and J. H. Martin, *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*, 2nd ed. in Prentice Hall series in artificial intelligence. Upper Saddle River, N.J: Pearson Prentice Hall, 2009.
- [162] A. F. van Woudenberg, “A Chatbot Dialogue Manager Chatbots and Dialogue Systems: A Hybrid Approach,” 2014.
- [163] U. Gnewuch, S. Morana, and A. Maedche, “Towards Designing Cooperative and Social Conversational Agents for Customer Service,” in *Proceedings of the International Conference on Information Systems (ICIS)*, 2017.
- [164] M. F. McTear, “Spoken dialogue technology: enabling the conversational user interface,” *ACM Comput. Surv.*, vol. 34, no. 1, pp. 90–169, Mar. 2002, doi: 10.1145/505282.505285.
- [165] S. Arora, K. Batra, and S. Singh, “Dialogue System: A Brief Review,” *ArXiv13064134 Cs*, Jun. 2013, Accessed: Sep. 16, 2021. [Online]. Available: <http://arxiv.org/abs/1306.4134>

- [166] N. M. Radziwill and M. C. Benton, "Evaluating Quality of Chatbots and Intelligent Conversational Agents," *ArXiv170404579 Cs*, Apr. 2017, Accessed: Sep. 03, 2021. [Online]. Available: <http://arxiv.org/abs/1704.04579>
- [167] OpenAI, "ChatGPT: Optimizing Language Models for Dialogue," Dec. 2022. [Online]. Available: <https://openai.com/blog/chatgpt>
- [168] C. Macdonald, D. Adeloje, A. Sheikh, and I. Rudan, "Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis," *J. Glob. Health*, vol. 13, p. 01003, Feb. 2023, doi: 10.7189/jogh.13.01003.
- [169] P. P. Ray, "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," *Internet Things Cyber-Phys. Syst.*, vol. 3, pp. 121–154, 2023, doi: 10.1016/j.iotcps.2023.04.003.
- [170] Sung Kim, "Prompt Engineering with OpenAI GPT-3 API: A Real-World Example," Jan. 2023. [Online]. Available: <https://blog.devgenius.io/prompt-engineering-with-openapi-gpt-by-example-23765ae0085e>
- [171] Y. Fan and X. Luo, "A Survey of Dialogue System Evaluation," in *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*, Baltimore, MD, USA: IEEE, Nov. 2020, pp. 1202–1209. doi: 10.1109/ICTAI50040.2020.00182.
- [172] Z. Lipton, X. Li, J. Gao, L. Li, F. Ahmed, and L. Deng, "BBQ-Networks: Efficient Exploration in Deep Reinforcement Learning for Task-Oriented Dialogue Systems," *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, Apr. 2018, [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/11946>
- [173] N. Asghar, P. Poupart, J. Hoey, X. Jiang, and L. Mou, "Affective Neural Response Generation," in *Advances in Information Retrieval*, vol. 10772, G. Pasi, B. Piwowarski, L. Azzopardi, and A. Hanbury, Eds., in *Lecture Notes in Computer Science*, vol. 10772, Cham: Springer International Publishing, 2018, pp. 154–166. doi: 10.1007/978-3-319-76941-7\_12.
- [174] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 2002, pp. 311–318.
- [175] S. Banerjee and A. Lavie, "METEOR: An automatic metric for MT evaluation with improved correlation with human judgments," in *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, 2005, pp. 65–72.
- [176] L. C. Rouge, "A package for automatic evaluation of summaries," in *Proceedings of Workshop on Text Summarization of ACL, Spain*, 2004.
- [177] J. Nielsen, "Usability inspection methods," in *Conference companion on Human factors in computing systems*, 1994, pp. 413–414.
- [178] S. Divya, Indumathi, S. Ishwarya, M. Priyasankari, and S. Kalpanadevi, "A Self-Diagnosis Medical Chatbot Using Artificial Intelligence," 2018.
- [179] A. Freedy, E. DeVisser, G. Weltman, and N. Coeyman, "Measurement of trust in human-robot collaboration," in *2007 International Symposium on Collaborative Technologies and Systems*, Orlando, FL: IEEE, May 2007, pp. 106–114. doi: 10.1109/CTS.2007.4621745.
- [180] B. F. Yuksel, P. Collisson, and M. Czerwinski, "Brains or Beauty: How to Engender Trust in User-Agent Interactions," *ACM Trans. Internet Technol.*, vol. 17, no. 1, pp. 1–20, Mar. 2017, doi: 10.1145/2998572.

- [181] J. E. Fox, "The effects of information accuracy on user trust and compliance," in *Conference companion on Human factors in computing systems common ground - CHI '96*, Vancouver, British Columbia, Canada: ACM Press, 1996, pp. 35–36. doi: 10.1145/257089.257119.
- [182] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman, "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 53, no. 5, pp. 517–527, Oct. 2011, doi: 10.1177/0018720811417254.
- [183] M. Rheu, J. Y. Shin, W. Peng, and J. Huh-Yoo, "Systematic Review: Trust-Building Factors and Implications for Conversational Agent Design," *Int. J. Human-Computer Interact.*, vol. 37, no. 1, pp. 81–96, Jan. 2021, doi: 10.1080/10447318.2020.1807710.
- [184] J. M. Kraus, F. Nothdurft, P. Hock, D. Scholz, W. Minker, and M. Baumann, "Human After All: Effects of Mere Presence and Social Interaction of a Humanoid Robot as a Co-Driver in Automated Driving," in *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Ann Arbor MI USA: ACM, Oct. 2016, pp. 129–134. doi: 10.1145/3004323.3004338.
- [185] R. van den Brule, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, and P. Haselager, "Do Robot Performance and Behavioral Style affect Human Trust?: A Multi-Method Approach," *Int. J. Soc. Robot.*, vol. 6, no. 4, pp. 519–531, Nov. 2014, doi: 10.1007/s12369-014-0231-5.
- [186] M. Skjuve and P. B. Brandtzæg, "Chatbots as a new user interface for providing health information to young people," 2018.
- [187] Y.-C. Lee, N. Yamashita, Y. Huang, and W. Fu, "'I Hear You, I Feel You': Encouraging Deep Self-disclosure through a Chatbot," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–12. doi: 10.1145/3313831.3376175.
- [188] M. D. Pickard, C. A. Roster, and Y. Chen, "Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions?," *Comput. Hum. Behav.*, vol. 65, pp. 23–30, Dec. 2016, doi: 10.1016/j.chb.2016.08.004.
- [189] J. Jang and J. Kim, "Healthier Life with Digital Companions: Effects of Reflection-Level and Statement-Type of Messages on Behavior Change via a Perceived Companion," *Int. J. Human-Computer Interact.*, vol. 36, no. 2, pp. 172–189, Jan. 2020, doi: 10.1080/10447318.2019.1615722.
- [190] A. Ravichander and A. W. Black, "An Empirical Study of Self-Disclosure in Spoken Dialogue Systems," in *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, Melbourne, Australia: Association for Computational Linguistics, 2018, pp. 253–263. doi: 10.18653/v1/W18-5030.
- [191] S. Kim, J. Lee, and G. Gweon, "Comparing Data from Chatbot and Web Surveys: Effects of Platform and Conversational Style on Survey Response Quality," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow Scotland Uk: ACM, May 2019, pp. 1–12. doi: 10.1145/3290605.3300316.
- [192] A. Ho, J. Hancock, and A. S. Miner, "Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot," *J. Commun.*, vol. 68, no. 4, pp. 712–733, Aug. 2018, doi: 10.1093/joc/jqy026.
- [193] M. Savin-Baden, G. Tombs, and R. Bhakta, "Beyond robotic wastelands of time: Abandoned pedagogical agents and new pedalled pedagogies," *E-Learn. Digit. Media*, vol. 12, no. 3–4, pp. 295–314, Jun. 2015, doi: 10.1177/2042753015571835.

- [194] R. Kang, S. Brown, and S. Kiesler, “Why do people seek anonymity on the internet?: informing policy and design,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Paris France: ACM, Apr. 2013, pp. 2657–2666. doi: 10.1145/2470654.2481368.
- [195] D. M. Law, J. D. Shapka, S. Hymel, B. F. Olson, and T. Waterhouse, “The changing face of bullying: An empirical comparison between traditional and internet bullying and victimization,” *Comput. Hum. Behav.*, vol. 28, no. 1, pp. 226–232, Jan. 2012, doi: 10.1016/j.chb.2011.09.004.
- [196] J. T. Hancock, J. Thom-Santelli, and T. Ritchie, “Deception and design: the impact of communication technology on lying behavior,” in *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, Vienna, Austria: ACM Press, 2004, pp. 129–134. doi: 10.1145/985692.985709.
- [197] R. Crutzen, G.-J. Y. Peters, S. D. Portugal, E. M. Fisser, and J. J. Grolleman, “An artificially intelligent chat agent that answers adolescents’ questions related to sex, drugs, and alcohol: an exploratory study,” *J. Adolesc. Health*, vol. 48, no. 5, pp. 514–519, 2011.
- [198] E. A. J. Croes and M. L. Antheunis, “36 Questions to Loving a Chatbot: Are People Willing to Self-disclose to a Chatbot?,” in *Chatbot Research and Design*, vol. 12604, A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, E. Luger, M. Goodwin, and P. B. Brandtzaeg, Eds., in *Lecture Notes in Computer Science*, vol. 12604, Cham: Springer International Publishing, 2021, pp. 81–95. doi: 10.1007/978-3-030-68288-0\_6.
- [199] K. Loveys, G. Sebaratnam, M. Sagar, and E. Broadbent, “The Effect of Design Features on Relationship Quality with Embodied Conversational Agents: A Systematic Review,” *Int. J. Soc. Robot.*, vol. 12, no. 6, pp. 1293–1312, Dec. 2020, doi: 10.1007/s12369-020-00680-7.
- [200] J. Kang and L. Wei, ““ Give Me the Support I Want!’ The Effect of Matching an Embodied Conversational Agent’s Social Support to Users’ Social Support Needs in Fostering Positive User-Agent Interaction,” in *Proceedings of the 6th International Conference on Human-Agent Interaction*, 2018, pp. 106–113.
- [201] A. Kerry, R. Ellis, and S. Bull, “Conversational Agents in E-Learning,” in *Applications and Innovations in Intelligent Systems XVI*, T. Allen, R. Ellis, and M. Petridis, Eds., London: Springer London, 2009, pp. 169–182. doi: 10.1007/978-1-84882-215-3\_13.
- [202] D. Pérez-Marín, “A Review of the Practical Applications of Pedagogic Conversational Agents to Be Used in School and University Classrooms,” *Digital*, vol. 1, no. 1, pp. 18–33, Jan. 2021, doi: 10.3390/digital1010002.
- [203] F. De Rosis, C. Pelachaud, I. Poggi, V. Carofiglio, and B. De Carolis, “From Greta’s mind to her face: modelling the dynamics of affective states in a conversational embodied agent,” *Int. J. Hum.-Comput. Stud.*, vol. 59, no. 1–2, pp. 81–118, 2003.
- [204] L. Ring, L. Shi, K. Totzke, and T. Bickmore, “Social support agents for older adults: longitudinal affective computing in the home,” *J. Multimodal User Interfaces*, vol. 9, no. 1, pp. 79–88, 2015.
- [205] P. B. Bae Brandtzæg, M. Skjuve, K. K. Kristoffer Dysthe, and A. Følstad, “When the Social Becomes Non-Human: Young People’s Perception of Social Support in Chatbots,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, May 2021, pp. 1–13. doi: 10.1145/3411764.3445318.
- [206] B. Abendschein, C. Edwards, and A. Edwards, “The influence of agent and message type on perceptions of social support in human-machine communication,” *Commun. Res. Rep.*, vol. 38, no. 5, pp. 304–314, Oct. 2021, doi: 10.1080/08824096.2021.1966405.

- [207] M. L. Antheunis, "Friendships and the Internet," *Int. Encycl. Interpers. Commun.*, pp. 1–8, 2015.
- [208] S. S. Sundar and J. Kim, "Machine Heuristic: When We Trust Computers More than Humans with Our Personal Information," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow Scotland Uk: ACM, May 2019, pp. 1–9. doi: 10.1145/3290605.3300768.
- [209] A. S. Miner, A. Milstein, S. Schueller, R. Hegde, C. Mangurian, and E. Linos, "Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health," *JAMA Intern. Med.*, vol. 176, no. 5, p. 619, May 2016, doi: 10.1001/jamainternmed.2016.0400.
- [210] National Sexual Assault, "Domestic family violence counselling service. . Sunny app | 1800 Respect." [Online]. Available: <https://www.1800respect.org.au/sunny>
- [211] G. McLean and S. G. Bocinski, "The Economic Cost of Intimate Partner Violence, Sexual Assault, and Stalking: Fact Sheet," 2017.
- [212] R. E. Morgan and J. L. Truman, "Criminal victimization, 2020," *Wash. DC Natl. Crime Victim. Surv. Bur. Justice Stat. Retrieved Jan*, vol. 4, p. 2022, 2021.
- [213] E. Gracia, "Unreported cases of domestic violence against women: towards an epidemiology of social silence, tolerance, and inhibition," *J. Epidemiol. Community Health*, vol. 58, no. 7, pp. 536–537, Jul. 2004, doi: 10.1136/jech.2003.019604.
- [214] United Nations, "Sustainable Development Goals," 2015, Accessed: Feb. 09, 2022. [Online]. Available: <https://www.un.org/sustainabledevelopment/gender-equality/>
- [215] T. Nadarzynski *et al.*, "Barriers and facilitators to engagement with artificial intelligence (AI)-based chatbots for sexual and reproductive health advice: a qualitative analysis," *Sex. Health*, vol. 18, no. 5, p. 385, 2021, doi: 10.1071/SH21123.
- [216] World Health Organization, *Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence*. Geneva: World Health Organization, 2013. Accessed: Jan. 06, 2022. [Online]. Available: <https://apps.who.int/iris/handle/10665/85239>
- [217] R. D. Nixon, P. A. Resick, and P. Nishith, "An exploration of comorbid depression among female victims of intimate partner violence with posttraumatic stress disorder," *J. Affect. Disord.*, vol. 82, no. 2, pp. 315–320, 2004.
- [218] K. Arroyo, B. Lundahl, R. Butters, M. Vanderloo, and D. S. Wood, "Short-Term Interventions for Survivors of Intimate Partner Violence: A Systematic Review and Meta-Analysis," *Trauma Violence Abuse*, vol. 18, no. 2, pp. 155–171, Apr. 2017, doi: 10.1177/1524838015602736.
- [219] H. Abrahams, *Rebuilding lives after domestic violence: Understanding long-term outcomes*. Jessica Kingsley Publishers, 2010.
- [220] H. Yeung, N. Chowdhury, A. Malpass, and G. S. Feder, "Responding to Domestic Violence in General Practice: A Qualitative Study on Perceptions and Experiences," *Int. J. Fam. Med.*, vol. 2012, pp. 1–7, Nov. 2012, doi: 10.1155/2012/960523.
- [221] A. L. Coker, P. H. Smith, D. J. Whitaker, B. Le, T. N. Crawford, and V. C. Flerx, "Effect of an In-Clinic IPV Advocate Intervention to Increase Help Seeking, Reduce Violence, and Improve Well-Being," *Violence Women*, vol. 18, no. 1, pp. 118–131, Jan. 2012, doi: 10.1177/1077801212437908.
- [222] S. Rajpara, "33 Chatbot Statistics and Trends to Watch in 2022." [Online]. Available: <https://acquire.io/blog/chatbots-trends/>

- [223] R. Baskerville, A. Baiyere, S. Gergor, A. Hevner, M. Rossi, and Aalto University, "Design Science Research Contributions: Finding a Balance between Artifact and Theory," *J. Assoc. Inf. Syst.*, vol. 19, no. 5, pp. 358–376, May 2018, doi: 10.17705/1jais.00495.
- [224] L. Chandra, S. Seidel, and S. Gregor, "Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions," in *2015 48th Hawaii International Conference on System Sciences*, HI, USA: IEEE, Jan. 2015, pp. 4039–4048. doi: 10.1109/HICSS.2015.485.
- [225] University of Mannheim *et al.*, "Designing a Requirement Mining System," *J. Assoc. Inf. Syst.*, vol. 16, no. 9, pp. 799–837, Sep. 2015, doi: 10.17705/1jais.00408.
- [226] J. G. Walls, G. R. Widmeyer, and O. A. El Sawy, "Building an Information System Design Theory for Vigilant EIS," *Inf. Syst. Res.*, vol. 3, no. 1, pp. 36–59, Mar. 1992, doi: 10.1287/isre.3.1.36.
- [227] R. Molich and J. Nielsen, "Improving a human-computer dialogue," *Commun. ACM*, vol. 33, no. 3, pp. 338–348, Mar. 1990, doi: 10.1145/77481.77486.
- [228] P. Meier, J. H. Beinke, C. Fitté, A. Behne, and F. Teuteberg, "FeelFit-Design and Evaluation of a Conversational Agent to Enhance Health Awareness.," in *ICIS*, 2019.
- [229] H. Park and J. Lee, "Can a Conversational Agent Lower Sexual Violence Victims' Burden of Self-Disclosure?," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–8. doi: 10.1145/3334480.3383050.
- [230] J. Feine, M. Adam, I. Benke, A. Maedche, and A. Benlian, "Exploring design principles for enterprise chatbots: An analytic hierarchy process study," in *International Conference on Design Science Research in Information Systems and Technology*, Springer, 2020, pp. 126–141.
- [231] J. Feine, U. Gnewuch, S. Morana, and A. Maedche, "A Taxonomy of Social Cues for Conversational Agents," *Int. J. Hum.-Comput. Stud.*, vol. 132, pp. 138–161, Dec. 2019, doi: 10.1016/j.ijhcs.2019.07.009.
- [232] S. Diederich, A. B. Brendel, and L. M. Kolbe, "Designing Anthropomorphic Enterprise Conversational Agents," *Bus. Inf. Syst. Eng.*, vol. 62, no. 3, pp. 193–209, Jun. 2020, doi: 10.1007/s12599-020-00639-y.
- [233] S. Stieglitz, L. Hofeditz, F. Brünker, C. Ehnis, M. Mirbabaie, and B. Ross, "Design principles for conversational agents to support Emergency Management Agencies," *Int. J. Inf. Manag.*, vol. 63, p. 102469, 2022.
- [234] B. Kuechler and V. Vaishnavi, "On theory development in design science research: anatomy of a research project," *Eur. J. Inf. Syst.*, vol. 17, no. 5, pp. 489–504, 2008.
- [235] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Q.*, pp. 75–105, 2004.
- [236] S. Gregor, L. Kruse, University of Liechtentsein, Liechtentsein, S. Seidel, and University of Liechtentsein, Liechtentsein, "Research Perspectives: The Anatomy of a Design Principle," *J. Assoc. Inf. Syst.*, vol. 21, pp. 1622–1652, Nov. 2020, doi: 10.17705/1jais.00649.
- [237] V. Braun and V. Clarke, *Thematic analysis*. American Psychological Association, 2012.
- [238] V. Vaishnavi and W. Kuechler, "Design research in information systems," 2004.
- [239] V. Braun and V. Clarke, *Successful qualitative research: A practical guide for beginners*. sage, 2013.
- [240] J. S. House, "Work stress and social support," *Addison-Wesley Ser. Occup. Stress*, 1981.

- [241] J. C. Campbell, D. W. Webster, and N. Glass, "The Danger Assessment: Validation of a Lethality Risk Assessment Instrument for Intimate Partner Femicide," *J. Interpers. Violence*, vol. 24, no. 4, pp. 653–674, Apr. 2009, doi: 10.1177/0886260508317180.
- [242] L. Wood, D. Hairston, R. V. Schrag, E. Clark, R. Parra-Cardona, and J. R. Temple, "Creating a Digital Trauma Informed Space: Chat and Text Advocacy for Survivors of Violence," *J. Interpers. Violence*, p. 088626052110435, Oct. 2021, doi: 10.1177/08862605211043573.
- [243] T. Gierman and A. Liska, "Shelter for Women and Girls at Risk of or Survivors of Violence," *U. N. Woman Virtual Knowl. Cent. End Violence Women Girls Httpwww Endvawnow Orguploadsmodulespdf1363538451 Pdf*, vol. 30, p. 2019, 2013.
- [244] S. E. Hobfoll, "Conservation of resources: a new attempt at conceptualizing stress.," *Am. Psychol.*, vol. 44, no. 3, p. 513, 1989.
- [245] S. E. Hobfoll, "The influence of culture, community, and the nested-self in the stress process: Advancing conservation of resources theory," *Appl. Psychol.*, vol. 50, no. 3, pp. 337–421, 2001.
- [246] R. Hoffart and N. A. Jones, "Intimate Partner Violence and Intergenerational Trauma Among Indigenous Women," *Int. Crim. Justice Rev.*, vol. 28, no. 1, pp. 25–44, Mar. 2018, doi: 10.1177/1057567717719966.
- [247] C. Menschner and A. Maul, *Key ingredients for successful trauma-informed care implementation*. Center for Health Care Strategies, Incorporated Trenton, 2016.
- [248] A. Cerekovic, O. Aran, and D. Gatica-Perez, "Rapport with virtual agents: What do human social cues and personality explain?," *IEEE Trans. Affect. Comput.*, vol. 8, no. 3, pp. 382–395, 2016.
- [249] K. M. Gallagher and J. A. Updegraff, "Health Message Framing Effects on Attitudes, Intentions, and Behavior: A Meta-analytic Review," *Ann. Behav. Med.*, vol. 43, no. 1, pp. 101–116, Feb. 2012, doi: 10.1007/s12160-011-9308-7.
- [250] Buffalo Center for Social Research, "What is trauma-informed care." 2021. [Online]. Available: <https://socialwork.buffalo.edu/social-research/institutes-centers/institute-on-trauma-and-trauma-informed-care/what-is-trauma-informed-care.html>
- [251] J. X. Chen *et al.*, "Trauma-Informed Computing: Towards Safer Technology Experiences for All," in *CHI Conference on Human Factors in Computing Systems*, New Orleans LA USA: ACM, Apr. 2022, pp. 1–20. doi: 10.1145/3491102.3517475.
- [252] Y. Ahn, Y. Zhang, Y. Park, and J. Lee, "A Chatbot Solution to Chat App Problems: Envisioning a Chatbot Counseling System for Teenage Victims of Online Sexual Exploitation," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–7. doi: 10.1145/3334480.3383070.
- [253] E. L. Pence and M. Paymar, "Education Groups for Men Who Batter: The Duluth Model," 1993. [Online]. Available: <https://www.ojp.gov/ncjrs/virtuallibrary/abstracts/education-groups-men-who-batter-duluth-model>
- [254] V. Ta *et al.*, "User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis," *J. Med. Internet Res.*, vol. 22, no. 3, p. e16235, Mar. 2020, doi: 10.2196/16235.



- [255] M. De Gennaro, E. G. Krumhuber, and G. Lucas, “Effectiveness of an Empathic Chatbot in Combating Adverse Effects of Social Exclusion on Mood,” *Front. Psychol.*, vol. 10, p. 3061, Jan. 2020, doi: 10.3389/fpsyg.2019.03061.
- [256] J. Kim, Y. Kim, B. Kim, S. Yun, M. Kim, and J. Lee, “Can a Machine Tend to Teenagers’ Emotional Needs?: A Study with Conversational Agents,” in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC Canada: ACM, Apr. 2018, pp. 1–6. doi: 10.1145/3170427.3188548.
- [257] A. Wahbeh, M. A. Al-Ramahi, O. El-Gayar, A. El Noshokaty, and T. Nasrallah, “Conversational Agents for Mental Health and Well-being: Discovering Design Recommendations Using Text Mining,” 2023.
- [258] N. M. Overstreet and D. M. Quinn, “The Intimate Partner Violence Stigmatization Model and Barriers to Help Seeking,” *Basic Appl. Soc. Psychol.*, vol. 35, no. 1, pp. 109–122, Jan. 2013, doi: 10.1080/01973533.2012.746599.
- [259] J. Zamora, “I’m Sorry, Dave, I’m Afraid I Can’t Do That: Chatbot Perception and Expectations,” in *Proceedings of the 5th International Conference on Human Agent Interaction*, Bielefeld Germany: ACM, Oct. 2017, pp. 253–260. doi: 10.1145/3125739.3125766.
- [260] S. Cormier, P. S. Nurius, and C. J. Osborn, *Interviewing and change strategies for helpers: Fundamental skills and cognitive-behavioral interventions*, Instructor’s ed. Brooks/Cole, 2009.
- [261] E. A. Bowen and N. S. Murshid, “Trauma-Informed Social Policy: A Conceptual Framework for Policy Analysis and Advocacy,” *Am. J. Public Health*, vol. 106, no. 2, pp. 223–229, Feb. 2016, doi: 10.2105/AJPH.2015.302970.
- [262] K. Kretschmar, H. Tyroll, G. Pavarini, A. Manzini, I. Singh, and NeurOx Young People’s Advisory Group, “Can Your Phone Be Your Therapist? Young People’s Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support,” *Biomed. Inform. Insights*, vol. 11, p. 117822261982908, Jan. 2019, doi: 10.1177/1178222619829083.
- [263] D. Griol, L. F. Hurtado, E. Segarra, and E. Sanchis, “A statistical approach to spoken dialog systems design and evaluation,” *Speech Commun.*, vol. 50, no. 8–9, pp. 666–682, 2008.
- [264] D. Griol, Z. Callejas, R. López-Cózar, and G. Riccardi, “A domain-independent statistical methodology for dialog management in spoken dialog systems,” *Comput. Speech Lang.*, vol. 28, no. 3, pp. 743–768, 2014.
- [265] J. F. Allen, D. K. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent, “Toward conversational human-computer interaction,” *AI Mag.*, vol. 22, no. 4, pp. 27–27, 2001.
- [266] D. C. Derrick, J. L. Jenkins, and J. F. Nunamaker Jr, “Design principles for special purpose, embodied, conversational intelligence with environmental sensors (SPECIES) agents,” *AIS Trans. Hum.-Comput. Interact.*, vol. 3, no. 2, pp. 62–81, 2011.
- [267] A. C. Elkins, D. C. Derrick, J. K. Burgoon, and J. F. Nunamaker Jr., “Predicting Users’ Perceived Trust in Embodied Conversational Agents Using Vocal Dynamics,” in *2012 45th Hawaii International Conference on System Sciences*, Maui, HI, USA: IEEE, Jan. 2012, pp. 579–588. doi: 10.1109/HICSS.2012.483.
- [268] D. Atkinson *et al.*, “Trust in computers and robots: The uses and boundaries of the analogy to interpersonal trust,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications Sage CA: Los Angeles, CA, 2012, pp. 303–307.

- [269] Y. Moon, “Intimate Exchanges: Using Computers to Elicit Self-Disclosure From Consumers,” *J. Consum. Res.*, vol. 26, no. 4, pp. 323–339, Mar. 2000, doi: 10.1086/209566.
- [270] C. E. Murray *et al.*, “Domestic violence service providers’ perceptions of safety planning: A focus group study,” *J. Fam. Violence*, vol. 30, no. 3, pp. 381–392, 2015.
- [271] H. Kondylakis *et al.*, “Smart recommendation services in support of patient empowerment and personalized medicine,” in *Multimedia Services in Intelligent Environments*, Springer, 2013, pp. 39–61.
- [272] “Trauma & Mental Health National Center on Domestic Violence. A trauma-informed approach to domestic violence advocacy.” 2011. [Online]. Available: [http://nationalcenterdvtraumamh.org/wp-content/uploads/2012/01/Tipsheet\\_TI-DV-Advocacy\\_NCDVTMH\\_Aug2011.pdf](http://nationalcenterdvtraumamh.org/wp-content/uploads/2012/01/Tipsheet_TI-DV-Advocacy_NCDVTMH_Aug2011.pdf)
- [273] N. Motalebi and S. Abdullah, “Conversational Agents to Provide Couple Therapy for Patients with PTSD,” in *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, New York NY USA: ACM, May 2018, pp. 347–351. doi: 10.1145/3240925.3240933.
- [274] T. A. Maniou and A. Veglis, “Employing a chatbot for news dissemination during crisis: Design, implementation and evaluation,” *Future Internet*, vol. 12, no. 7, p. 109, 2020.
- [275] R. M. Schuetzler, G. M. Grimes, and J. Scott Giboney, “The impact of chatbot conversational skill on engagement and perceived humanness,” *J. Manag. Inf. Syst.*, vol. 37, no. 3, pp. 875–900, Jul. 2020, doi: 10.1080/07421222.2020.1790204.
- [276] E. Svikhnushina, A. Placinta, and P. Pu, “User Expectations of Conversational Chatbots Based on Online Reviews,” in *Designing Interactive Systems Conference 2021*, Virtual Event USA: ACM, Jun. 2021, pp. 1481–1491. doi: 10.1145/3461778.3462125.
- [277] A. Muresan and H. Pohl, “Chats with bots: Balancing imitation and engagement,” in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–6.
- [278] C. L. Mason and S. Magnet, “Surveillance studies and violence against women,” *Surveill. Soc.*, vol. 10, no. 2, pp. 105–118, 2012.
- [279] N. E. Glass *et al.*, “Longitudinal Impact of the myPlan App on Health and Safety Among College Women Experiencing Partner Violence,” *J. Interpers. Violence*, vol. 37, no. 13–14, pp. NP11436–NP11459, Jul. 2022, doi: 10.1177/0886260521991880.
- [280] S. E. Ullman, “Psychometric Characteristics of the Social Reactions Questionnaire: A Measure of Reactions to Sexual Assault Victims,” *Psychol. Women Q.*, vol. 24, no. 3, pp. 257–271, Sep. 2000, doi: 10.1111/j.1471-6402.2000.tb00208.x.
- [281] J. Pereira and Ó. Díaz, “Using Health Chatbots for Behavior Change: A Mapping Study,” *J. Med. Syst.*, vol. 43, no. 5, p. 135, May 2019, doi: 10.1007/s10916-019-1237-1.
- [282] M. Lee, S. Ackermans, N. van As, H. Chang, E. Lucas, and W. IJsselsteijn, “Caring for Vincent: A Chatbot for Self-Compassion,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow Scotland Uk: ACM, May 2019, pp. 1–13. doi: 10.1145/3290605.3300932.
- [283] K. M. Staller and D. Nelson-Gardell, “‘A burden in your heart’: Lessons of disclosure from female preadolescent and adolescent survivors of sexual abuse,” *Child Abuse Negl.*, vol. 29, no. 12, pp. 1415–1432, Dec. 2005, doi: 10.1016/j.chiabu.2005.06.007.
- [284] J. Gupta *et al.*, “A nurse-delivered, clinic-based intervention to address intimate partner violence among low-income women in Mexico City: findings from a cluster randomized

- controlled trial,” *BMC Med.*, vol. 15, no. 1, p. 128, Dec. 2017, doi: 10.1186/s12916-017-0880-y.
- [285] M. M. Geyser, “An exploratory case study analysis of a restorative justice intervention project that mediated domestic violence cases in KwaZulu-Natal, South Africa,” 2017.
- [286] S. Siedlikowski, L.-P. Noël, S. A. Moynihan, and M. Robin, “Chloe for COVID-19: Evolution of an Intelligent Conversational Agent to Address Infodemic Management Needs During the COVID-19 Pandemic,” *J. Med. Internet Res.*, vol. 23, no. 9, p. e27283, Sep. 2021, doi: 10.2196/27283.
- [287] B. Kim, J. Park, and J. Suh, “Transparency and accountability in AI decision support: Explaining and visualizing convolutional neural networks for text information,” *Decis. Support Syst.*, vol. 134, p. 113302, Jul. 2020, doi: 10.1016/j.dss.2020.113302.
- [288] R. Ahmad, D. Siemon, U. Gnewuch, and S. Robra-Bissantz, “Designing Personality-Adaptive Conversational Agents for Mental Health Care,” *Inf. Syst. Front.*, vol. 24, no. 3, pp. 923–943, Jun. 2022, doi: 10.1007/s10796-022-10254-9.
- [289] J. Balakrishnan and Y. K. Dwivedi, “Conversational commerce: entering the next stage of AI-powered digital assistants,” *Ann. Oper. Res.*, Apr. 2021, doi: 10.1007/s10479-021-04049-5.
- [290] M. Allouch, A. Azaria, and R. Azoulay, “Conversational Agents: Goals, Technologies, Vision and Challenges,” *Sensors*, vol. 21, no. 24, p. 8448, Dec. 2021, doi: 10.3390/s21248448.
- [291] D. Freed, J. Palmer, D. Minchala, K. Levy, T. Ristenpart, and N. Dell, “‘A Stalker’s Paradise’ How Intimate Partner Abusers Exploit Technology,” in *Proceedings of the 2018 CHI conference on human factors in computing systems*, 2018, pp. 1–13.
- [292] C. Nass, J. Steuer, and E. R. Tauber, “Computers are social actors,” in *Conference companion on Human factors in computing systems - CHI '94*, Boston, Massachusetts, United States: ACM Press, 1994, p. 204. doi: 10.1145/259963.260288.
- [293] S. S. Sundar, *The MAIN model: A heuristic approach to understanding technology effects on credibility*. MacArthur Foundation Digital Media and Learning Initiative Cambridge, MA, 2008.
- [294] Y. Kim and S. S. Sundar, “Anthropomorphism of computers: Is it mindful or mindless?,” *Comput. Hum. Behav.*, vol. 28, no. 1, pp. 241–250, Jan. 2012, doi: 10.1016/j.chb.2011.09.006.
- [295] M. Kumar and C. H. Noble, “Beyond form and function: Why do consumers value product design?,” *J. Bus. Res.*, vol. 69, no. 2, pp. 613–620, 2016.
- [296] E. Stoeckli, F. Uebernickel, and W. Brenner, “Exploring affordances of slack integrations and their actualization within enterprises-towards an understanding of how chatbots create value,” in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.
- [297] B. Hine, S. Wallace, and E. A. Bates, “Understanding the Profile and Needs of Abused Men: Exploring Call Data From a Male Domestic Violence Charity in the United Kingdom,” *J. Interpers. Violence*, p. 088626052110280, Jun. 2021, doi: 10.1177/08862605211028014.
- [298] “FACEBOOK MESSENGER USERS, STATS, DATA & TRENDS.” May 2023. [Online]. Available: <https://datareportal.com/essential-facebook-messenger-stats>
- [299] S. Nguyen and Daniel, “Telegram Updated Statistics In 2024 | Everything You Need To Know.” Nov. 2023. [Online]. Available: <https://avada.io/articles/telegram-statistics/>

- [300] K. A. Baikie and K. Wilhelm, "Emotional and physical health benefits of expressive writing," *Adv. Psychiatr. Treat.*, vol. 11, no. 5, pp. 338–346, Sep. 2005, doi: 10.1192/apt.11.5.338.
- [301] M. L. Benson, G. L. Fox, A. DeMaris, and J. Van Wyk, "Neighborhood disadvantage, individual economic distress and violence against women in intimate relationships," *J. Quant. Criminol.*, vol. 19, no. 3, pp. 207–235, 2003.
- [302] M. M. Chilton, J. R. Rabinowich, and N. H. Woolf, "Very low food security in the USA is linked with exposure to violence," *Public Health Nutr.*, vol. 17, no. 1, pp. 73–82, 2014.
- [303] C. A. Byrne, H. S. Resnick, D. G. Kilpatrick, C. L. Best, and B. E. Saunders, "The socioeconomic impact of interpersonal violence on women.," *J. Consult. Clin. Psychol.*, vol. 67, no. 3, p. 362, 1999.
- [304] D. B. Jackson, K. R. Lynch, J. J. Helton, and M. G. Vaughn, "Food Insecurity and Violence in the Home: Investigating Exposure to Violence and Victimization Among Preschool-Aged Children," *Health Educ. Behav.*, vol. 45, no. 5, pp. 756–763, Oct. 2018, doi: 10.1177/1090198118760683.
- [305] R. C. Davis, C. D. Maxwell, and B. Taylor, "Preventing repeat incidents of family violence: analysis of data from three field experiments," *J. Exp. Criminol.*, vol. 2, no. 2, pp. 183–210, Jun. 2006, doi: 10.1007/s11292-006-9002-0.
- [306] P. C. Friday, V. B. Lord, M. L. Exum, and J. L. Hartman, "Evaluating the impact of a specialized domestic violence police unit, final report," *Wash. DC Natl. Inst. Justice*, 2006.
- [307] S. B. Scott, G. K. Rhoades, S. M. Stanley, E. S. Allen, and H. J. Markman, "Reasons for divorce and recollections of premarital intervention: Implications for improving relationship education.," *Couple Fam. Psychol. Res. Pract.*, vol. 2, no. 2, pp. 131–145, 2013, doi: 10.1037/a0032025.
- [308] K. M. Kitzmann, N. K. Gaylord, A. R. Holt, and E. D. Kenny, "Child witnesses to domestic violence: a meta-analytic review.," *J. Consult. Clin. Psychol.*, vol. 71, no. 2, p. 339, 2003.
- [309] J. Campbell and N. Glass, "Safety planning, danger, and lethality assessment.," in *Intimate partner violence: A health-based perspective.*, New York, NY, US: Oxford University Press, 2009, pp. 319–334.
- [310] K. B. Eden *et al.*, "Patients were more consistent in randomized trial at prioritizing childbirth preferences using graphic-numeric than verbal formats," *J. Clin. Epidemiol.*, vol. 62, no. 4, pp. 415–424.e3, Apr. 2009, doi: 10.1016/j.jclinepi.2008.05.012.
- [311] K. Y. A. McKenna and J. A. Bargh, "Coming out in the age of the Internet: Identity 'demarginalization' through virtual group participation.," *J. Pers. Soc. Psychol.*, vol. 75, no. 3, pp. 681–694, Sep. 1998, doi: 10.1037/0022-3514.75.3.681.
- [312] R. Crutzen, J. de Nooijer, W. Brouwer, A. Oenema, J. Brug, and N. K. de Vries, "Qualitative assessment of adolescents' views about improving exposure to internet-delivered interventions," *Health Educ.*, vol. 108, no. 2, pp. 105–116, Feb. 2008, doi: 10.1108/09654280810855568.
- [313] I. Rowlands *et al.*, "The Google generation: the information behaviour of the researcher of the future," *Aslib Proc.*, vol. 60, no. 4, pp. 290–310, Jul. 2008, doi: 10.1108/00012530810887953.
- [314] R. Voth Schrag, D. Hairston, M. L. Brown, and L. Wood, "Advocate and Survivor Perspectives on the Role of Technology in Help Seeking and Services with Emerging Adults

- in Higher Education,” *J. Fam. Violence*, vol. 37, no. 1, pp. 123–136, Jan. 2022, doi: 10.1007/s10896-021-00279-0.
- [315] “ChatGPT.” [Online]. Available: <https://chat.openai.com>
- [316] T. W. Bickmore, D. Utami, R. Matsuyama, and M. K. Paasche-Orlow, “Improving access to online health information with conversational agents: a randomized controlled experiment,” *J. Med. Internet Res.*, vol. 18, no. 1, p. e1, 2016.
- [317] A. S. Miner, L. Laranjo, and A. B. Kocaballi, “Chatbots in the fight against the COVID-19 pandemic,” *NPJ Digit. Med.*, vol. 3, no. 1, p. 65, 2020.
- [318] P. B. Brandtzaeg and A. Følstad, “Why people use chatbots,” in *Internet Science: 4th International Conference, INSCI 2017, Thessaloniki, Greece, November 22-24, 2017, Proceedings 4*, Springer, 2017, pp. 377–392.
- [319] I. Ucuz, A. Ari, O. O. Ozcan, O. Topaktas, M. Sarraf, and O. Dogan, “Estimation of the development of depression and PTSD in children exposed to sexual abuse and development of decision support systems by using artificial intelligence,” *J. Child Sex. Abuse*, vol. 31, no. 1, pp. 73–85, 2022.
- [320] H. L. Walker *et al.*, “Reliability of Medical Information Provided by ChatGPT: Assessment Against Clinical Guidelines and Patient Information Quality Instrument,” *J. Med. Internet Res.*, vol. 25, p. e47479, Jun. 2023, doi: 10.2196/47479.
- [321] S. Spallek, L. Birrell, S. Kershaw, E. K. Devine, and L. Thornton, “Can we use ChatGPT for Mental Health and Substance Use Education? Examining Its Quality and Potential Harms,” *JMIR Med. Educ.*, vol. 9, p. e51243, Nov. 2023, doi: 10.2196/51243.
- [322] Y.-C. Lee, N. Yamashita, and Y. Huang, “Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional,” *Proc. ACM Hum.-Comput. Interact.*, vol. 4, no. CSCW1, pp. 1–27, May 2020, doi: 10.1145/3392836.
- [323] A. S. Miner, A. Milstein, and J. T. Hancock, “Talking to machines about personal mental health problems,” *Jama*, vol. 318, no. 13, pp. 1217–1218, 2017.
- [324] L. Liao, G. H. Yang, and C. Shah, “Proactive Conversational Agents in the Post-ChatGPT World,” in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Taipei Taiwan: ACM, Jul. 2023, pp. 3452–3455. doi: 10.1145/3539618.3594250.
- [325] A. Choudhury and H. Shamszare, “Investigating the Impact of User Trust on the Adoption and Use of ChatGPT: Survey Analysis,” *J. Med. Internet Res.*, vol. 25, p. e47184, Jun. 2023, doi: 10.2196/47184.
- [326] F. Fui-Hoon Nah, R. Zheng, J. Cai, K. Siau, and L. Chen, “Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration,” *J. Inf. Technol. Case Appl. Res.*, vol. 25, no. 3, pp. 277–304, Jul. 2023, doi: 10.1080/15228053.2023.2233814.
- [327] “Growth forecast of monthly active users of ChatGPT in December 2022 and January 2023 (in millions),” CBS News, [Graph], Feb. 2023. [Online]. Available: <https://www.statista.com/statistics/1368657/chatgpt-mau-growth/>
- [328] N. Yamawaki, M. Ochoa-Shipp, C. Pulsipher, A. Harlos, and S. Swindler, “Perceptions of Domestic Violence: The Effects of Domestic Violence Myths, Victim’s Relationship With Her Abuser, and the Decision to Return to Her Abuser,” *J. Interpers. Violence*, vol. 27, no. 16, pp. 3195–3212, Nov. 2012, doi: 10.1177/0886260512441253.
- [329] T. LaLiberte, J. Bills, N. Shin, and J. L. Edleson, “Child welfare professionals’ responses to domestic violence exposure among children,” *Child. Youth Serv. Rev.*, vol. 32, no. 12, pp. 1640–1647, Dec. 2010, doi: 10.1016/j.childyouth.2010.06.025.

- [330] N. T. Feather, "Domestic violence, gender, and perceptions of justice," *Sex Roles*, vol. 35, no. 7, pp. 507–519, Oct. 1996, doi: 10.1007/BF01544134.
- [331] J. Finch, "The Vignette Technique in Survey Research," *Sociology*, vol. 21, no. 1, pp. 105–114, Feb. 1987, doi: 10.1177/0038038587021001008.
- [332] C. Atzmüller and P. M. Steiner, "Experimental Vignette Studies in Survey Research," *Methodology*, vol. 6, no. 3, pp. 128–138, Jan. 2010, doi: 10.1027/1614-2241/a000014.
- [333] S. E. Ullman, "Social support and recovery from sexual assault," *Aggress. Violent Behav.*, vol. 4, no. 3, pp. 343–358, Sep. 1999, doi: 10.1016/S1359-1789(98)00006-8.
- [334] K. M. Edwards, C. M. Dardis, K. M. Sylaska, and C. A. Gidycz, "Informal Social Reactions to College Women's Disclosure of Intimate Partner Violence: Associations With Psychological and Relational Variables," *J. Interpers. Violence*, vol. 30, no. 1, pp. 25–44, Jan. 2015, doi: 10.1177/0886260514532524.
- [335] T. P. Sullivan, J. A. Schroeder, D. N. Dudley, and J. M. Dixon, "Do Differing Types of Victimization and Coping Strategies Influence the Type of Social Reactions Experienced by Current Victims of Intimate Partner Violence?," *Violence Women*, vol. 16, no. 6, pp. 638–657, Jun. 2010, doi: 10.1177/1077801210370027.
- [336] Y. W. Lee, D. M. Strong, B. K. Kahn, and R. Y. Wang, "AIMQ: a methodology for information quality assessment," *Inf. Manage.*, vol. 40, no. 2, pp. 133–146, 2002.
- [337] J. Brooke, "SUS-A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, no. 194, pp. 4–7, 1996.
- [338] K. Coperich, E. Cudney, and H. Nembhard, "Continuous improvement study of chatbot technologies using a human factors methodology," in *Proceedings of the 2017 industrial and systems engineering conference*, 2017.
- [339] D. M. Hite, T. Voelker, and A. Robertson, "Measuring perceived anonymity: The development of a context independent instrument," *J. Methods Meas. Soc. Sci.*, vol. 5, no. 1, pp. 22–39, 2014.
- [340] Anonymous, "To Reveal or Not to Reveal: A Theoretical Model of Anonymous Communication," *Commun. Theory*, vol. 8, no. 4, pp. 381–407, Nov. 1998, doi: 10.1111/j.1468-2885.1998.tb00226.x.
- [341] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *Int. J. Cogn. Ergon.*, vol. 4, no. 1, pp. 53–71, 2000.
- [342] K. M. Sherin, J. M. Sinacore, X. Q. Li, R. E. Zitter, and A. Shakil, "HITS: a short domestic violence screening tool for use in a family practice setting," *Fam. Med.*, vol. 30, no. 7, pp. 508–512, 1998.
- [343] C. Lewis, P. G. Polson, C. Wharton, and J. Rieman, "Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces," in *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90*, Seattle, Washington, United States: ACM Press, 1990, pp. 235–242. doi: 10.1145/97243.97279.
- [344] M. H. Blackmon, P. G. Polson, M. Kitajima, and C. Lewis, "Cognitive walkthrough for the web," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2002, pp. 463–470.
- [345] A. M. Aldkheel and L. Zhou, "How to Support Domestic Violence Survivors with Conversational Agents: Meta Requirements and Design Principles," 2023.
- [346] J. Sauro and J. R. Lewis, "When designing usability questionnaires, does it hurt to be positive?," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011, pp. 2215–2224.

- [347] A. Ait-Mlouk and L. Jiang, “KBot: A Knowledge Graph Based ChatBot for Natural Language Understanding Over Linked Data,” *IEEE Access*, vol. 8, pp. 149220–149230, 2020, doi: 10.1109/ACCESS.2020.3016142.
- [348] R. Bapat, P. Kucherbaev, and A. Bozzon, “Effective Crowdsourced Generation of Training Data for Chatbots Natural Language Understanding,” in *Web Engineering*, vol. 10845, T. Mikkonen, R. Klamma, and J. Hernández, Eds., in Lecture Notes in Computer Science, vol. 10845. , Cham: Springer International Publishing, 2018, pp. 114–128. doi: 10.1007/978-3-319-91662-0\_8.
- [349] S. Prasad Uprety and S. Ryul Jeong, “The Impact of Semi-Supervised Learning on the Performance of Intelligent Chatbot System,” *Comput. Mater. Contin.*, vol. 71, no. 2, pp. 3937–3952, 2022, doi: 10.32604/cmc.2022.023127.
- [350] R. Al Farisi, R. Ferdiana, and T. B. Adj, “The Effect of Anthropomorphic Design Cues on Increasing Chatbot Empathy,” in *2022 1st International Conference on Information System & Information Technology (ICISIT)*, Yogyakarta, Indonesia: IEEE, Jul. 2022, pp. 370–375. doi: 10.1109/ICISIT54091.2022.9873008.
- [351] Y. Li, T. Yao, R. Hu, T. Mei, and Y. Rui, “Video ChatBot: Triggering Live Social Interactions by Automatic Video Commenting,” in *Proceedings of the 24th ACM international conference on Multimedia*, Amsterdam The Netherlands: ACM, Oct. 2016, pp. 757–758. doi: 10.1145/2964284.2973835.
- [352] W. Maeng and J. Lee, “Designing a chatbot for survivors of sexual violence: Exploratory study for hybrid approach combining rule-based chatbot and ml-based chatbot,” in *Asian CHI Symposium 2021*, 2021, pp. 160–166.
- [353] J. Castellà-Roca, A. Viejo, and J. Herrera-Joancomartí, “Preserving user’s privacy in web search engines,” *Comput. Commun.*, vol. 32, no. 13–14, pp. 1541–1551, Aug. 2009, doi: 10.1016/j.comcom.2009.05.009.
- [354] J. M. van der Zwaan, V. Dignum, and C. M. Jonker, “Social Support Strategies for Embodied Conversational Agents,” *Emot. Model. Pragmatic Comput. Models Affect. Process.*, pp. 134–147, 2014.
- [355] L. E. Ross *et al.*, “Barriers and facilitators to primary care for people with mental health and/or substance use issues: a qualitative study,” *BMC Fam. Pract.*, vol. 16, no. 1, p. 135, Dec. 2015, doi: 10.1186/s12875-015-0353-3.
- [356] J. J. Prochaska *et al.*, “A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study,” *J. Med. Internet Res.*, vol. 23, no. 3, p. e24850, Mar. 2021, doi: 10.2196/24850.
- [357] S. N. Jumarali, N. Nnawulezi, S. Royson, C. Lippy, A. N. Rivera, and T. Toopet, “Participatory Research Engagement of Vulnerable Populations: Employing Survivor-Centered, Trauma-Informed Approaches,” *J. Particip. Res. Methods*, vol. 2, no. 2, Jul. 2021, doi: 10.35844/001c.24414.

## APPENDICES

### APPENDIX A: INTERVIEW PROTOCOL FOR STUDY 1

Introduction: My name is Abdulrahman Aldkheel, and I'm a Ph.D. student at UNCC. I'm currently working on my dissertation on technology-supported domestic violence intervention. The primary objective of my research is to design a conversational agent (chatbot) that assists victims of domestic violence who choose to disclose their victimization experience on social media.

Objective: The goal of this interview is to understand the needs of domestic violence victims based on your experience of interacting with them. This understanding will help inform the features and interaction design of the conversational agent.

Answers to any of the questions are neither right nor wrong. I would like to hear your perspectives and opinions based on your own experiences. Additionally, I would like your permission to audio record this session to have a copy of our notes. Once the transcriptions have been completed, all identifying information will be deleted from our recordings, and the original recording will be destroyed. Would you mind if we recorded the meeting?

Before we start the interview, we will begin with brief introductions about your job experiences and background.

- Kindly, can you tell us your job title, as well as the number of years you have worked with domestic violence survivors?
- In what capacity do your work with those survivors? Please describe your most typical experiences.



- Have you encountered survivors who have used tech-related intervention (online intervention, social media, mobile app, etc.)? If so, please share an example with us.

We have three dimensions to discuss. The needs and goals of victims, The design of the chatbot, Self-disclosure to social media, and the chatbot.

- What are the ways for you to get in touch with the survivors? Who typically initiates the contact?
- What questions do DV victims typically ask? What are your typical responses to each of those questions?
- What are the most common types of abuse that you have observed?
- What are the most stages of domestic violence relationships that you have observed?
- What are the most common types of support that the victims are seeking?
- What are the expectations of DV victims from interacting with you?
- What are your expectations of DV victims from interacting with them?
- What main outcomes do victims expect to achieve through interacting with police departments and advocacy groups?
- What are the most challenging aspects of interacting with a DV victim?
- How the online counseling differs from face-to-face counseling?

In this part, we want to discover some items regarding the conversation and features that could help us structure the design of the chatbot to be more helpful to the survivors.

Description of the chatbot: A chatbot is an automated system that is designed to simulate human communication or "chatter" using texts or voices. It allows people to interact directly with digital devices as if they were conversing with a human being. A chatbot could act as a counselor and

can be used as an effective tool for addressing health and social issues such as domestic violence. The chatbot is not intended to replace human experts or counselors like you. However, it can be viewed as a complementary tool to existing domestic violence services.

- What procedure do you follow in interacting with DV victims? Or how do you structure the flow of conversation?
- How do you start a conversation? How do you end a conversation?
- How to design the script of questions? Do you have an existing template you use?
- What are the best practices for establishing rapport with DV victims?
- What are the major questions that you recommend asking survivors before referring them to support?
- Are there any principles or guidelines that you typically follow in preparing responses to victims' questions?
- How to build a relationship between the situation and the direction of the conversation? Any available guidance?
- What language do you recommend using when designing the chatbot (i.e., Friendly, trauma-informed language)?
- What are the main capabilities/features that the user needs from a chatbot?
- What types of interventions, counseling, and support could a chatbot potentially provide (i.e., informational support, emotional support, instrumental support)?
- When we make the chatbot more intelligent, what intelligence features would the survivor expect to see?
- In designing a chatbot, which types of questions are more useful to ask users, open-ended questions, or close-ended questions?

- Do you perceive any strengths of using a chatbot to support DV victims compared to a human expert?
- What questions the chatbot should avoid when interacting with the user?
- What types of DV victims are most likely to benefit from the chatbot? In terms of abuse type? Stage of relationship? Age? Gender?
- What main outcomes do you expect that victim can achieve through interacting with a chatbot?
- If we had a chatbot that can interact with DV victims as a human counselor does, do you think the victim would prefer to interact with the chatbot or the human expert?
- What do you think about implementing safety decision aids?

In the last part, we want to discuss the self-disclosure behavior of survivors when they disclose their stories on different technological platforms.

- What are the main challenges and burdens that DV victims might face in disclosing their experience on social media?
- What are the main benefits/advantages that users might get after the disclosure on social media?
- What are the main challenges and burdens that DV victims might face in disclosing their experience to a chatbot?
- What are the main benefits/advantages that users might get from disclosure to a chatbot?
- How likely are the victims to disclose their stories on social media?
- How likely are the victims to disclose their stories to a chatbot?

- How do the degrees of disclosure to chatbots compare with social media, and disclosure to a chatbot compare with human experts)?
- In general, when the disclosure is beneficial, and why?

## APPENDIX B: SURVEY INSTRUMENTS

### Social Reaction Questionnaire (SRQ) [280]:

#### Emotional support

- Assuring the survivor that they were not to blame.
- Telling the survivor that they did not do anything wrong.
- Telling the survivor it is not their fault.
- Comforting the survivor by telling them it would be all right.
- Showing an understanding of the survivor's experience.
- Reframing the survivor's experience.
- Seeing the survivor's side of things and not making judgments.
- Being able to accept the survivor's account of their experience.
- Feeling sorry for the survivor.
- Seeming to understand the survivor's feelings.

#### Informational and instrumental support

- Helping the survivor obtain health care or shelter.
- Sharing information and offering options.
- Assisting the survivor in obtaining information regarding how to cope with the traumatic experience
- Encouraging the survivor to seek counseling services.

### AIMQ information quality assessment [336]:

- Accessibility: This information is easily accessible.
- Appropriate Amount: This information is of sufficient volume for our needs.

- Believability: This information is believable.
- Completeness: This information is sufficiently complete for our needs.
- Concise Representation: This information is presented concisely.
- Consistent Representation: This information is represented in a consistent format.
- Ease of Operation: This information is easy to manipulate to meet our needs.
- Free of Error: This information is correct.
- Interpretability: It is easy to interpret what this information means.
- Relevancy: This information is useful to our work.
- Timeliness: This information is sufficiently timely.
- Understandability: This information is easy to understand.

The System Usability Scale SUS [337]:

- I think that I would like to use this system (chatbot/ search engine/ ChatGPT) frequently.
- I found the system (chatbot/ search engine/ ChatGPT) unnecessarily complex.
- I thought the system (chatbot/ search engine/ ChatGPT) was easy to use.
- I think that I would need the support of a technical person to be able to use this system (chatbot/ search engine/ ChatGPT).
- I found the various functions in this system (chatbot/ search engine/ ChatGPT) were well integrated.
- I thought there was too much inconsistency in this system (chatbot/ search engine/ ChatGPT).
- I would imagine that most people would learn to use this system (chatbot/ search engine/ ChatGPT) very quickly.

- I found the system (chatbot/ search engine/ ChatGPT) very cumbersome to use.
- I felt very confident using the system (chatbot/ search engine/ ChatGPT).
- I needed to learn a lot of things before I could get going with this system (chatbot/search engine/ ChatGPT).

Perceived anonymity [339]:

- I am confident that others do not know who I am when I use the chatbot (search engine/ ChatGPT).
- I believe that my personal identity remains unknown to others when I use the chatbot (search engine/ ChatGPT).
- I am easily identified as an individual by others when I use the chatbot (search engine/ ChatGPT) (Reverse coded).
- Others are likely to know who I am when I use the chatbot (search engine/ChatGPT) (Reverse coded).
- My real personal identity can be guessed or known by people when I use the chatbot (search engine/ ChatGPT) (Reverse coded).

Perceived trust [341]:

- The system is deceptive.
- The system behaves in an underhanded manner.
- I am suspicious of the system's intent, action, or outputs.
- I am wary of the system.
- The system's actions will have a harmful or injurious outcome.
- I am confident in the system.

- The system provides security.
- The system has integrity.
- The system is dependable.
- The system is reliable.
- I can trust the system.
- I am familiar with the system.



## APPENDIX C: SCREENING TOOL FOR DV

HITS screening tool [342]

How often does your partner:

- physically hurt you?
- insult you or talk down to you?
- threaten you with harm?
- scream or curse at you?

5-point frequency format: never, rarely, sometimes, fairly often, frequently. A score of 1-5 is assigned to each item. Therefore, the score ranges from 4 to 20. Positive scores are those greater than 10.

#### APPENDIX D: INTERVIEW PROTOCOL FOR STUDY 2

The objective of the interview: The goal of this interview is to discuss your perception and experience with SafeHaven and online search/or ChatGPT for supporting domestic violence survivors.

Let's start with the questions now.

- In your opinion, what are the advantages of using SafeHaven compared with the search engine (or ChatGPT) when seeking information or support about domestic violence?
- What are the disadvantages of using SafeHaven compared with the search engine (or ChatGPT) when seeking information or support about domestic violence?
- Would you/or someone you know prefer to use SafeHaven or a search engine (or ChatGPT) in the future when seeking information or support about domestic violence? Why?
- How difficult was it to complete tasks using SafeHaven compared with a search engine (or ChatGPT)? Did you encounter any technical difficulties when using SafeHaven?
- In general, what do you think about using a chatbot for domestic violence intervention as compared to other forms of technology (such as mobile apps or social media)?
- What are some domestic violence situations that can benefit more from using a chatbot than a search engine (or ChatGPT)?

- In your opinion, how can chatbots be used to promote long-term healing and recovery processes for domestic violence survivors, beyond just providing immediate support?
- What privacy or security concerns do you have when using SafeHaven to search for help or information regarding domestic violence?
- What privacy or security concerns do you have when using search engines (or ChatGPT) to search for help or information regarding domestic violence?
- What are your recommendations or things that need to be improved on SafeHaven?

## APPENDIX E: COGNITIVE WALKTHROUGH: AN ILLUSTRATION

Survey

Task: Find a DV shelter

	Did you expect to achieve the desired outcome?		Did you recognize that the desired action is available?		Did you connect the description of the desired action to what was being attempted?		Have you observed progress toward solving the task if the desired action is performed?		Action Success (S) / Failure (F)	
	Yes	No	Yes	No	Yes	No	Yes	No	S	F
Action: Identify where to click/write to find a shelter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Action: Click the designated button/ or write a text to find a shelter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Action: Share the location	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Action: View a list of nearby shelters	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Actions Notes: