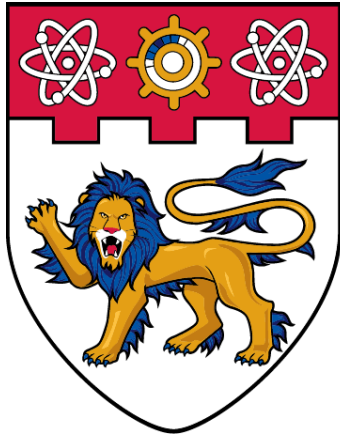




**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**Exploring the use of conversational agents in healthcare:
an overview, conceptual analysis and case study**

Dhakshenya Ardhithy Dhinakaran

Lee Kong Chian School of Medicine

(2022)

Exploring the use of conversational agents in healthcare: an overview, conceptual analysis and case study

Dhakshenya Ardhithy Dhinakaran

Lee Kong Chian School of Medicine

A thesis submitted to the Nanyang Technological University in partial fulfilment of the requirements for the degree of Doctor of Philosophy (2022)

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

13/01/2022

Date

Dhakshenya

Dhakshenya Ardhithy Dhinakaran

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

13/01/2022

Date

NTU NTU NTU NTU NTU NTU NTU NTU
NTU NTU NTU NTU NTU NTU NTU NTU
Lorraine Tudor Car
NTU NTU NTU NTU NTU NTU NTU NTU
NTU NTU NTU NTU NTU NTU NTU NTU
Lorraine Tudor-Car

Authorship Attribution Statement

The thesis contains materials from three papers published in the peer-reviewed journals listed below. I was the first and/or corresponding author for these publications.

Chapter 4 is published as Lorainne Tudor Car, Dhakshenya Ardhithy Dhinakaran, Bhone Myint Kyaw, Tobias Kowatsch, Shafiq Joty, Yin-Leng Theng, Rifat Atun. Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. J Med Internet Res 2020. Available from: DOI: 10.2196/17158.

The contributions by co-authors were as follows:

- Asst. Prof Lorainne Tudor-Car conceptualized the overall study; obtained the funding; supervised the team; provided critical input into all stages of the study and supported the writing of the manuscript.
- I (DAD) planned the search strategy, prepared the study protocol, screened articles, extracted and analysed data, and wrote the manuscript.
- Dr. Bhone Myint Kyaw (BK) screened the articles and extracted and analyzed the data.
- Asst. Prof Tobias Kowatsch (TK), Asst. Prof Shafiq Joty (JR), Prof Rifat Atun (RA), and Prof Yin-Leng Theng (YLT) provided guidance and revised the manuscript critically.

Chapter 6 is published as Dhakshenya Ardhithy Dhinakaran, Thirunavukkarasu Sathish, Tobias Kowatsch, Konstadina Griva, James Donovan Best, Lorainne Tudor Car. Public Perceptions of Diabetes, Healthy Living, and Conversational Agents in Singapore: Needs Assessment. JMIR Form Res 2021. Available from: DOI: 10.2196/30435.

The contributions by co-authors were as follows:

- I (DAD) performed the interviews, conducted the analysis, wrote and revised the manuscript.
- Dr. Thirunavukkarasu Sathish (TS) provided inputs and supervision on data analysis
- Asst. Prof Tobias Kowatsch, Assoc. Prof Konstadina Griva and Prof James Donovan Best provided guidance and revised the manuscript critically.
- Asst. Prof Lorainne Tudor-Car conceptualized the overall study; obtained the funding; supervised the team; provided critical input into all stages of the study and critically reviewed the draft manuscript.

Chapter 7 is published as Dhakshenya Ardhithy Dhinakaran, Thirunavukkarasu Sathish, Aijia Soong, Yin-Leng Theng, James Best, Lorainne Tudor Car. Conversational agent for healthy lifestyle behavior change: an online feasibility study. JMIR Form Res 2021. Available from: DOI: [10.2196/27956](https://doi.org/10.2196/27956).

- I (DAD) facilitated and conducted the study, conducted the analysis, wrote and revised the manuscript.
- Dr. Thirunavukkarasu Sathish (TS) provided inputs and supervision on data analysis
- Ms. Aijia Soong (AS) supported the development of the intervention and commented on the manuscript
- Prof Yin-Leng Theng (YLT) and Prof James Donovan Best (JDB) provided guidance and revised the manuscript critically.
- Asst. Professor Lorainne Tudor-Car conceptualized the overall study; obtained the funding; supervised the team; provided critical input into all stages of the study and critically reviewed the draft manuscript.

13/01/2022

Date

ITU NTU NTU NTU NTU NTU NTU NTU
NTU NTU NTU NTU NTU NTU NTU NTU
ITU NTU NTU NTU NTU NTU NTU NTU
ITU NTU NTU NTU NTU NTU NTU NTU

Dhakshenya

Dhakshenya Ardhithy Dhinagaran

Acknowledgments

The work presented in this thesis would not have been possible without the support and guidance of my supervisor, Assistant Professor Lorainne Tudor Car. I'd like to thank her for being such an excellent teacher, for her patience, positivity, encouragement, for giving me the freedom to explore my ideas and new techniques, room to learn new things, and opportunities to present at prestigious conferences. It has been a privilege to have worked with her, learned from her, and to have been in her company for the last four years. This is a relationship I will cherish for a lifetime.

I would like to thank my colleagues Dr. Bhone Myint Kyaw and Miss Ai Jia Soong for supporting my research, as well as myself, during the first few years of my Ph.D. when I needed the most hand-holding. They were always so forthcoming with their help and produced work of such high quality. Thank you to Mauricette and Lidia, my fellow Ph.D. students, for being my companions and an unending source of positivity and support during this Ph.D.

Next, I am so indebted to the family medicine and primary care group at LKC medicine for giving me a beautiful and conducive space to work in for the last four years. Thanks to Prof Helen Smith for being such a solid head of the department.

I would like to extend my gratitude to the Saheki lab for being the first point of support for me at LKC medicine. I'd like to thank Assistant Professor Yasunori Saheki for being my first mentor at LKC and for allowing me to experience research in a wet lab. My lab mates Darshini Jeyasimman, JingBo Sun, Raihanah Harion, and Loges Krshnan for their words of encouragement, guidance in learning new laboratory techniques, and advice, both personal and professional. These acquaintances have been a constant source of support over the last four years and for that, I am very grateful. Thank you.

I would also like to extend my thanks to NTU for granting me the NTU research scholarship which has been such a valuable gift and I believe that the high quality of work I have strived to generate will reflect my gratitude for this grant. Thank you to the graduate office at LKC medicine for being so prompt with administrative details pertaining to assessments, exams, and reimbursement claims.

Finally, I would like to take a moment to acknowledge my family. I have been very lucky to pursue this Ph.D. in Singapore, my hometown. Being able to go home to my family every day has been a blessing as I have received the love, support, encouragement, and all the homecooked meals I've needed to comfort me through this challenging journey!

Table of contents

Statement of originality	3
Supervisor declaration statement	4
Authorship attribution statement	5
Acknowledgments.....	7
Table of contents	8
List of tables	13
List of figures	14
List of abbreviations	15
List of publications	17
Summary	18
Chapter 1 – Conversational agents	20
1.1 Digital health and the relevance of CAs	20
1.2 Defining conversational agents.....	21
1.3 Types of conversational agents	22
1.4 Embodied vs non-embodied conversational agents	24
1.5 Evolution of conversational agents in the literature	24
1.6 Conversational agents in healthcare	26
1.7 Challenges associated with CAs in healthcare	26
1.8 Mitigating these challenges	27
Chapter 2 – Study context	29
2.1 Overview of Singapore	29
2.1.1 Healthcare in Singapore	29
2.1.2 Healthy living challenges in Singapore	30
2.1.3 Digital health ecosystem in Singapore	31
2.2 Aims and objectives	32
Chapter 3 – Methodology	34
3.1 Evidence-based medicine	34
3.2 Mixed methods design	35
3.3 Scoping existing evidence	36
3.4 Conceptual framework methodology	37

3.5 Implementation science	38
3.6 Pilot feasibility study	39
3.7 Qualitative approaches	39
Chapter 4 –Scoping review and conceptual analysis	42
4.1 Introduction	42
4.2 Methods	43
4.2.1 Search Strategy	43
4.2.2 Inclusion and exclusion criteria	43
4.2.3 Screening, data extraction, and analysis	43
4.3 Results	44
4.3.1 Search findings	44
4.3.2 Characteristics of included studies	45
4.3.3 Characteristics of CAs in included studies	47
4.3.4 Conversational agent content analysis	50
4.3.5 Conversational agent evaluation	52
4.4 Discussion	68
4.4.1 Key Findings	68
4.4.2 Comparisons with existing literature	68
4.4.3 Strengths and limitations	70
4.4.4 Conclusion	70
Chapter 5 – Conceptual framework	71
5.1 Introduction	71
5.1.1 Objective	72
5.2 Methods	72
5.2.1 Mapping the data sources	72
5.2.2 Literature review of mHealth frameworks	72
5.2.3 Scoping review of CAs	73
5.2.4 Deconstructing, integrating and synthesizing concepts	73
5.2.5 Validating the framework	73
5.3 Results	74
5.3.1 Literature review of mHealth frameworks	74
5.3.2 Literature review of rule-based CAs	82

5.3.3 Conceptual framework	88
5.3.4 Validation study	101
5.4 Discussion	103
5.4.1 Key findings	103
5.4.2 Rule-based and AI CAs	103
5.4.3 Comparisons with existing literature	103
5.4.4 Implications for future research	104
5.4.5 Strengths and limitations	105
5.4.6 Conclusions	106
Chapter 6 – Needs Assessment	107
6.1 Introduction	107
6.1.1 Objectives	108
6.2 Methods	108
6.2.1 Participant and recruitment	108
6.2.2 Data collection	109
6.2.3 Data analysis	111
6.3 Results	111
6.3.1 Participant demographics	111
6.3.2 Thematic analysis	112
6.4 Discussion	125
6.4.1 Key findings	125
6.4.2 Comparisons with existing literature	126
6.4.3 Implications for future work	127
6.4.4 Strengths and limitations	129
6.4.5 Conclusions	130
Chapter 7 – Conversational Agent pilot feasibility study, Singapore	131
7.1 Introduction	131
7.1.1 Objective	132
7.2 Methods	132
7.3 Results	139
7.4 Discussion	147
7.4.1 Principal findings	147

7.4.2 Comparisons with existing literature	147
7.4.3 Implications and future research	148
7.4.4 Strengths and limitations	149
7.4.5 Conclusions	151
Chapter 8 – Follow up interview qualitative analysis	152
8.1 Introduction	152
8.1.1 Objectives	153
8.2 Methods	153
8.2.1 Participants and recruitment	153
8.2.2 Data collection	154
8.2.3 Data analysis	155
8.3 Results	155
8.3.1 Participant demographics	155
8.3.2 Thematic analysis	156
8.3.3 Trustworthiness in qualitative research	167
8.4 Discussion	168
8.4.1 Key findings	168
8.4.2 Comparisons with existing literature	168
8.4.3 Main recommendations	169
8.4.4 Strengths and limitations	171
8.4.5 Conclusions	172
Chapter 9 – Systematic assessment of healthy living conversational agent apps	173
9.1 Introduction	173
9.1.1 Objectives	174
9.2 Methods	174
9.2.1 Inclusion and exclusion criteria	175
9.2.2 Assessment criteria	175
9.2.3 Data analysis	176
9.3 Results	177
9.3.1 App selection	177
9.3.2 General app characteristics	178
9.3.3 Individual app analysis	179

9.3.4 mHON code guidelines	185
9.4 Discussion	186
9.4.1 Key findings	186
9.4.2 Comparisons with existing literature	187
9.4.3 Strengths and limitations	189
9.4.4 Conclusions	190
Chapter 10 – Discussion and conclusions	191
10.1 Aims of the thesis	191
10.2 Summary of findings and contributions	192
10.3 Comparisons with existing literature	195
10.4 Impact of this work	197
10.5 Potential implications on policies in Singapore	197
10.6 Potential challenges to wide adoption	198
10.7 Limitations and methods for improvement	200
10.8 Impact of the research on the researcher	201
10.9 Quality and trustworthiness of the qualitative research	202
10.10 Recommendations for the future	208
10.11 Concluding remarks	209
References	210
Appendices	255

List of Tables

Table 4.1 Personality codes for the conversational agents, adapted from Haan *et al.*

Table 4.2 Characteristics of studies included in scoping review

Table 5.1 mHealth, text messaging, and smartphone app frameworks and their key features.

Table 5.2 Conversational agent frameworks and their key features.

Table 5.3 Characteristics of studies on rule-based conversational agents

Table 6.1 Basic demographic information of interview participants

Table 6.2 Recommendations on content and delivery for future CAs for healthy lifestyle behavior change

Table 7.1 Overall summary of results from pilot study questionnaires

Table 7.2 Characteristics of all the enrolled participants who completed baseline assessments for the pilot study

Table 8.1 Basic demographic information of interview participants

Table 9.1 Characteristics of included CA apps from 42 matters database search

Table 9.2 The BCTs and personalities used by each app and CA to achieve the healthy living aim

Table 10.1 Quality criteria to ensure trustworthiness in qualitative research

Table 10.2 Reflexivity diary for qualitative studies

List of Figures

Figure 1.1 An example of a text-based conversational agent

Figure 1.2 Demonstration of conversations with conversational agent ELIZA

Figure 1.3 Outline of chatbots and their corresponding year of development

Figure 4.1 Scoping Review PRISMA flowchart

Figure 4.2 Bubble plots showing results from scoping review

Figure 4.3 Type of input in CAs from scoping review

Figure 4.4 Type of output in conversational agents from scoping review

Figure 5.1 PRISMA flowchart of mHealth development frameworks search

Figure 5.2 Key steps prioritized in the collated mHealth frameworks

Figure 5.3 PRISMA flowchart for literature on rule-based healthcare CA development

Figure 5.4 A five-step conceptual framework for rule-based healthcare CA development with two overarching themes.

Figure 6.1 The order of events participants underwent, from (1) recruitment, (2) interviews for needs assessment and subsequent participation in a 4-week feasibility study.

Figure 6.2 The four main themes and subthemes derived from the needs assessment interviews with 20 participants.

Figure 7.1 Example of conversations between user and CA *Precilla*.

Figure 7.2 Overview of feasibility study workflow from screening to follow up.

Figure 7.3 Numbers of participants who were initially recruited, completed baseline and follow up assessments for feasibility pilot study.

Figure 7.4 Usability scores for *Precilla*

Figure 7.5 Number of complete, incomplete, and absent interactions during the 4-week feasibility study period.

Figure 7.6 Number of immediate and delayed responses for the 16 interactions over 4 weeks with *Precilla*.

Figure 8.1 The main five themes and sub-themes identified from follow up interview transcripts.

Figure 8.2 Main recommendations on how to enhance a healthy living conversational agent

Figure 9.1 Flowchart for systematic app selection in Android and iOS stores

Figure 9.2 The 20 BCTs used and associated frequencies in the included apps

Figure 9.3 Percentage compliance of included apps with mHON code guidelines

Figure 10.1 Reflexive considerations for studies in this PhD project

List of Abbreviations

CA: conversational agent
AI: artificial intelligence
HIC: high-income country
TDV: technical development
HS: health care support
US: Unites States
EDU: education
IQR: interquartile rate
BC: behavior change
N/A: not applicable
LMIC: low- and middle-income country
T&M, treatment, and monitoring
DG: diagnosis
RCT: randomized controlled trial
CBT: cognitive behavioral therapy
UK: United Kingdom
MOL: methods of levels
PHQ-9: patient health questionnaire-9
FS: Flourishing scale (for wellbeing)
STI: sexually transmitted infection
mASMAA: mobile phone-based asthma self-management aid for adolescents
A1c: glycated hemoglobin. Indication of blood sugar levels
COREQ: Consolidated criteria for reporting qualitative research
FFQ: food frequency questionnaire
MET: metabolic equivalent of task
PSQI: Pittsburgh Sleep Quality Index
PSS: perceived stress scale
QOL: quality of life
REDCAP: research electronic data capture
SD: standard deviation
CI: confidence interval
HCP: healthcare professional
HIV/AIDS: Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome
STIs: Sexually transmitted infections

EMAs: Ecological momentary assessments

mHON: Mobile Health On Net (mobile app certification guidelines)

COVID-19: Coronavirus disease of 2019

List of Publications

Published papers

Car, L.T., **Dhinakaran, D.A.**, Kyaw, B.M., Kowatsch, T., Joty, S., Theng, Y.L. and Atun, R., 2020. Conversational agents in health care: scoping review and conceptual analysis. *Journal of medical Internet research*, 22(8), DOI: [10.2196/17158](https://doi.org/10.2196/17158).

Dhinakaran, D.A., Sathish, T., Kowatsch, T., Griva, K., Best, J.D. and Car, L.T., 2021. Public Perceptions of Diabetes, Healthy Living, and Conversational Agents in Singapore: Needs Assessment. *JMIR Formative Research*, 5(11), DOI: [10.2196/30435](https://doi.org/10.2196/30435)

Dhinakaran, D.A., Sathish, T., Soong, A., Theng, Y.L., Best, J.D. and Tudor Car, L., 2021. Conversational agent for healthy lifestyle behavior change: an online feasibility study. *JMIR Formative Research*, 5(12), DOI: [10.2196/27956](https://doi.org/10.2196/27956)

Submitted papers

Dhinakaran, D.A., Martinengo, L. and Car, L.T., 2022, Designing and implementing a mobile phone delivered rule-based conversational agent: a conceptual framework. *Journal of medical Internet research*.

Dhinakaran, D.A. and L. and Car, L.T., 2022, Public perception of a healthy lifestyle change conversational agent “Precilla” in Singapore: A qualitative study. *BMJ Health & Care Informatics*.

Poster presentations

Dhinakaran, D.A., Car, L.T. A framework for designing & implementing a rule-based conversational agent in healthcare. STEMM MIT CSAIL AI in healthcare summit 2020. Virtual Summit. [Poster presentation]

Dhinakaran, D.A., Car, L.T., Kyaw, B.M., Kowatsch, T., Joty, S., Theng, Y.L. and Atun, R., Conversational agents in healthcare: A scoping review and conceptual analysis. STEMM MIT CSAIL AI in healthcare summit 2020. Virtual Summit. [Poster presentation]

Dhinakaran, D.A., Car, L.T. A framework for designing and implementing a rule-based conversational agent (chatbot) in healthcare. Evidence and Implementation Summit (EIS 2021), Sydney, Australia. Centre for Evidence and Implementation & Monash University. Virtual Summit. [Poster presentation].

Dhinakaran, D.A., Car, L.T., Kyaw, B.M., Kowatsch, T., Joty, S., Theng, Y.L. and Atun, R., Conversational agents in healthcare: A scoping review and conceptual analysis. Evidence and Implementation Summit (EIS 2021), Sydney, Australia. Centre for Evidence and Implementation & Monash University. Virtual Summit. [Poster presentation].

Awards

Dhinakaran, D.A., Thirunavukkarasu S, Soong A.J., Theng Y.L., Best J.D., Car, L.T. Conversational agent for healthy lifestyle behaviour change: an online feasibility study. Singapore Health & Biomedical Congress 2021, Singapore. [Oral Presentation]. SHBC Student Awards (Open Category). Merit Award.

Summary

Conversational agents (CAs) or chatbots are computer programs designed to participate in human-like conversations with their users. The growing interest in these agents is evident in their employment in numerous fields including business, retail, and even healthcare. The need for new initiatives to promote healthy living is apparent in Singapore given its highly stressful and fast-paced way of life, insufficient time for physical activity, and frequent dining out. Singapore is technologically astute and always on the lookout for novel initiatives, such as the use of CAs, to improve patient care and support healthcare staff.

This project was grounded in the underlying principles of evidence-based medicine and employed a multitude of study designs and methodologies. Firstly, a (1) scoping review was conducted to map out and evaluate the existing evidence on CAs in healthcare. Secondly, a (2) conceptual framework for rule-based healthcare CA design and development was proposed based on existing mHealth and CA development frameworks. Thirdly, a (3) healthy lifestyle change CA was developed and piloted based on guidance from the proposed conceptual framework and (4) co-design/needs assessment with members of the public through a series of semi-structured interviews. (5) Follow-up interviews were then conducted with participants from the pilot study who reported on their experience interacting with the healthy living CA and their recommendations for its improvement in the future. Finally, a (6) systematic assessment of CA apps was performed in Google Play (android) and Apple (iOS) stores to map out and assess the characteristics and behavior change techniques used in those healthcare CAs currently available to the public for healthy lifestyle behavior change.

The results indicated that this field is still in its nascent stages with a predominance of small case studies and a dearth of large-scale randomized controlled trials. Most of the CAs reported in the literature were standalone smartphone apps that were text-based and employed machine learning. The conceptual framework used existing frameworks as a foundation and culminated in a five-step framework guiding the user through defining the aim, creating an identity, developing the content, building the conversation flow, and piloting the CA. The CA designed and developed for a healthy lifestyle change, *Precilla*, aimed to include those recommendations made by volunteer participants in the needs assessment interviews. Their suggestions included providing educational information on diabetes and prediabetes, risk factors, and risk prevention. They also suggested a friendly personality and simple, layman terms to be used in the conversations with *Precilla*. Some participants were very satisfied with *Precilla*, found her informative, and thought the information shared was very valuable and a good reminder to make healthier choices with

regards to healthy eating, physical activity, sleep, and stress. Conversely, others were disappointed in the lack of artificial intelligence and would have liked to have the opportunity to “chit-chat” with Precilla without any specific endpoint. One of the main recommendations for improvement suggested introducing some personalization to the CA in terms of more tailored content and message delivery timings. Incentives such as rewards and point collection were also of interest and participants had high regard for formal and informal group support in enabling them to achieve their healthy lifestyle change goals. Finally, the systematic assessment of healthy living apps from the app stores indicated that the pool of CA apps in this area is fairly small. The apps focus on just one area of healthy living either diabetes prevention, wellness and physical activity, sleep improvements, or healthier eating. The apps were all free with only a couple offering in-app purchases and no advertisements were featured. CA personality was sometimes neglected in most apps which came across as very rigid, factual, and not human-like.

This Ph.D. project has made several valuable contributions such as a timely update of the current progress and status of this field of research on healthcare CAs through the review and synthesis of evidence mapped in the scoping review and apps assessment. It has also proposed some guidelines on CA development which currently have not been strictly defined. Finally, a test of the feasibility and acceptability of a healthcare conversational agent for healthy living in Singapore revealed the applicability and/or necessity of such an agent and could propose future potential applications in Singapore’s healthcare system. This project has laid a good foundation for future research to tap into.

Chapter 1 – Introduction

This thesis opens with a series of introductions to the concepts grounding this PhD project. In the introductory chapter, the reader is presented with the definitions for digital health and conversational agents, the relevance of these to healthcare and their evolution over time, in the literature. What follows is a setting of the scene for the geographic context in which this project was conducted. Here, Singapore is introduced, along with the way healthcare is conducted, the key areas of healthcare challenges and how digital health fits into this environment. Next, the various forms of methodology employed in the thesis are outlined, with a short description of their applicability and necessity. The next six chapters form up the main crux of the thesis, describing the studies undertaken, the findings and the conclusions drawn from them. The thesis then concludes with a discussion chapter which consolidates the findings from the studies, discusses the implications of the findings and presents a critical appraisal of the PhD project.

Chapter 1 provides an introduction to the thesis by first introducing digital health and CAs in Section 1.1. Sections 1.2, 1.3, and 1.4 define CAs, explore the types of CAs that exist, and highlight the distinction between embodied CAs (ECAs) and Non-ECAs, respectively. In Section 1.5, the evolution of CAs in the literature is presented. Sections 1.6-1.8 discuss CAs in healthcare specifically, the challenges associated with their widespread usage, how these issues might be addressed and the relevance of this Ph.D. project.

1.1 Digital health and the relevance of CAs

Digital health can be defined (by the US Food and Drug Administration) as those technologies which use “computing platforms, connectivity, software, and sensors for health care and related uses” [1]. These technologies can range from promoting lifestyle changes to electronic health devices.

Digital health research has been around since the 1970s when digital means were used to support biological research in the form of bioinformatics and computational genomics [2]. Since then, technology usage has only grown in healthcare as evidenced by electronic health records (EHRs), electrocardiogram monitoring devices, and more recently, telecommunication with HCPs and health tracking apps. The maturing digital health arena is supported by the increased uptake of digital health solutions by both patients and physicians [2]. *Oncoassist*, for example, is a smartphone-based tool that supports oncologists in their selection of available treatments. Similarly, *SideKick Health*, employs gamification and behavioral models to promote engagement and positive health outcomes in patients with a range of chronic conditions. It appears that multiple patient-facing digital health products support behavior and lifestyle changes either on

their own or in conjunction with other treatments to improve patient outcomes [2]. Due to the wide reach possible with digital health, in the future, digital health would most logically aim to support population-level prevention over treatment by promoting health and lifestyle changes.

To further reinforce the idea that patients and the public are keen on digital health, a report by Accenture in 2019 indicated that consumers were more likely to select those HCPs who had digital capabilities [3]. This was evidenced by an obvious increase from 2016 to 2019 in the percentage of individuals who employed these digital health services: requested prescription refills electronically (67% in 2016, 77% in 2019), asked for reminders through emails or text messages for follow up or preventative care (57%, 70%), communicated with the HCP via email (53%, 69%), or video conferencing (36%, 49%), booked or changed appointments online (58%, 68%) and used remote devices to track their health (39%, 53%) [3].

Within digital health, mobile health (mHealth) is concerned with the use of smartphone devices to deliver healthcare services, for example, in the form of education, appointment booking, remote monitoring or even teleconsultations [4]. Conversational agents (CAs) in particular reduce the need for face-to-face consultations and resultantly reduce exposure to air-borne or nosocomial infections, which individuals are at risk of acquiring when frequenting healthcare institutions for non-essential purposes [5]. Furthermore, they enable the maintenance of frequent communication between patient and HCP as they are easily accessible from the comfort of one's home through smartphones which are ubiquitous today [5]. CAs pose the additional benefit of enabling healthcare services to be tailored to an individual's personalised needs and preferences. In addition, the semi-automated (or fully automated) nature of the conversation greatly alleviates some stress from the HCPs as they are not required for every single interaction [5].

As such, the communicative nature of CAs makes them highly beneficial agents for situations targeting communication, feedback, monitoring, and remote healthcare delivery.

1.2 Defining conversational agents

Conversational agents (CAs) are sometimes referred to as chatbots or voice/virtual assistants or and are computer programs designed to re-enact human-human conversations with users [6]. They have been successfully employed in many different areas namely finance, retail, marketing, and fitness promotion [7]. CAs can be implemented via different platforms such as smartphone apps, through existing messaging platforms, websites, or even individual devices such as speakers. They can communicate through different modalities using text, images, or speech. CAs that use speech as the mode of input and output are also known as voice assistants. These agents depend on speech recognition technology (amongst other technicalities) to fulfil their functions.

Some popular examples include Siri, Alexa, Cortana, and Google assistant, and these function through smart speakers. They are increasingly popular and utilized to perform remote tasks such as online shopping, entertainment, news or to control smart home devices [8-10].

Conversely, there are other CAs that communicate predominantly via text, sometimes supplemented with emojis, images, videos, or other visual aids, to interact with the user (Figure 1.1). This Ph.D. project focuses on text-based conversational agents.

1.3 Types of conversational agents

CAs can be categorized in different ways, with “smart” and “simple” agents being one mode [7]. Simple CAs are sometimes referred to as “rule-based” as their conversations are dictated by pre-defined keywords and rules programmed by the developer. The user is restricted to a finite number of answer options presented to them and there is very little opportunity for free flow in the conversation.

Conversely, smart CAs are not envisioned to be as rigid, and users can have a more free and easy conversation that is not contained by pre-written rules. This is enabled by artificial intelligence (AI), the ability of computer systems to perform processes normally requiring human intelligence [11].

AI CAs are often preferred by users because of their promise for free flow, slick conversations which makes the user feel like they are having an interaction with another human being. However, when looked at more closely, AI CAs are not always able to live up to the expectations of the user. What’s more, in some cases, the complexity involved in reaching the response generated by the AI agent may be so convoluted that the developers themselves are unable to explain it. This phenomenon is known as the “Black box effect” and can be especially dangerous with vulnerable populations in healthcare if miscommunication occurs and the CA generates a response, not in the user’s best interest [12]. Research has also indicated that some of the common reasons behind the failure of mHealth and health apps are a lack of evidence-based, transparent, and step wise approaches in their design [13]. As appealing and as promising as AI CAs are for all purposes, including healthcare, plenty of work still needs to be done to gain a clearer picture of the safety of using such agents for healthcare.

It is also interesting to note that some of the most successful mHealth applications employ mobile phones for simple purposes such as text messaging or voice calls for example, in smoking cessation programs [13, 14]. It has been posited that healthcare researchers may sometimes be over-engineering for the intended purpose and AI may not be a requirement for every healthcare intervention [13-15]. A potentially safer and effective alternative is rule-based CAs. They can address some of the limitations of the older mHealth methods of text messaging-based interventions by overcoming barriers associated with cost, user preferences, easy availability, low barriers to adoption due to high smartphone usage, possibility to integrate other functionalities than just a chat (such as health tracking, reminders, etc.) and real-time bidirectional communication. These advantages are all possible whilst also avoiding the severe challenges associated with AI technology for healthcare.

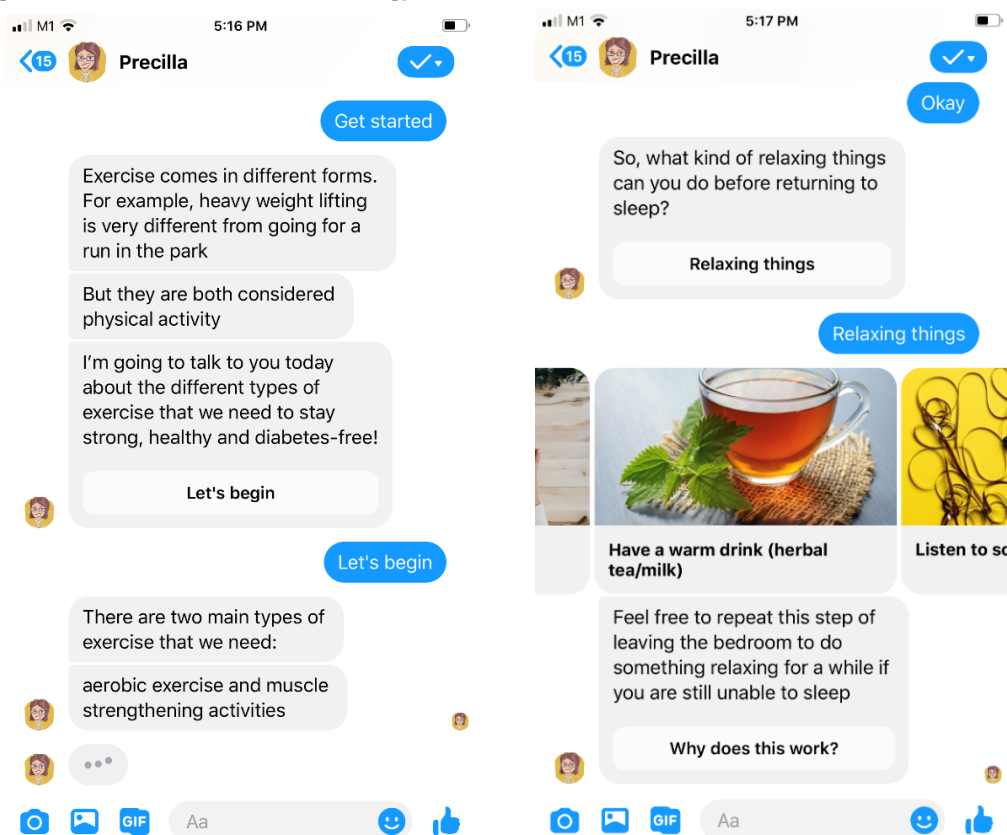


Figure 1.1 An example of a text-based conversational agent

Microsoft published a report highlighting an alternative way to categorize chatbots, mainly as goal-oriented or chit-chat bots [16, 17]. Goal-oriented conversational agents are designed to be very focused and to help the user achieve a particular goal or complete a task [17]. On the flip side, chit-chat bots are much more general and do not have one area of focus. They have the main aim of keeping the user engaged and conversing with them for as long as possible. They do this by sharing jokes, memes, chatting, etc., and don't tend to delve very deep into any one particular topic [17]. Eliza, the first chatbot developed in the 1960s to mimic a Rogerian psychotherapist was a chit-chat-based conversational agent (Figure 1.2) [16].

```
Welcome to

EEEEEE LL      IIII ZZZZZZZZ AAAAA
EE      LL      II      ZZ  AA  AA
EEEEEE LL      II      ZZZ  AAAAAA
EE      LL      II      ZZ  AA  AA
EEEEEE LLLLLL IIII ZZZZZZZZ AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:   █
```

Figure 1.2 Demonstration of conversations with conversational agent ELIZA [16]

1.4 Embodied vs non-embodied conversational agents

CAs can be embodied or unembodied. Embodied conversational agents or ECAs are those which present themselves as avatars to users. In other words, there is a digital representation of a human which the user can see when they communicate with the CA [18]. Thus, the user experience involves the exchange of text and/or speech as well as non-verbal cues like facial expressions from the ECA. Unembodied CAs lack this avatar so, the conversation is restricted to just the text and/or speech that is transmitted between the user and CA. The CAs discussed in this thesis were predominantly unembodied.

1.5 Evolution of conversational agents in the literature

The first CA “ELIZA” was the brainchild of Joseph Weizenbaum who introduced the world to a person-centered Rogerian psychotherapist (Figure 1.2) [19]. This was a ground-breaking discovery especially for AI and the CA showed some success with the patients it conversed with [20]. Following ELIZA was CA “PARRY”, which modelled a paranoid schizophrenic [21, 22]. These

CAs were known as “chatterbots” at the time of development (1966, 1972, respectively) and demonstrated the ability of virtual agents to recreate human-human conversations. At their best, they could even pass the Turing test. The Turing test assesses the artificial intelligence of a machine whereby under specific conditions, the assessor is unable to distinguish it from an actual human [23].

In the following few decades, the term “talking computers” was used to discuss and present new CAs [24-28]. These were a simpler version of the CAs we see today and their duties involved performing patient satisfaction surveys [24], healthy eating behavior change [25], diagnostic support [26], and streamlining patient-physician communication [27]. The advent of Apple’s “Siri” in 2010 was the next noteworthy step for CAs (Figure 1.3). The attention directed to CAs increased rapidly as Google now, Alexa and Cortana were all released closely after Siri as representatives of Google, Amazon, and Microsoft, respectively [29].

The year 2016 earned the title “Year of the Chatbot” as many major IT companies like Facebook, Google, LinkedIn and Viber started to advocate for CAs too [30-32].



Figure 1.3 Outline of chatbots and their corresponding year of development to indicate the evolution of conversational agents in the past decade.

More recently, health-specific conversational agents like Babylon health have become a mainstay in the UK’s NHS system as well as in less developed countries like Rwanda because of its promise for remote healthcare for those in need [33]. Woebot is one of the many mental health chatbots that has been addressing the growing issue of anxiety and depression which is of global importance at the moment [34]. Lark is another a more recently developed healthcare conversational agent designed to target many facets of healthy living [35].

1.6 Conversational agents in healthcare

Globally, 80.63% of the population owns a smartphone and as of October 2021, two billion individuals were using the most popular messaging app – WhatsApp [36, 37]. To emphasize the point that global internet users are spending the majority of their time online on messaging apps, the Statista Research Department revealed that 90.7% of global internet users were accessing chat apps [36]. The rising tendency for individuals to communicate via messaging apps supports the notion that CAs may be preferably adopted for a variety of applications, including healthcare, given their familiarity with messaging apps like WhatsApp, Telegram, and Facebook Messenger.

Furthermore, the healthcare industry may be a suitable candidate for CA-delivered interventions given its success with a simpler counterpart – text messaging. CAs allow for bidirectional communication and their flexibility in input/output modalities make them attractive to several target populations ranging from children to older adults. The use of mobile phone messaging for healthcare has been present and increasingly adopted in healthcare research since 2002 [38]. This is evidenced by numerous systematic reviews exploring text messaging as an intervention and showing its success and acceptability for multiple health conditions [39, 40]. However, text messaging, although proven to be effective, has its shortcomings, such as minimal features, associated costs and only a limited number can be issued as dictated by mobile plans. A potential alternative is messenger conversational agents. Although applications of conversational agents in healthcare are gaining traction in several medical fields, (eg. healthcare service provision, chronic disease management, patient education etc.) research into this area is still in its infancy [5]. For example, a systematic review by Laranjo et al. identified 17 healthcare CAs with the ability for unrestrained natural language input and concluded that CAs are still in the nascent stages of optimisation as evidenced by a predominance of quasi-experimental studies which did not always report on efficacy and safety [41]. As such, the literature indicates a tacit need for more research to be conducted around the design, development, evaluation and optimisation of CAs in healthcare. Furthermore, in a scoping review conducted in our research team, it was noted that there is an increasing trend for healthcare CAs to evolve from performing simple transactional tasks to more involved process with a longer-term goal in mind [42]. As such, more research needs to be conducted into the evolution of CA usage applications to capitalise on their strengths, to explore their reception by target populations and to establish some consistency in the guidelines for their development and evaluations.

1.7 Challenges associated with CAs in healthcare

Having said that, CA usage in healthcare doesn't come without its challenges. As healthcare deals with vulnerable populations, safety is a major concern with any form of mHealth. The relevance of this issue to CAs was publicly reported when a physician in the NHS raised his concerns about

the safety of the triage and diagnoses chatbot “Babylon” [43, 44]. The CA which was reported to have undergone robust clinical testing and was backed by venture capitalists, appeared to underperform in its accuracy of diagnoses [43]. The physician specifically cited a case where the CA misdiagnosed a critical state of chest pain as gastritis or sickle cell crisis [44]. This also raised the issue around the ethics of mis promoting a medical device which still has its flaws and has yet to be optimised for what it claims to be an expert in [43].

Another challenge associated with CAs in healthcare is that they are not suitable for all usage applications. In a survey study of 100 practising physicians in the US, they believed that the inability for CAs to display human emotion deemed them unfit to cater for all of a patient’s needs [45]. Furthermore, CAs were thought to be unsuitable for diagnosis and treatment as they are unable to perform a thorough examination of the patient which involves physical examination in addition to reviewing a medical history [45]. Physicians most consistently agreed that CAs would be beneficial for administrative tasks, directing users to health clinics and providing medical information [45].

1.8 Mitigating these challenges

These issues around safety, ethics and the limits of CAs can be mitigated by avoiding over-provision. Researchers often tend to seek out the latest gadgets and software when designing mHealth interventions without realising that that level of complexity is truly not needed to make a meaningful change to the user’s health. Text messaging interventions have demonstrated this very well [14, 15]. Furthermore, as beneficial as CAs can be, they will not single handedly become the be-all, end all solution for any healthcare initiative. They do not aim to replace HCPs but rather, CAs can help to (1) alleviate the burden a little from HCPs (eg. by automating patient reminders, booking appointments, efficient information transfer etc.), (2) support patients in navigating the healthcare system (keep track of appointments, ask questions about health prevention, be guided to helpful resources online), and (3) empower them to manage their own health [5]. Pragmatically, CAs could be well integrated into a blended programme where a CA supports pre-existing face-to-face appointments so as not to overstep into a person’s life before enough research has been done to establish healthcare CAs as safe and respectful of boundaries. Using CAs to complement one’s healthcare plan (instead of being a standalone intervention) can also reduce uptake barriers by easing the digital agents into an individual’s lifestyle gradually. Secondly, when exploring CA usage in a new population, the intervention targeted can be broad and relevant to the majority. This means the target users piloting the CA will be fairly diverse but will still find the CA relevant. This type of study population can be highly valuable in providing insight into how a healthcare CA will be received in a novel setting and what their strengths and

weaknesses are to help optimise them for future usage applications. CA applications can also be selected to avoid risks associated with safety of patients whilst still maintaining some efficacy.

The following chapter will elaborate on the study setting which informed the CA usage application of focus in this thesis and also helped define the aims and objectives of the project.

Chapter 2 - Study context

Chapter 2 provides an overview of the study setting and its relationship with CAs. Section 2.1 describes Singapore, with specific details on how healthcare works in Singapore (Section 2.1.1), the challenges to healthy living (Section 2.1.2), and the current digital health landscape in the country (2.1.3). The chapter rounds off in Section 2.2 by tying the study setting together with the applicability of CAs for the intended purpose of healthy lifestyle behavior change and presents the aims and objectives of this Ph.D. project.

2.1 Overview of Singapore

Singapore is a small island, situated in Asia, previously a part of “Malaya”. It is well-known as a tiny, yet thriving country often referred to, endearingly, as the “tiny red dot”. The nation prides itself on its multiracial population and spirit of collectiveness which celebrates all four races (Indian, Malay, Chinese, and Eurasian) and their associated religions [46]. Singapore achieved its independence in 1965 and has surpassed plenty of its neighbors, becoming one of the richest, most literate, successful, and productive countries on the planet, housing one of UNESCO’s world heritage sites (the botanic gardens) and the globally acclaimed Marina Bay Sands [47].

2.1.1 Healthcare in Singapore

Singapore has a reputation for its highly efficient and proficient healthcare system in both the private and public sectors. There are two main branches of healthcare: (1) subsidized public healthcare and (2) unsubsidized private healthcare [48]. Unlike some western countries like the USA and the UK where healthcare is heavily subsidized for all or paid for from government funds, healthcare in Singapore is paid for by each individual. Whilst there are financial schemes that offer subsidies for households with lower incomes and mandatory healthcare savings in the form of “Medisave,” citizens are very much responsible for their health and means of paying for their healthcare and treatments [48]. Insurance policies have become incredibly important but are not taken up by all and the government doesn’t foster a culture of over-reliance on third-party insurance schemes but would rather the citizens prioritize healthier living styles [49].

A lack of awareness and education around important health issues has been identified in the country and efforts are being spent on raising awareness in these areas. In 2016, a war was declared against diabetes as Singapore was found to have the second-highest rates of diabetes amongst developed countries [50]. This gave way to the national step challenge which considers behavioral insights and technology to encourage its citizens to track their daily step count [51]. This incorporation of physical activity into their daily lives was positively reinforced with rewards and incentives in the form of point collection [51]. Similarly, hawker centers and other

food courts frequented by patrons have subscribed to the “Healthier dining program” whereby meals are adorned with “healthy option” symbols and occasionally with markings of the calorie count to direct consumers towards healthier food choices [52].

2.1.2 Healthy living challenges Singapore

Whilst such efforts exist, the fast-paced lifestyle in Singapore is still reiterated by the fact that Singapore is the most fatigued country in the world and the high-stress levels are more apparent than ever [53]. As mentioned, Singapore has been placed second-highest for its diabetes prevalence amongst developed countries with 600, 000 adults living with diabetes and its obesity levels are also on the rise [54]. Additionally, 34% of men and 39% of women in Singapore do not reach the weekly target of 150 minutes of moderate-intensity activity per week, leaving people at higher risk of developing conditions like diabetes [55]. People with diabetes live shorter lives, by at least 10 years, and have a lower quality of life compared with those without diabetes [56, 57]. Prediabetes, a precursor stage to diabetes, affected 15.5% of Singaporean adults in 2010 (i.e. one in seven), and this figure is estimated to increase to 24.9% by 2035 [58]. Prediabetes increases the risk of heart disease, and if untreated, over time, the majority transition to diabetes [59]. Research has shown that lifestyle interventions (such as increasing physical activity and eating a healthy diet) delivered by trained healthcare professionals can help to promote healthy weight loss [60-62]. In addition, lifestyle changes for high-risk populations were also directly associated with a reduction in the incidence of type 2 diabetes, accentuating the efficacy of healthy lifestyle changes for diabetes prevention [60].

To support this notion, a WHO report indicated that lifestyle diseases can result from the prolonged engagement in modifiable lifestyle behaviours namely smoking, unhealthy eating and lack of physical activity [63]. This can result in chronic diseases such as heart disease, diabetes, obesity etc. The number of global deaths due to chronic diseases are expected to increase to 70% by 2030 [64]. The World Health Assembly proposed a resolution for the prevention and control of such diseases [63]. It was noted that the amendment of lifestyle factors (such as having a healthy weight, engaging in regular exercise, healthy eating and smoking cessation) could potentially reduce the risk of chronic disease development by as much as 80% [65].

Achieving healthy lifestyle changes independently can be challenging and support from experts such as dietitians or exercise physiologists has been shown to add to the efficacy [60]. However, access to experts at a population level may not be feasible or affordable. A potentially more accessible alternative to in-person support and supervision could be digital health interventions, such as conversational agents.

2.1.3 Digital health ecosystem in Singapore

Singapore is a smart country where Wi-Fi, smartphones, and apps are ever-present [66]. Delivering disease prevention methods through such digital means such as CAs ensures a wider reach, could be more affordable for the layman and the techniques for a healthier lifestyle can be seamlessly integrated into an existing lifestyle if delivered through a smartphone. The government and the citizens are then at liberty to focus on staying healthy rather than investing in sickness.

In addition, Singapore's Ministry of Health has proposed the increasing use of CAs in healthcare in tackling issues such as rising chronic disease burden, and the aging population [67]. They envision a near future where a CA can collect a history from a patient before their consultation, streamlining the doctor's visit and thus cutting down waiting times. This makes Singapore an ideal place for the evaluation of novel mobile health interventions such as conversational agents.

The implementation of CAs for healthcare in Singapore has been of interest for the past few years, spurred on by the announcements at the National Health IT summit in 2018 [68]. Singapore's health technology agency "Integrated Health Information Systems" (iHiS) envisioned a future where unwell Singaporeans could chat with a CA which would direct them to the right hospital and even triage their symptoms. The probable uptake of novel healthcare CAs by the population is encouraged by the widespread interest and use of the smartphone apps and wearable step-trackers for the National Steps Challenge – as mentioned above, an initiative to get the population moving and more physically active [68].

More recently, during the COVID-19 pandemic, the iHiS, in collaboration with SingHealth and the Agency for Science, Technology and Research (A*STAR) developed a CA called "Doctor Covid" [69]. This multilingual CA was created to support the HCPs who were overwhelmed with the sheer number of patients all needed varying levels of care and attention. The CA relieved some of the burden by circulating important reminders and critical updates, conducted clinical and psychosocial surveillance to ensure the wellbeing of patients [69]. Social content was also shared for entertainment and to alleviate some feelings around isolation. The success of the CA and likely acceptability by users was evidence by more than 3000 individuals subscribing to the Telegram CA whilst residing at SingHealth run community care centres [69]. In the near future, "Doctor Covid" is envisioned to become an AI CA which will use machine learning to answer more complex questions related to COVID-19 [69].

Beyond COVID-19 the iHiS aims to trial AI CAs to support the rise in antibiotic resistance by helping to minimise incorrect prescriptions [70]. Additionally, with mental health issues increasing exponentially, Singapore partnered with the health-tech start-up company Touchkin

to bring the well-known chatbot “Wysa” onto the “mindline.sg” website in Singapore for one year of free access for its citizens (since the service usually requires a fee) [71].

Besides healthcare, CAs have been featured in multiple customer service and, food delivery and banking platforms [72]. The CA “Bus Uncle” is well-known amongst Singaporeans as it provides timely updates for commuters on bus timings [73].

Given the widespread exposure to CAs in Singapore and constant discussion about it in the news, it is safe to say that the population is at least conscious or subconsciously acquainted with CAs.

This research is especially timely, as now, more than ever, the public has shown an increased preference to communicate using messaging apps [74]. More than in-person connection and communication, online and specifically messaging app-based conversations are spearheading forwards [74]. The advantages of remote modes of healthcare delivery are also a contributor to their likely uptake. This research is relevant even globally where people in difficult-to-reach regions can gain access to healthcare support remotely and expenses are saved on travel-associated costs. High-risk individuals can be monitored without a feeling of intrusion and judgment can be saved for those with highly stigmatized conditions (esp. mental health or STDs) [75].

Research has indicated that healthcare CAs have the greatest potential in helping patients to navigate health services, seek out healthcare resources and engage in conversations regarding healthy living and positive wellness [5]. In Singapore in particular, CAs then provide the possibility of supporting healthy lifestyle changes which are of high importance given the rising rates of diabetes and barriers to healthy living posed by the urbanized lifestyle in Singapore.

As such, this thesis capitalises on the most promising and risk averse versions of healthcare CAs, namely rule-based healthcare CAs for an application highly relevant in Singapore and deemed appropriate by the limits invoked by physicians and literature – healthy lifestyle change.

2.2 Aims and objectives

The main aim of this Ph.D. project was to develop a healthy lifestyle change conversational agent for the general population in Singapore. This required a review of existing evidence to solidify the health CA landscape and to identify the existing recommendations on how to develop such a CA with relevance and fidelity. Empirical research was also required to understand the views of the study population, study setting and test the CA. This generated qualitative and quantitative feedback which contributes greatly to the evidence base currently lacking in Singapore with regards to health CA research. This research is timely as Singapore strives towards becoming a

smart nation and increases its adoption of digital health solutions such as remote monitoring, telecommunication with HCPs, and health tracking.

The specific objectives were to:

1. Map and analyze the current uses of conversational agents in healthcare and identify knowledge gaps by performing a **scoping review**
2. Propose a systematic list of steps on how to develop a rule-based text healthcare conversational agent and present it in a **conceptual framework**, using evidence from the literature and a validation study
3. Understand the public's perceptions and needs with regards to conversational agents for healthy living and diabetes prevention in Singapore, using **semi-structured interviews** and subsequent **qualitative analysis**
4. Develop and test the feasibility and acceptability of a healthy lifestyle behavior change conversational agent for the general population in Singapore in an **online feasibility study**
5. Understand the public's views on a conversational agent for healthy lifestyle change and obtain their feedback on areas for improvement using **semi-structured interviews** in the follow-up phase of a pilot study
6. Identify and describe the main characteristics and evaluate the reliability, transparency, and BCTs used in healthy lifestyle change conversational agents through a **systematic assessment of apps**

Chapter 3 - Methodology

This chapter discusses the methodology adopted for each of the studies and objectives of this Ph.D. project. Section 3.1 considers the concept of evidence-based medicine and highlights its importance in furthering knowledge and research related to CAs in healthcare. Section 3.2 discusses the rationale for mixed-methods and the chosen paradigms to guide the research. Section 3.3 explains the methods of scoping literature and smartphone app stores to map usage applications of CAs. Section 3.4 discusses the methodology for a conceptual framework. Section 3.5 introduces implementation science and how it has motivated the methodology selected for this PhD project, namely the pilot study design for empirical data collection in 3.6 and the need for qualitative methods in 3.7. The thesis is predominantly written in the third person, except in cases where I (the PhD student) discuss the selection of certain methodologies and elaborate on reflexivity (in Chapters 8 and 10).

3.1 Evidence-based medicine

The foundations of this Ph.D. project are rooted in the theories and methods of evidence-based medicine (EBM). EBM involves the application of the best available evidence along with patients' preferences and clinicians' experience in clinical decision-making [76]. This Ph.D. project aimed to contribute to the evidence base around healthcare CAs to support such decision-making in healthcare.

As medical knowledge grows so rapidly, a need for regular updates in the form of reviews and primary research is necessary to keep the evidence timely and relevant [76]. Considering the nascent nature of this field on healthcare CAs, the literature is still limited. The growing number of relatively small-scale studies and dearth of larger-scale studies reiterates the novelty of this research area and the need for more evidence [41]. A systematic review of conversational agents reported that there are inconsistencies in the reporting of outcomes and that only one study in the review used a randomized controlled trial (RCT) study design [41]. Hence, the need for more EBM-driven research in this area.

The urgent need for EBM in CA research is also supported by the issues raised by a lack of such practice. For example, the Babylon AI chatbot is used widely in the healthcare system in Rwanda however, a lot of controversies have been raised regarding safety issues associated with the chatbot [77]. The lack of rigorous testing, evaluating (with actual patients and not just patient vignettes) and improving, prior to employment was raised as an issue with the Babylon chatbot.

The government and healthcare sector in Singapore have indicated an interest to implement CAs in healthcare [68]. However, the evidence for such implementation is lacking hence, why this research project is necessary. It will contribute heavily to the evidence base around CAs in healthcare in general and specifically in Singapore too.

3.2 Mixed methods & Paradigmatic approach

Mixed method research involves the conduct of both qualitative and quantitative research and combining the strengths of both methods to answer specific research questions [78]. Employing a variety of research methods enables a more comprehensive and reliable set of findings to be presented than just a single method [79]. The mixed-methods approach is especially useful for cross-cultural groups [78], which is appropriate in this context considering the multi-racial nature of the population in the study setting, Singapore.

In an attempt to employ research methodologies which best address the needs of the research project, I (the PhD student) did not weigh the strengths and limitations of the post positivist approach (pure objectivity) with the constructivist approach (pure subjectivity) to pick just one method. Instead, I adopted the stance of employing multiple research paradigms because, as cautioned by Creswell, sometimes these paradigms may be in tension with one another, but “such tension is good [80, 81].” Thus, the amalgamation of post positivism and constructivism in this PhD project resulted in a pragmatic paradigm [80]. Instead of arguing over the incompatibility of quantitative and qualitative research, as Green and Caracelli said, pragmatism celebrates that “the dialectic between opposing ideas can contribute to new insight and new understandings [82].” Pragmatist epistemology is of the view that research methodologies, concepts and ideas should be selected based on “what works” for the selected research question, over what is considered to be objectively “right” or “true” as this varies for each individual [81]. Pragmatism thus connects the induction of qualitative research with the deduction of quantitative research with the idea of abduction; rejects the traditional philosophical dualism of subjectivity and objectivity with inter-subjectivity; and reconciles context and generalisability with transferability [83].

Abduction refers to interchanging between qualitative and quantitative approaches in a sequential manner based on inductive goals from one method informing the deductive goals of the other, and vice versa [83]. Similarly, pragmatists do not believe that it is possible for any research to be exclusively objective or subjective but instead acknowledges that each researcher has their own inherent beliefs and politics which will certainly affect the conduct and analysis of research and there needs to be this mutual understanding with research participants and recipients of the research –intersubjectivity [83]. Furthermore, pragmatism effectively

strengthens transferability of data due to the connection between the depth and breadth of data afforded by the marriage of quantitative and qualitative approaches [80, 83].

Thus, this PhD project utilised a combination of quantitative and qualitative methods, selected through abduction, to explore the aim of developing a healthy lifestyle change conversational agent for the general population in Singapore. After careful consideration, back and forth reasoning and planning, the following methods were finalised. A scoping review was first selected to map out the existing evidence, which highlighted a gap in the literature for a conceptual framework to inform the design, development and evaluation of healthcare CAs. This inspired and guided the development of a healthy lifestyle change CA which employed a qualitative needs assessment prior to the feasibility study. Upon completion, the feasibility trial called for qualitative feedback from participants. Finally, a systematic assessment of apps was planned, to assess how the CA Precilla compared and contrasted with existing CAs accessible to the general population.

3.3 Scoping existing evidence

Systematic reviews are the cornerstone of all reviews and have been around since the 1970s and 80s [84, 85]. They have increased in popularity especially with the emergence of prestigious institutions such as Cochrane and the Joanna Briggs Institute and since have grown to be valuable tools that can influence policy and practice today [86]. Systematic reviews follow a very rigorous and methodical approach which usually involves a team of expert researchers who set out to identify and critically appraise international literature relevant to a specific research question [87]. They aim to minimize bias as far as possible and if possible, conduct a meta-analysis if there is minimal heterogeneity in the quantitative data [88]. These reviews are expected to be pillars of evidence-based healthcare and can inform clinical guidelines [89]. However, researchers and peer reviewers noticed that such reviews may not be applicable in every circumstance. In cases where the volume of literature on a topic is unclear or the applications and methods used in an emerging field remain unknown, a scoping review would be more applicable. In this sense, scoping reviews are more geared towards identifying knowledge gaps, mapping out the scope of a body of literature, or clarifying concepts, research conduct, and methods [90, 91]. Scoping reviews may also pose as good pre-cursors to systematic reviews by pinpointing the right inclusion criteria and appropriate research questions to prioritize [92].

Considering the nascent nature of the research around CAs in healthcare, the need to map out existing usage applications, the dearth of robust evidence, the presence of knowledge gaps and no clear indication of a necessary niche research question to be addressed in this area, scoping review methodology was selected. Chapter 4 describes a scoping review conducted to map the

current usage applications of CAs in healthcare along with their basic characteristics, evaluation measures, input and output modalities, etc. An updated version of the Arskey and O'Malley framework was employed, and two researchers worked in parallel to screen, extract and present data in the review [91]. The five stages of conducting a scoping review were followed, namely: (1) identifying the research question, (2) identifying relevant studies by searching both grey and bibliographic literature databases, (3) study selection (based on established inclusion criteria), (4) charting the data in a predefined data extraction sheet with relevant points of interest and (5) collating, summarizing, synthesizing and reporting the findings [91].

Another method of scoping existing evidence was employed for the systematic assessment of smartphone apps. Here, Apple and Google Play stores were searched, as opposed to literature databases, for healthcare conversational agents. Although multiple other app stores exist, (eg. Blackberry or Amazon app store) the Google Play store (android), and the Apple app store (iOS) are the main app stores searched in the published studies in the research literature, hence, these stores were selected. In some studies, app stores were searched manually using search terms and relevant categories [93], whilst in other cases, an app database, *42 matters*, was employed [94, 95]. To conduct a robust, comprehensive search, as similar as possible to a PubMed search, *42 matters* was employed and the methodology presented in existing apps assessments was used for guidance [96, 97]. The systematic assessment was based on systematic review methodology as employed in other studies [96, 98, 99]. The methodology required the establishment of the following: a list of relevant search terms, assessment criteria (healthcare-related and general app characteristics), eligibility criteria for app selection, and criteria for app assessment.

For the app's assessment in this thesis, apps were evaluated based on mobile app certification (mHON code) guidelines [100], and to study behavior change, Michie's behavior change taxonomy was used [101]. The methodology is elaborated on in more detail in Chapter 9.

3.4 Conceptual framework methodology

According to Jabareen (2009), a conceptual framework can be defined as a network of interlinking concepts which collectively provide a sound understanding of a phenomenon [102]. They are typically developed through a process of qualitative analysis [102]. Similarly, in 1994, Miles and Huberman posited that a conceptual framework should include written and diagrammatic representations of the main variables to be studied on a topic and a comprehensive explanation of how these factors and concepts relate to one another [103]. In this thesis, the interweaving concepts related to text-based healthcare CA development were

identified and presented in a conceptual framework to solidify the CA design and development process for researchers.

Jabareen proposed a new method of qualitative analysis in conceptual framework development which involves: (1) mapping the selected data sources, (2) reading and characterizing relevant data, (3) defining and naming concepts, (4) deconstructing and categorizing concepts, (5) integrating concepts, (6) synthesizing and resynthesizing concepts for optimization and (7) validating the conceptual framework [102]. Jabareen's framework is relevant as other research groups have cited this methodology in the development of conceptual frameworks for health and mHealth related applications [104, 105].

The framework presented in Chapter 5 indicates how the stages relate to one another for the successful development of a healthcare CA. The relevance of each stage, and how they translate to the real world have also been emphasized.

3.5 Implementation science

Existing evidence has shown that novel interventions, tested rigorously in RCTs with exceptionally promising results have often failed to translate into practice. This implies that there are other factors beyond just effectiveness which contribute to the successful implementation of an intervention [106]. As such, this PhD project was highly motivated by Implementation science. Implementation science is concerned with factors in addition to effectiveness such as barriers and facilitators to uptake. The delay between evidence and widespread uptake of treatments was reported by Colditz and Emmons (2018) [107]. They compared the relatively slower uptake of smallpox vaccine, penicillin and insulin to the more rapid uptake of anti-retroviral treatments for HIV/AIDS accentuating that contextual factors sometimes take precedence over effectiveness in determining how quickly a clinical innovation is adopted.

The investigations in implementation science may be exploratory or explanatory culminating in qualitative and/or quantitative data [106]. "Exploratory" aims to plan for future implementation, while "explanatory" aims to identify reasons for the failure of past implementations. This PhD project was concerned with the development and preliminary evaluation of a rule-based CA promoting healthy lifestyle change for its potential implementation for the general public in Singapore. As such, a more "exploratory" approach was employed. A pilot feasibility study was chosen as the study design to support the investigations around the feasibility of implementing such a CA, and its potential acceptability by the target population (Chapter 3.6). Leading up to the pilot study, a needs assessment was performed to obtain the views, opinions, needs and characteristics of the target population. This would enable easier uptake of a CA, designed to be relevant to the needs and wants of the target users (Chapter 3.7). A further qualitative study was

conducted to understand the target users' experience of using the CA, identify the barriers and facilitators to the CA usage and collate suggestions for improvements to improve the uptake up such a CA when it is implemented in practice (Chapter 3.7).

3.6 Pilot feasibility study

Due to resource constraints, it would not be plausible to test every novel intervention for efficacy and effectiveness, at the get go [108]. There is a need to systematically evaluate and prioritise those interventions which show the greatest likelihood of being efficacious. This is where feasibility studies pose as an indispensable method to generate findings which help determine whether an intervention should be recommended for efficacy testing [108].

A pilot feasibility study was designed and implemented per the Medical Research Council's (MRC) guidelines on developing and evaluating complex interventions [109]. An intervention qualifies as 'complex' if a range of behaviors is targeted, broad expertise and skills are necessary for those delivering the intervention, there is an involvement of multiple groups and settings, amongst other criteria [109]. In this project, the research team had to equip themselves with a wide range of expertise on healthcare CA development, and feasibility and acceptability were tested at multiple points with varied outcomes measures. Furthermore, the pilot study entailed three large stages: a needs assessment, a 4-week pilot study, and follow-up interviews. All these factors qualify it as a complex intervention. The MRC framework identifies four main stages to be followed by complex interventions: (1) development, (2) testing of feasibility, (3) evaluation, and (4) implementation [109]. Resultantly, as described in Chapter 7, following the development of the CA (phase 1), the pilot study embodies the feasibility phase (phase 2) of the MRC guidelines. Chapter 7 also proposes recommendations and points for improvement for the CA which can subsequently be evaluated (phase 3) for effectiveness before being implemented (phase 4) in the near future.

3.7 Qualitative approaches

In certain circumstances, qualitative research transcends quantitative research in that it enables researchers to explore the views and opinions of target users and stakeholders without pre-set boundaries. They can provide very detailed insights and strong opinions which cannot be collated or measured by quantitative scales. This was especially relevant when exploring the views of the public on CAs in general and their application for healthy living and diabetes prevention in Singapore as the evidence in this area is currently lacking. Semi-structured interviews were selected to understand the needs of the public when it comes to using a CA for healthy lifestyle behavior change as a user-centered approach for CA development was adopted [110, 111]. These interviews denote a flexible structure whereby the interviewer has a predefined interview guide

but is free to ask additional questions, probe, or comment to elaborate on interesting and relevant points raised by the interviewee. To then follow up on the experience of interacting with the CA, another set of interviews was conducted. These were invaluable in revealing the publics' perceptions of the prototype, their likes, and dislikes, as well as receiving their suggestions on improvements for the future. To conduct these interviews effectively and with fidelity, the recommendations made by DeJonckheere and Vaughn on semi-structured interviewing were followed [112]. The pivotal steps were: (1) defining the purpose of the studies, (2) obtaining suitable participants, (3) considering ethical issues (like informed consent and confidentiality), (4) planning logistics (like interview locations to maintain the privacy of participants or telephone or in-person interviews to suit their convenience), (5) developing the interview guide, (6) establishing trust and rapport with volunteers, (7) conducting the interview, (8) taking field notes, (9) analyzing the data, (10) demonstrating the trustworthiness of the research and (11) presenting the findings in a paper [112].

There are several methods of qualitative analysis including content analysis, narrative analysis, discourse analysis, grounded analysis, and thematic analysis [113]. Thematic analysis was the chosen method for the needs assessment and follow-up interviews. Although time-consuming, thematic analysis allows the researcher to explore the data thoroughly, either inductively or deductively [113]. In this context, an inductive approach was used whereby the data determined the themes organically [114]. Without much research to inform the publics' perceptions of healthcare CAs in Singapore, there were no preconceived themes to approach the data with [114]. Identifying themes from the data was especially helpful in informing the conversational prompts to be included in the healthy lifestyle CA Precilla which was to be developed for the study population. Furthermore, some of the themes represented the participants' experiences and opinions on CAs and healthy living. Having multiple themes and codes meant the data could be very comprehensive, covering a broad range of opinions both positive and negative. This was very helpful in directing which areas of the CA would flourish and which areas need improvement in the future.

Reflexivity is an important component in qualitative research. I (the PhD candidate) familiarised myself with interview techniques using the literature and online courses on Coursera. Furthermore, I attended workshops run by the Family Medicine and Primary Care (FMPC) network on coding and qualitative data analysis in preparation for these studies. I kept field notes during the interviews and reflected on the rapport built with participants, the reviewer working in parallel and other noteworthy points which may have affected the conduct of the interviews or analysis of the findings. The interviews were semi-structured and adapted to suit the flow of each participant and where they chose to emphasise or share more of their views. The participants did

not build a relationship with any other the researchers prior to or after the interviews. A detailed account of the researcher's reflexivity is presented in Chapter 10.9.1, Table 10.2 and Figure 10.1.

This thesis presents the projects undertaken in this PhD for CA development and evaluation in healthcare, with a lot of optimism. However, in reality, there will be multiple challenges to field beyond design and development, in terms of implementation barriers, privacy and security concerns, digital inequity as well as ambivalence about potential effectiveness. The researcher was not blind to these issues and they are discussed in more detail in chapter 10.

Chapter 4 - Conversational Agents in Health Care: Scoping Review and Conceptual Analysis

This chapter addresses the first aim to “map and analyze the current uses of conversational agents in healthcare and identify knowledge gaps by performing a scoping review.” Section 4.1 provides some background information on existing reviews on healthcare CAs, what is missing from the literature, and the specific objectives of this review. The methodology is described in section 4.2 and the results are presented in section 4.3 both narratively and diagrammatically. The implications of the findings, strengths, and limitations of this review as well as suggestions for future research are outlined in section 4.4.

4.1 Introduction

Health care, which has been exposed to a decade of text messaging on smartphones, is an ideal candidate for conversational agent–delivered interventions. Conversational agents (CAs) enable interactive, bi-directional communication, and their text- or speech-based method of communication makes it suitable for a variety of target populations, from young children to older adults. The concept of using mobile phone messaging as a health care intervention has been present and increasingly explored in health care research since 2002 as evidenced by multiple systematic reviews. Existing reviews on conversational agents specifically, focus on a certain subtype of agents such as virtual coaches [41, 115, 116] and embodied conversational agents (ECAs) [117] or on specific functionalities of these agents such as behavior change [118] and mental health applications [119, 120]. Other reviews solely discuss technical aspects of conversational agents such as system architecture and dialogues [121] or comment on the funding component of health care conversational interfaces [122].

4.1.1 Objective

In contrast to the more niche reviews conducted thus far, this scoping review aimed to provide a more comprehensive, timely, bird’s eye overview of the existing research literature on the use of healthcare CAs. This review examined how CAs have been employed and evaluated in the literature to date and mapped out their characteristics. Finally, in line with the observed gaps in the literature, recommendations were provided for future conversational agent research, design, and applications. As such, the research questions for this scoping review were:

“For what applications have CAs been used in healthcare? In which healthcare areas have CAs been used and how have these CAs been evaluated?”

4.2 Methods

4.2.1 Search Strategy

This scoping review was conducted using the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist [123] and the updated Arksey and O'Malley framework [124]. To address the research questions presented in 4.1.1, a broad literature search needed to be conducted to identify the various usage applications of CAs. This required a comprehensive search of several databases using a multitude of synonyms. A broad literature search was conducted in April 2019 in MEDLINE (Medical Literature Analysis and Retrieval System Online; Ovid), EMBASE (Excerpta Medica database; Ovid), PubMed, Scopus, and Cochrane Central. Given the novelty of the field, and to increase comprehensiveness, gray literature was also searched for in the OCLC WorldCat database, ResearchGate, Google Scholar, OpenGrey, and the first 10 pages of Google. There was an extensive list of 63 search terms, which were various iterations of the term “conversational agents” (Appendix 4.1). These synonyms were generated using a web-based search and by identifying specific terms or phrases used in the titles of articles discussing health care conversational agents. Snowball searching of relevant articles and systematic reviews was also done to identify more articles related to the review. The PRISMA checklist for scoping reviews was completed and is presented in Appendix 4.2.

4.2.2 Inclusion and Exclusion Criteria

Primary research studies that had evaluated a CA implemented for healthcare were included. Articles that just presented a proposal for CA development, articles that mentioned conversational agents briefly or as an insignificant part of a review, as well as opinion pieces and articles where primary research was not conducted, were excluded. A further point of exclusion was articles with poorly reported data on CA assessments where there was minimal or no evaluation data. In addition, articles concerning ECAs, relational agents, animated conversational agents, or other conversational agents with a visual or animated component were excluded. ECAs are computer-generated virtual avatars that interact with users with both verbal and non-verbal communication [125]. Relational agents are a type of ECA designed to create long-term deep and meaningful relationships with individuals [126]. As the user's interaction may be affected by nonverbal behaviors, graphics, and layout of the program, it was decided that the complexities associated with ECAs were beyond the scope of this review and were therefore excluded.

4.2.3 Screening, Data Extraction, and Analysis

Following an initial screening of titles and abstracts, full texts were obtained and screened by 2 reviewers. From the included studies, 2 reviewers independently extracted relevant information in an Excel (Microsoft) spreadsheet. The data extraction covered the first author, year of publication, source of literature, the title of article, type of literature, study design and methods,

geographic focus, health care sector, conversational agent name, accessibility of conversational agent, dialogue technique, input and output modalities, and nature of conversational agent's end goal. This data extraction sheet was piloted on at least five articles. Any discrepancies in the extracted data were discussed between the authors and resolved through discussion and consensus. A narrative synthesis of the included literature was conducted, and findings were presented on (1) study specifics, such as study design, geographic focus, and type of literature; (2) conversational agent specifics (ie, conversational agent delivery channel, dialogue technique, personality, etc); (3) conversational agent content analysis; and (4) evaluation findings. Thematic analysis was used to analyze the content, scope, and personality traits of the conversational agents. Two researchers familiarized themselves with the literature identified, generated the initial codes concerning personality and content analysis, applied the codes to the included studies, compared their findings, and resolved any discrepancies via discussion. The exploration of CA personalities was motivated by the concepts presented in the study by de Haan et al. [127], which posited that personalities can be extended to nonhuman artifacts to explain their actions and behavior [127]. Furthermore, it was emphasized that personality traits are especially important in the design of socially interactive robots, which CAs are. The 5 dimensions of personality used to guide the analysis of CAs presented in this paper were: extraversion, agreeableness, conscientiousness, emotional stability, and culture. Multiple codes were sometimes assigned to the same agent where necessary, but this was limited to a maximum of 3 codes to maintain some degree of specificity.

4.3 Results

4.3.1 Search Findings

The initial database searches yielded 11,401 records, and another 28 records were retrieved from gray literature sources and reference lists of relevant studies. 196 duplicates were identified and removed, leaving 11,233 titles and abstracts to be screened. Title and abstract screening led to the exclusion of 11,099 records, resulting in 134 full texts to be assessed for eligibility. Of these, 87 were excluded, resulting in a final pool of 47 reports made up of 45 studies and 2 ongoing trials (Figure 4.1). A list of the excluded studies and the reasons for exclusion are presented in Appendix 4.3.

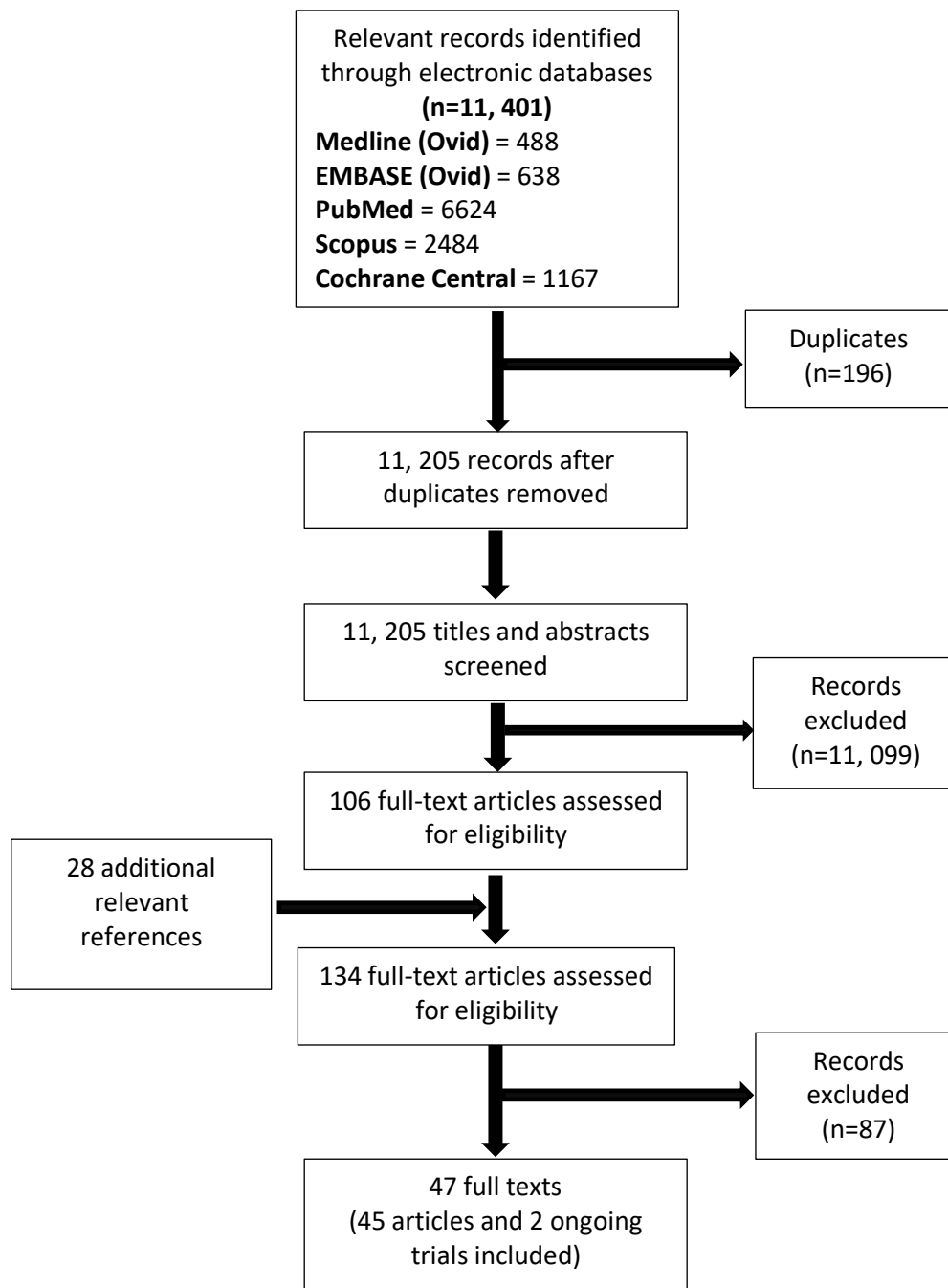


Figure 4.1 Scoping review PRISMA flow chart

4.3.2 Characteristics of included studies

In this scoping review, 40 included studies were from high-income countries (HIC), six from low and middle-income countries (LMIC), and the study location was not specified in one study [128]. Twenty one studies were from European countries, including Italy [129, 130], Switzerland [131-137], France [138, 139], Portugal [140], The Netherlands [141], The United Kingdom [142-146], Spain [147, 148] and Sweden [149]. Eight studies originated in Asian countries: Philippines [150],

China [151], Japan [75, 152], Pakistan [153], India [154, 155] and Hong Kong [156]. Other geographic regions covered were Australia [157, 158], Canada [159], New Zealand [160, 161], South Africa [162], and the United States of America [34, 163-172].

A variety of study designs were used in the included studies: 20 case studies [129, 131, 132, 146-148, 151, 153, 155, 157-163, 165, 167, 168, 172], four surveys [140, 141, 144, 150], three observational studies [138, 169, 170], eleven randomized controlled trials (RCT) [34, 128, 133, 135, 142, 149, 152, 156, 164, 166, 171], three diagnostic accuracy studies [75, 143, 145], three controlled before and after studies [130, 134, 154], two ongoing trials [136, 139], and one pilot study [137] (Figure 4.2).

The types of literature included: 25 journal articles [34, 129, 132, 140-142, 146-150, 152, 153, 156, 158-160, 164-170, 172], 11 conference abstracts [130, 131, 133, 135, 137, 144, 154, 155, 157, 162, 163], four conference papers [128, 134, 151, 161], one poster abstracts [75], four electronic pre-prints [138, 143, 145, 171] and two clinical trial protocols [136, 139].

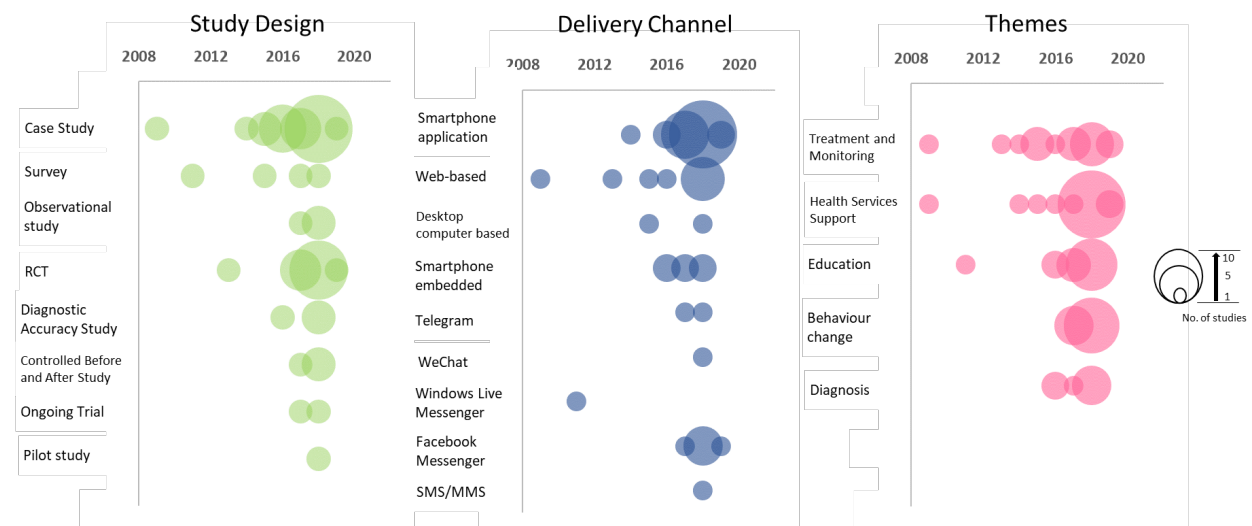


Figure 4.2 Bubble plots showing the distribution of study designs, CA delivery channels, and healthcare themes in the included literature, plotted against the year of the publication. (The scale on the right indicates that the size of the bubble corresponds with the number of studies whereby, the smallest denotes 1 study and the largest, 10 studies.)

There was an increasing number of publications in consecutive years from three in 2015, five in 2016, 10 in 2017, and 23 in 2018. Some author groups were highly productive and published at least two papers within two years. Kowatsch et al. have published three papers between 2017 and 2018 based on their open-source behavioral intervention platform MobileCoach that allows authors to design a text-based healthcare CA for obesity and behavior change [128, 134, 173]. Griol et al. are another group that published articles on CAs for chronic conditions including

chronic pulmonary disease [148] and Alzheimer's disease [147] in 2015 and 2016, respectively. Such productive teams reiterate the research interest in this area of conversational agents. Furthermore, the high frequency of publications also emphasizes the feasibility and support to conduct research successfully in this area.

4.3.3 Characteristics of the conversational agents in the included studies

i. Conversational agent delivery channel

Conversational agents were delivered through a variety of means in the included studies. Most (n=23) were smartphone applications [128, 132-135, 137, 138, 140, 143-146, 149, 152, 154, 155, 159, 161, 164, 166, 168, 169, 171], web-based (n=5) [142, 151, 157, 158, 165], desktop computer based (n=2) [150, 163], used smartphone embedded software (n=6) (e.g. Siri, Google assistant or Alexa etc.) [129, 136, 147, 160, 167, 170], Telegram [130, 162], WeChat [156], SMS/MMS [172], Windows Live Messenger [141] or Facebook Messenger [34, 131] and four were made available on more than one platform [75, 138, 144, 166]. Three studies did not specify the CA delivery channel [139, 148, 153].

ii. Technical development approach

Eight studies mentioned technical details related to CA development. Specific tools were named such as C and MS Access [150]. Others discussed the application of well-known concepts, to conversational agent development such as the Computers are Social Actors (CASA) paradigm to develop a health advice CA, or the conversion of the structure association technique (SAT) into a digital format for implementation on a LINE platform [152, 166]. Some emphasized dataset creation and sources for the knowledge base [129]. Four studies provided a step-by-step explanation of the technical development of the CA. Cheng et al. broke this down and explained it in parts: whereby application development happened on Google's home device, webhook, and internal logic, and web interface [163]. Galescu et al. described the CARDIAC system architecture including knowledge base, task models, dialogue management, speech recognition, language generation, etc. [165]. Griol et al. presented a spoken dialogue system with specific details of the proposed emotion recognizer [148]. For example, it considered pitch, frequency, energy, and rhythm of speech input from the user [148]. Joerin et al. provided a less technically dense explanation for CA development but mentioned technologies used such as emotion algorithms and machine learning techniques [159].

iii. Input and output modalities

The conversational agents could be categorized according to whether the user input was fixed (i.e. predetermined) or unrestricted (i.e. free text/speech). Ten studies employed fixed text user inputs [128, 131, 133-135, 137, 139, 143, 166, 171], with two additional studies enabling fixed text and image inputs [75, 152]. Nineteen studies allowed free text inputs [130, 132, 136, 141, 142, 145, 151, 153, 154, 156, 161, 162, 164, 168, 169, 172, 174-176], four studies used both fixed and free text user inputs [138, 149, 157, 177]. Speech was enabled in eight studies [129, 148, 155, 160, 163, 165, 178, 179] while free text and speech together were employed in three studies [147, 159, 180]. The method of user input was unspecified in one study [144] (Figure 4.3).

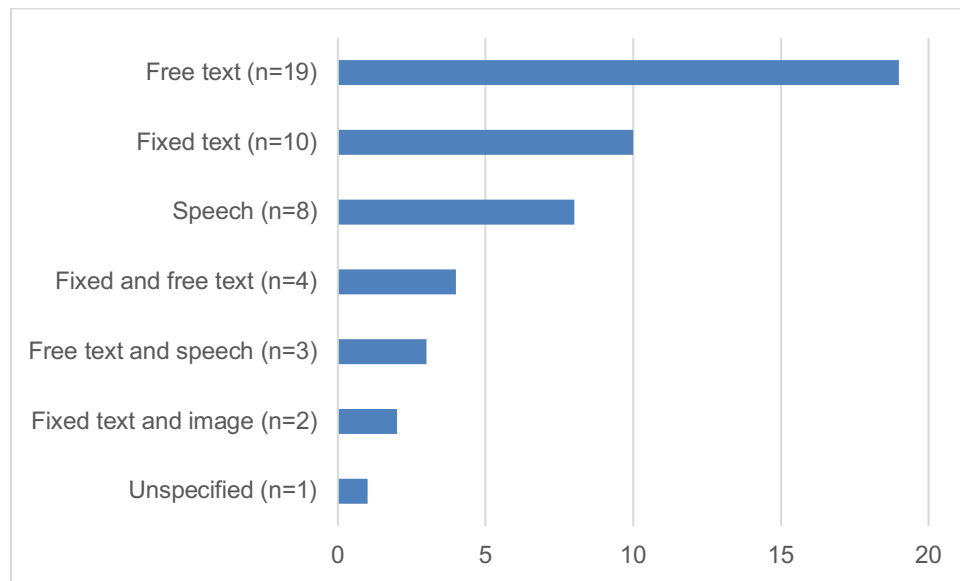


Figure 4.3 Type of input in CAs from the scoping review

Similarly, output modalities were largely text alone (n=30) [34, 75, 128, 130, 133, 135-139, 141-143, 145, 146, 149-151, 153, 154, 156-158, 161, 162, 164, 166, 168, 171, 172], text and speech (n=5) [132, 140, 148, 155, 170], speech alone (n= 4) [129, 163, 165, 167], text and images (n= 4) [134, 152, 159, 169], text, speech and images [147] or text, speech, images and videos [131, 160]. The input and output method was not specified in one study [144] (Figure 4.4).

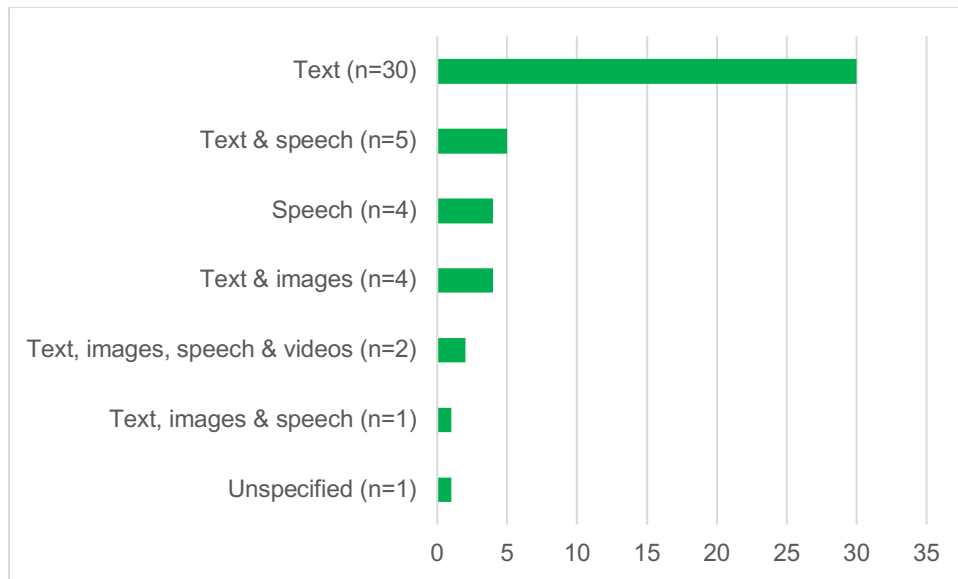


Figure 4.4 Type of output in conversational agents from scoping review

iv. Conversational agent personality

The descriptive terms for CA personalities were condensed into a list of nine relevant personality traits (Table 4.1).

The CAs in included studies were healthcare professional-like [142, 143, 147, 151, 155, 157, 158, 169], informal [128, 131, 138, 141, 146, 150, 164, 168], coach-like [34, 131, 133, 137, 149, 151, 154], knowledgeable [75, 141, 145, 156, 172], human-like [132, 162, 163, 171], culture specific [131, 132, 138], factual [75, 160], gender specific [128, 162] and/or an obvious conversational agent identity [128, 150].

One article [162] reported on a conversational agent personality which was criticized for being overly formal and some articles did not report on the personality of the conversational agent at all [129, 130, 134-136, 139, 140, 144, 148, 152, 153, 159, 161, 165-167, 170].

Table 4.1 Personality codes derived for the conversational agents included in this review, adapted from Haan et al. [127].

Personality Codes	
Coach-like:	encouraging, motivating, nurturing
Conversational agent identity:	explicitly identifies as a conversational agent
Culture-specific:	speaks the native language or has native names
Factual:	non-judgmental, no personal opinions, responses based on fact or observations
Gender-specific:	male and female versions available

Healthcare professional-like:	designed to be a doctor, expert. I.e. mimics a healthcare professional
Human-like:	tries to emulate humans. e.g. participants reported feeling like they were talking to another human or researchers used features like "typing" to make the conversation more human-like
Informal:	informal, like talking to a friend. Uses exclamations, abbreviations, and emoticons
Knowledgeable:	content created or informed by medical experts

v. Human Involvement

Seven studies reported on human involvement in the conversation and the remaining articles did not [128, 134, 137, 154, 156, 162, 168]. Human involvement took many forms. In some studies, a healthcare administrator or professional was available via a chat channel for the user to communicate with. In other cases, the individual's role was to monitor the user's activity on the conversational agent and provide personalized feedback to them.

vi. Conversational agent goals

All the conversational agents in this review were "goal-oriented". Goal-directed CAs have a conspicuous endpoint and guide the user towards achieving it. In contrast, "chit-chat" agents have no specific end goal, do not delve into the details of any topic, and have a primary aim of merely keeping the conversation going [181]. The goal-oriented conversational agents were further characterized as having long or short-term outcomes. 22 articles were focused on conversational agents with long-term goals and 23 with short-term goals (See Appendix 4.4). CAs with short-term scope provided users with a response or service almost instantaneously such as answering health-related queries [167]. Conversely, those with long-term scope needed to build a rapport or relationship with the user, over time, to help them overcome a health-related issue such as smoking cessation [156] or working through a mental health problem [34]. Two studies reported on conversational agents with both short-term and long-term goals [130, 141], such as answering immediate queries (short) and providing education and building up the user's knowledge on the topic over time (long) [141].

4.3.4 Conversational agent content analysis

CA content was categorized into five themes: treatment and monitoring, health service support, education, lifestyle behavior changes, and diagnosis. A few CAs spanned several different themes. (See Table 4.2 and Appendix 4.4).

i. Treatment & Monitoring

Seventeen articles reported on CAs targeting treatment, monitoring, or rehabilitation of patients with specific health conditions. One study described a CA which supported the preservation of cognitive abilities in those with Alzheimer's disease [147]. Two other studies focused on CAs supporting the treatment for metabolic conditions like Type 2 diabetes [154] and obesity [128]. Eight studies presented CAs for mental health management using techniques like counseling [152], cognitive-behavioral therapy (CBT) [34, 149] method of levels (MOL) therapy [142], positive psychology [146], provision of a virtual companion [151] and a combination of modalities such as CBT with mindfulness-based therapy, emotionally focused therapy, motivational interviewing, etc. [159, 164] One study each explained the use of a CA for monitoring patients with asthma [168], HIV [130], heart failure [165] and chronic respiratory disease management [148]. Non-disease-specific conversational agents were used as health information advisors [166] and pediatric generic medicine consultants [150].

ii. Healthcare services support

Nineteen studies discussed CAs used to complement existing healthcare services. Their tasks covered remote delivery of healthcare services for mental health support [152, 159, 164], breast cancer [138, 139], dysarthria [129], obesity [135], diabetes management [163], chronic respiratory diseases [148], asthma [168], heart failure [165] and HIV management [130]. Other studies mentioned CAs automating healthcare services such as patient history taking [132, 161], providing health advice [166], symptom checking [143], triaging, and providing diagnosis support [145, 153, 158].

iii. Education

In thirteen articles, CAs were used primarily for educating the users. The topics addressed were sexual health [144, 160] including information on HIV [162], overcoming unhealthy habits such as alcohol misuse [157] and smoking cessation [156], improving wellbeing [171], diabetes management [163], breast cancer [138, 139] and medication-related queries [140], as well as general health for adolescents on sex, drugs, and alcohol [141, 167, 170].

iv. Lifestyle behavioral changes

Twelve studies were identified whereby CAs were used for healthy lifestyle behavior change in the general population as well as more niche populations. Two studies discussed conversational agents for the management of obesity in younger patients, including adolescents [128, 135]. They employed a CA with a coach-like personality to promote physical activity [136] and healthy eating [131], sometimes with incentives, and provided techniques on how to reverse obesity [133, 134, 137]. Other behavior change interventions used a social-media-driven CA for smoking cessation [156], a health coach for diabetes prevention [169], a reflection companion to encourage physical

activity in adults [172], and emotionally intelligent agents to improve mental health [146] and wellbeing [171].

V. Diagnosis

Six articles presented healthcare CAs with a primary aim of diagnosis. Three articles reported on CAs conducting triage, diagnosis, or a combination of both, using a symptom checker [143, 145, 158]. Three studies reported solely on the general diagnostic accuracy of two conversational agents [153, 155, 161]. One article reported on a conversational agent for diagnosing sexually transmitted infections to overcome barriers such as social stigma, embarrassment, and discomfort associated with traditional diagnostic approaches [75].

4.3.5 Conversational agent evaluation

Included studies that evaluated conversational agents reported on their accuracy, user acceptability, and effectiveness. Some studies reported more than one outcome, for example, acceptability and effectiveness. The evaluation data was mostly positive, with a few studies reporting limitations of the CA or technical issues experienced by users. Nineteen studies presented self-reported data from participants in the form of surveys, questionnaires, etc. whilst in 16 studies, the data were objectively assessed in the form of changes in BMI, the number of user interactions, etc. In nine studies, there was a mixture of self-reported and objectively assessed outcomes (see Table 4.2).

i. Accuracy: information, diagnosis, and triage

Eleven studies reported on the accuracy of conversational agents [75, 129, 143, 145, 151, 153, 155, 158, 160, 161, 165] (see Table 4.2). Middleton and Razzaki et al. evaluated two versions of the Babylon conversational agent: “Babylon check” and “Babylon chatbot for triage and diagnosis,” respectively [143, 145]. In both studies, the diagnostic accuracy of the CA was assessed based on its ability to diagnose using clinical vignettes and their performance was compared with doctors. The CAs were found to diagnose more accurately, rapidly, and safely compared to doctors and nurses. Similarly, Ghosh and Danda et al. assessed conversational agents on their general diagnostic accuracy, and they had a precision rate of 82% and 86%, respectively [155, 158]. Ni et al. assessed chatbot MANDY, designed to automate patient intake and symptom checking, on its ability to adequately diagnose the patient [161]. The accuracy of prediction was 100%, 64%, 25%, and 14% for respiratory issues, chest pain, headache, and dizziness, respectively [161]. Another two studies tested the accuracy of CAs employed for sexual health purposes [75, 160]. The conversational agent in Kobori et al. diagnosed sexually transmitted infections (STIs) with an accuracy of 77.7% and had high effectiveness (97.7%) in encouraging

patients to seek help at a clinic earlier [75]. In contrast, Wilson et al. compared smart assistants: Google Assistant, Siri, and Google search to determine their accuracy in responding to queries around sexual health. The Google search option provided the best answers and also had the lowest failure rate [160]. Another study compared three known virtual assistants Siri, Google Assistant, and Amazon Alexa on their abilities to recognize speech from individuals with Dysarthria [129]. They all performed similarly (50-60% recognition) with Siri being the only agent attempting to parse all the dialogue inputted [129]. Two studies discussed the accuracy of two conversational agents in diagnosing stress, achulophobia, and autism in adolescents and children [151, 153]. Teenchat had a 78.34% precision rate for stress [151], Aquabot had an accuracy of 85%, 86.64% and 87.2% (three groups aged between 18-28) for achulophobia and 88%, 87.6% and 87.53% (three patient groups aged 1-7) for autism [153]. Finally, Galescu et al. discussed the accuracy of a conversational agent “CARDIAC” in speech recognition for patients with heart failure [165]. A significant number of errors were detected and attributed to insufficient vocabulary coverage in the language model given that there was an “out-of-vocab” rate of 3% [165].

ii. Effectiveness

The effectiveness of healthcare conversational agents was assessed in eight studies [131, 137, 142, 154, 159, 164, 176, 179]. A further 10 studies reported on both effectiveness and acceptability of which five are presented here [133, 149, 152, 169, 174] and the remainder are presented under “Acceptability” (see Table 4.2). Five studies described CAs for healthy lifestyle change specifically for healthy eating [131], active lifestyle [133], obesity [137], and diabetes management [154, 169]. Casas et al. reported improvements in food consumption [131] while Stasinaki and Heldt et al. noted increases in physical activity [133, 137] with high compliance [133, 137]. Shaikh et al. focused on the successful reduction in HbA1c levels post-engagement with “Wellthy diabetes” [154] whilst Stein et al. reported on successful weight loss (2.38%) and high levels of satisfaction (87/100) for the diabetes prevention chatbot [169].

Eight studies noted the effectiveness of conversational agents for mental health applications [34, 142, 146, 149, 152, 159, 164, 167]. In Fulmer et al. a conversational agent “Tess” imbued a statistically significant improvement in depression and anxiety compared to the control group [164]. Two studies explored machine learning-based CAs for CBT in young adults [34, 149]. The CA was both effective (reduced levels of depression, perceived stress, and improved psychological well-being) and well-received (high engagement with the chat app and high levels of satisfaction) [34, 149]. This positive effect was reproduced in Joerin et al. where emotional support from Tess decreased symptoms of anxiety and depression by 18% and 13%, respectively

[159]. Inkster et al. employed the PHQ-9 self-reported depression scale to note significant improvements in depression scores in the high user group compared to the low user group [146]. Moreover, 67.7% of users found the app usage to be helpful and encouraging [146]. In Kamita et al., the counseling bot encouraged significant improvements in users' self-esteem, anxiety and depression compared to the control condition [152]. Aside from being effective, user ratings of acceptability, using the technology acceptance model, were also higher in the CA condition compared to the control [152]. Gaffney et al. proposed a conversational agent "MYLO" which was significantly better than the existing conversational agent ELIZA in problem-solving and helpfulness but both were equally effective in lowering distress [142]. Miner et al. compared Apple's Siri, Microsoft's Cortana, Samsung's S Voice, and Google Now on their abilities to respond to questions about mental health, interpersonal violence, and physical health [167]. Siri responded appropriately and empathetically to issues concerning depression and physical health and Cortana for matters involving interpersonal violence [167].

iii. Acceptability

Twenty-six studies commented on the acceptability of conversational agents (see Table 4.2). Five studies commenting on acceptability and effectiveness were discussed above [133, 149, 152, 169, 174] (see Effectiveness) so, the remaining 21 are presented here [128, 130, 132, 134, 135, 138, 141, 144, 147, 148, 156, 157, 162, 163, 166, 168, 171, 172, 177, 178, 180]. Six studies were targeted at children or adolescents. Three studies discussed CAs for health education regarding medication, asthma management, drugs, sex, and alcohol [141, 150, 168]. Acceptability was exemplified by high response rates and scores like "strongly agree" or "agree" for user-friendliness, appropriateness, consistency, and speed of response [150]. Additionally, users in Crutzen et al. favored the CA over existing methods of information provision [141]. In another three studies, CAs were implemented for the management of obesity in young adults [128, 134, 135]. Acceptability was high in all studies evidenced by the enjoyment of the chats, bonding, and formation of social and emotional relationships, high perceived ease of use, usefulness, and intention to use [128, 134, 135]. In L'Allemand et al. acceptability was measured as high compliance, which was attributed to the rewarding game system [135].

In four studies, the healthcare CAs targeted chronic conditions like Alzheimer's disease, diabetes, heart failure, and chronic respiratory disease [140, 147, 148, 163]. In Cheng et al. users responded positively, particularly to features for personalization and the conversational agent's ability to understand and respond to natural conversational language [163]. Some difficulties included learning commands, restricted answer options, slow processing speed, and some problematic responses [163]. Lobo et al. reported acceptability as usability where the CA had a system

usability score of 88 which corresponded to “very good” [140]. Griol et al. considered the Alzheimer’s patients’ caregiver’s perspective when judging the acceptability of the conversational agent. The global rate for the system (on a scale from 0 to 10) was 8.6 and the application was thought to be attractive, adequate, and appropriate for its purpose [147]. Griol et al. (2015) employed an emotionally sensitive conversational agent for chronic respiratory disease patients who rated this agent significantly higher for interaction rate, frequency, and empathy than the baseline version [148].

A further three studies were concerned with sexual health and/or HIV management [130, 144, 162]. They indicated that in this field, CAs could be used for a variety of functions such as booking an appointment, getting test results, therapy, event reminders, etc. [130]. In addition, the conversational agent in Heerden et al. was well-received when used as a counseling tool because it was given an avatar-like profile image and the conversation was embedded in a familiar chat interface which users associated with talking to another human being [162]. In Nadarzynski et al. users appreciated the CA because of its easy accessibility as a smartphone application and the availability of remote services such as video consultation, potentially alleviating any reservations users may have around discussing sexual health in-person [144].

Two studies employed an emotionally sensitive conversational agent for mental health counseling and general health information advice [166, 171]. In Liu et al. the sympathetic CA was rated more positively than the advice-only condition [166]. Another CA for wellbeing improvement procured positive feedback from participants who thought it was “*an interesting experience*”, “*pretty quick*” and “*fun*” [171].

In three studies, CAs were used for healthy behavior change, specifically to halt smoking, binge drinking and promote physical activity [156, 157, 172]. For smoking cessation, participants indicated enjoyment when conversing with the CA, and effectiveness was also demonstrated by 38.3% not having smoked in the past week and 69.4% admitting to a reduction in smoking frequency [156]. In Elmasri et al. the participants had a higher satisfaction rate with the use of the CA to manage and treat alcohol misuse [157]. For physical activity promotion through the use of a reflection companion, response rates were high (96% at baseline, 90% at follow-up) indicating high engagement throughout the study. Furthermore, sixteen out of 33 participants opted to continue without any reward suggesting participants found some added value in using the conversational system [172].

Two studies looked at the acceptability of conversational agents for healthcare service support [132, 170]. Outcomes were reported qualitatively including comments on ease of use, the

humanity of the chatbot, and users' comfort with the input functionalities available to them as well as limitations regarding technical difficulties [132]. Bickmore et al. more specifically compared conversational assistants Siri, Alexa, and Google assistant on their provision of health information and found satisfaction to be lowest with Alexa and highest with Siri. Overall, there was a neutral rating for satisfaction with a median score of 4 (IQR 1-6) [170].

One study discussed a CA targeted at improving the quality of life and medication adherence of breast cancer patients [138]. Participants indicated a positive experience when interacting with the CA whereby 88% said it provided them with support in tracking their treatment and mentioned they would recommend the conversational agent to their friends. There was an overall satisfaction of 94% [138].

Table 4.2 Characteristics of studies included in scoping review

Study ID, Study Design, Location, Income	Field of study	User input/agent output modality	Theme, types of participants, and & number of participants	Aim	Outcomes and findings (Self-reported vs Objectively reported)
Ballati <i>et al.</i> (2018); Case study, Italy, HIC	Dysarthria	Free Speech/ Speech	TDV; people with dysarthria (n=17)	To investigate how far individuals with dysarthria can interact with and be understood by Siri, Google Assistant, and Amazon Alexa.	Accuracy: Siri, Google Assistant, and Amazon Alexa performed similarly with a recognition accuracy of 50-60%; Objective.
Bickmore <i>et al.</i> (2018); Observational study, USA, HIC	Medical information	Free Text & Speech/ Speech, Text	EDU; general population (n=54)	To explore the use of conversational assistants to access medical information and assess the risks of such usage.	Effectiveness: Of the 168 tasks executed on the conversational agents, 29.2% could have had hazardous outcomes including 16.1% which could have resulted in death. Alexa failed most tasks (125/394, 91.9%). Siri had the highest task completion rate (365, 77.6%) and the highest likelihood of inflicting harm for the test scenarios (27, 20.9%); Objective. Acceptability: Satisfaction was neutral with a median rating of 4 (IQR 1-6); Self-reported.
Casas <i>et al.</i> (2018), Case study, Switzerland, HIC	Healthy eating	Fixed Text/ Text, image, video	BC; general population (n=36)	To present the design for a conversational agent which aims to coach people on how to improve their food lifestyle.	Effectiveness: Two choices are available to the user: i) reduce his consumption of meat or ii) consume more fruits and vegetables. User tests were conducted with 36 participants and only 11% of challenges on diary intake were successful with the use of the conversational agent. However; there was improved consumption in 65% of cases compared to before the intervention; Self-reported.
Chaix <i>et al.</i> (2018), Observational study, France, HIC	Breast cancer	Fixed & Free Text/ Text	EDU, HS; patients (n=958)	To evaluate a conversational agent (Vik) designed to improve the quality of life of patients with breast cancer by addressing their fears and concerns. An additional observation was the performance of the conversational agent's medication adherence function	Acceptability: The majority (94%) of the participants were satisfied with the agent's support; Self-reported. Effectiveness - Vik increased patient compliance with their treatment (medication adherence) thanks to the reminder function; Objective.

Cheng <i>et al.</i> (2018), Case study, USA, HIC	Diabetes management	Free Speech/ Speech	EDU, HS: type 2 diabetic patients (n=10)	To propose a "Healthy coping with diabetes" conversational agent designed to support elderly patients in independently managing their type 2 diabetes.	Acceptability: The majority (80%) of the participants preferred the intervention and the users were satisfied with the Healthy Coping functionalities; Self-reported.
Clinical trial NCT03384550 (2017); Ongoing trial, Switzerland; HIC	Physical activity	Free Text/ Text	BC; general population (n=274)	To test the efficacy of a smartphone app with a conversational agent component targeted at increasing physical activity in healthy adults through the use of coaching and incentive provision.	N/A
Clinical trial NCT (2018); Ongoing trial, France, HIC	Breast Cancer Patients' Information	Fixed Text/ Text	HS, EDU, patients with breast cancer (n=150)	To present the ability of the conversational agent Vik to respond to breast cancer patients' questions as a group of multidisciplinary doctors would.	N/A
Comendador <i>et al.</i> (2015); Case study, Philippines; LMIC	Pediatric generic medicine consultation	Fixed & Free Text/ Text	T&M; parents and children (n=18)	To develop the software application (Pharmabot) that will educate the user on which medication to take for an ailment.	Acceptability: Students rated user-friendliness and consistency of the CA as "Strongly agree" and the suitability of answers and the speed of responses from Pharmabot as "Agree". Experts rated all the above parameters as "Agree"; Self-reported.
Crutzen <i>et al.</i> (2011); Case study, The Netherlands; HIC	Health education	Free Text/ Text	EDU; adolescents (n=737)	To test and evaluate a chat agent designed to " <i>answer adolescents' questions related to sex, drugs, and alcohol</i> "	Acceptability: The conversational agent was favored (48%) over existing modes of information acquisition including search engines (40%), telephone lines (2%), and others (10%); Self-reported. The average duration of chatbot use was 45 minutes; Objective.
Danda <i>et al.</i> (2016), Case study, India, LMIC	Symptom diagnosis	Free Speech/ Text, speech	DG; general population (n=10)	To present a spoken dialogue system Vaidya which can provide a rough diagnosis based on symptoms input by the user in the form of speech via a laptop and/or hand-held device.	Accuracy: 86% of the dialogs led to successful completion with the correct diagnosis. On average, the system used 3.8 steps to reach a diagnosis; Objective.
Denecke <i>et al.</i> (2018), Case Study, Switzerland, HIC	Self-Anamnesis/taking patient history	Free Text/ Text, speech	HS; patients (n=22)	To outline a self-anamnesis conversational agent which allows users to input their patient history before a doctor's appointment	Acceptability: The participants reported mixed responses regarding ease of use, the humanity of conversational agent, and input functionalities; Self-reported.

Elmasri <i>et al.</i> (2016), Case study, Australia, HIC	Alcohol misuse in young adults	Fixed & Free Text/ Text	EDU; young adults (n=17)	To investigate the suitability of conversational agents for a mental health intervention, specific to alcohol drinking habits assessment in young adults	Acceptability: The participants reported higher satisfaction with the use of mental health CA. Positive comments included: informative knowledge base, simple conversation, quick response time, comprehensible, easy to use, personalized, human-like, reliable, accurate, and used layman language; Self-reported.
Fitzpatrick <i>et al.</i> (2017); RCT, USA, HIC	Mental health	Free Text/ Text	T&M; youth with depression and anxiety (n=70)	To test a conversational agent which delivers strategies for self-help using cognitive behavioral therapy, targeted at university students suffering from anxiety and depression	Effectiveness: All the participants in the Woebot group (100%) confirmed that they learned something new whilst only 77% in the control group indicated that they learned something new; Self-reported. Acceptability: Participants in the Woebot condition checked in with the bot on average, 12.14 times over the 2 weeks; Objective.
Fulmer <i>et al.</i> (2018); RCT, USA, HIC	Mental health	Free Text/ Text	HS, T&M: general population (n=74)	To assess the ability of the CA, Tess, to effectively reduce symptoms of anxiety and depression	Effectiveness: Significant differences were found between the control and intervention groups for symptoms of depression and anxiety at baseline and 2-4 weeks after the intervention; Self-reported.
Gaffney <i>et al.</i> (2013); RCT, UK, HIC	Mental health	Free Text/ Text	T&M; adults with mental health (n=42)	To test a conversational agent which employs techniques from MOL (methods of levels) therapy to help participants solve their problems	Effectiveness: Problem-solving and helpfulness scores were significantly higher for a conversational computer-based agent (MYLO) than ELIZA (control). Distress lowering scores were significantly lowered for both conditions with no differential effects between computer programs; Self-reported.
Galescu <i>et al.</i> (2009); Case study, USA, HIC	Health monitoring	Free Speech/ Speech	HS, T&M; patient with heart failure (n=14)	To give an overview of CA CARDIAC, used for health monitoring of heart failure patients and to convey the speech recognition results obtained in a feasibility trial of CARDIAC.	Accuracy: A noteworthy number of errors occurred due to insufficient vocabulary coverage in the language models. Overall, the out-of-vocabulary rate was about 3% during the checkup phase. Nine of the fourteen participants had word error rates under 21% during the checkup phase. The other five subjects had error rates above 30%; Objective.

Ghandeharioun <i>et al.</i> (2018);	Improving wellbeing	Fixed Text/Text	EDU, BC; general population (n=39)	To introduce an emotionally intelligent CA for the general public	Effectiveness: The number of positive emotions experienced by participants was significantly higher for the EMMA condition (M=80.55, SD=3.65) versus the control group (M=69.08, SD=4.16). Additionally, the EMMA group reported a lower percentage of negative emotions. Acceptability: Engagement with the agent was also found to be higher in the EMMA condition; Self-reported.
Ghosh <i>et al.</i> (2018); Case study, Australia, HIC	Triage and diagnosis	Free Text/ Text	DG, HS; general population (n=30 clinical scenarios/vignettes)	To test a conversational agent with symptom checker functions and the ability to provide a medical summary of the patient based on the input information	Accuracy: The conversational agent had a precision rate of 82% in diagnosis and provided a more streamlined alternative to completing patient data entry forms; Objective.
Griol <i>et al.</i> (2015), Case study, Spain, HIC	E-therapy for patients with chronic respiratory diseases	Free Speech/ Text, Speech	HS, T&M, chronic respiratory patients (n=30)	To develop an emotionally sensitive conversational system for patients suffering from chronic pulmonary disease	Acceptability: Emotionally sensitive version was rated significantly higher for procuring requested information, interaction rate, and frequency as well as empathy; Self-reported.
Griol and Callejas (2016); Case study, Spain, HIC	Alzheimer's disease	Free Text & Speech/ Text, Speech, Image	T&M; patients with Alzheimer's disease (n= 25 participants and 6 caregivers)	To propose a framework for the development of a multimodal, context-aware, and user-centric CA for personalized service provision.	Acceptability: Users preferred the multimodal version of the conversational agent (eg. with tactile sensors) and reported higher satisfaction due to more flexibility in usage. Caregivers were satisfied with the technical aspects and believed the activities prompted by the agent were relevant and could stimulate cognitive abilities; Self-reported.
Heldt <i>et al.</i> (2018); RCT, Switzerland, HIC	Overweight management	Fixed Text/ Text	BC; overweight adolescent (n=22)	To explore the use of a text-based healthcare conversational agent to motivate overweight adolescents to make lifestyle changes such as increased relaxation and physical activity	Effectiveness: 43% of adolescents completed their daily challenges successfully. Acceptability: 67% of patients had more than four interactions with the conversational agent per day within 5.5 months and an ongoing RCT on the conversational agent indicates high compliance for over half a year; Objective.
Huang <i>et al.</i> (2015); Case study, China, LMIC	Mental Health	Free Text/ Text	T&M; adolescents (n=10)	To pilot a conversational agent which simulates a virtual companion, available	Accuracy: Conversational agent detected the existence of stress with a precision rate of 78.34%; Objective.

				to listen to and comfort adolescents in stressful situations	Effectiveness: The study reported that 60% of conversations improved the user's level of stress; Self-reported.
Inkster <i>et al.</i> (2018);	Mental health	Free Text/ Text	BC, T&M; individuals with self-reported depression (n=129)	To explore the effectiveness of a text-based conversational agent to deliver positive psychology and mental well-being techniques to individuals with self-reported symptoms of depression	Effectiveness: Improvement in self-reported depression scores (PHQ-9) was significantly higher in the high user's group (mean 5.84, SD 6.66) compared to the low user's group (mean 3.52, SD 6.15). Effect size was moderate = 0.63. An unpaired t-test was also conducted, and this also produced significant results (P=.028). Acceptability: 67.7% of users commented that the app experience was helpful and encouraging; Self-reported.
Joerin <i>et al.</i> (2019), Case study, Canada and USA, HICs	Mental Health Support	Free Text & Speech/ Text, image	HS, T&M; mental health patients and caregivers (n=675)	To present a conversational agent which provides customized psychological support for caregiving professionals, patients, and family caregivers	Effectiveness: Emotional support from the CA decreased depressive symptoms and anxiety respectively by 13 and 18% while costing considerably less than a single therapy session; Self-reported & Objective.
Kamita <i>et al.</i> (2019), randomized controlled trial, Japan, HIC	Mental Health (counselling)	Fixed Text & Image/ Text, image	HS, T&M, adults (n=27)	To propose a new system that uses a smartphone and conversational agent to boost user motivation and stress reduction	Effective: The study reported higher scores in the conversational agent course than the Web course for self-esteem, anxiety, and depression. Acceptability: Scores for the technology acceptance model were higher in the chatbot course; Self-reported.
Kobori <i>et al.</i> (2018); Diagnostic accuracy study; Japan; HIC	Screening for Sexually transmitted infection	Fixed Text & Image/ Text	DG; general population (n=70)	To test the feasibility of a novel conversational agent system to diagnose sexually transmitted infections (STIs) using a smartphone	Accuracy: Conversational agent had 77.7% accuracy in diagnosing Sexually Transmitted Infections (STIs). The accuracy rates of diagnosing gonorrheal urethritis, chlamydial urethritis, syphilis, and condyloma acuminatum were 65%, 70%, 85%, and 95%, respectively; Objective. Effectiveness: After using conversational agent, 97.7% of patients thought to visit the clinic earlier; Self-reported.

Kocielnik et al. (2018); Case study; USA; HIC	Physical activity	Free text/ Text	BC; general population (n=33)	To encourage physical activity in users by employing a deep reflection model with a conversational assistant to induce behavior change	Acceptability: response rates were high (96% at baseline, 90% at follow-up, 11 participants answered all questions) insinuating high engagement throughout the study. There was a decline in response length from 170.1 characters in the first week (SD=31.8) to 138.1 characters in the second week (SD=17.0). Furthermore, usage of the system beyond the 2-week study period was an indicator of viability. 16 out of 33 participants opted to continue without any reward
Kowatsch et al. (2017) [a]; RCT; Switzerland; HIC	Childhood obesity	Fixed Text/ Text	T&M, BC; overweight children (n=15)	To test a text-based healthcare conversational agent/conversational agent (THCB) for use in a childhood obesity intervention	Acceptability: Almost 70% of the patients had at least 4 conversational turns with the THCB per day and 13 patients reported that they enjoyed using the THCB chat; Objective. The findings on attachment bonds indicated a high degree of social and emotional relationships at the end of the intervention; Self-reported.
Kowatsch et al. (2017) [b]; Controlled before and after study, Switzerland; HIC	Behavior change	Fixed Text/ Text, image	BC; adolescents (n=11)	To test a rule-based conversational agent employed to coach adolescents on how to manage obesity	Acceptability: the acceptability by patients was generally very high. Items measured were: Perceived ease of use (mean 6.7, SD 0.7, p<0.001), perceived enjoyment (mean 6.2, SD 1.5, p<0.01), perceived usefulness (mean 5.9, SD 1.1, p<0.05), intention to use (mean 5.6, SD 1.4, p<0.05) and some usability problems; Self-reported.
L'Allemand et al. (2018), randomized controlled trial, Switzerland, HIC	Childhood obesity	Fixed Text/ Text	HS, BC; overweight adolescents (n=20)	To test whether a novel CA app for overweight adolescents, can motivate its users to make lifestyle changes regarding relaxation and physical activity	Acceptability: 70% of the participants had >4 text-based healthcare conversational turns per day, and 37% of the daily challenges were completed; Objective.
Liu et al. (2018), RCT, USA, HIC	Health information advisor	Fixed Text/ Text	HS, T&M; general population (n=158)	To evaluate a conversational agent which offers both emotional and informational support	Acceptability: Support with sympathy, affective empathy, and cognitive empathy were more positively evaluated than advice-only support in Study 1. Similar findings were reported in Study 2. In both studies, expression of sympathy or

					affective empathy did not increase eeriness compared with advice-only support; Self-reported.
Lobo <i>et al.</i> (2017)	Heart medication information	Free speech/ Text, speech	EDU; Heart failure patients (n=11)	To present a conversational medication assistant designed to provide advice and information on medication-related queries for heart failure patients	Acceptability: Average scores were given from 1-4 for usability (3, Mean deviation = 0.87), Coherence (4, MD = 0.63), Naturalness (3, MD = 0.68), Quality of information (3, 0.74), HF-directed usability (3, MD = 0.80). In general, user engagement with CARMIE was successful. The responses were coherent, short, and easy to understand. System usability scale (SUS) score of 88, which is equivalent to a 'very good' or 'B' rank score; Self-reported.
Ly <i>et al.</i> (2017); RCT; Sweden; HIC	Mental Health	Fixed & Free Text/ Text	T&M; young adults (n=28)	To study how effective a smartphone integrated conversational agent could be in conveying advice and education regarding positive psychology and CBT to young adults	Effectiveness: There was a significant difference in the score changes of self-esteem (p=0.024), state-trait anxiety inventory (STAI) (p=0.038), and Self-rating depression (SDS) (p=0.043) in favor of conversational agent course; Self-reported. Acceptability: The participants were highly engaged with the CA during the 2-week study period, evidenced by an average open app ratio of 17.71 throughout the study; Objective.
Middleton <i>et al.</i> (2016); Diagnostic accuracy study; UK; HIC	Automatic symptoms checker	Fixed Text/ Text	DG, HS; general population (n= 12 clinicians and 17 nurses)	To assess the safety, efficacy, and speed of a conversational agent designed for medical triage as compared to doctors and nurses using clinical vignettes	Accuracy: Conversational agent achieved 88.2% accuracy in triaging compared to 75.5% in doctors and 73.5% in nurses. The conversational agent provided a clinically safe outcome in 100% of cases and was faster than healthcare professionals in 89% of cases; Objective.
Miner <i>et al.</i> (2016); Case study; USA; HIC	Health Education	Free Speech/ Speech	EDU; general population (n=68 phones)	To compare the replies provided by four well-known conversational agents (Siri, Cortana, S Voice, Google Now) when asked questions about "mental health, interpersonal violence, and physical health"	Effectiveness: Siri's responses were empathetic for the topics depression and physical health and Cortana for matters involving interpersonal violence; Objective.

Mujeeb <i>et al.</i> (2017), Case study, Pakistan, LMIC	Diagnostic Conversational agent for Achluophobia and Autism	Free Text/ Text	DG, HS; patients with autism disorders (n=30)	To evaluate the use of a conversational agent in the diagnosis of achluophobia (fear of darkness) and autism disorder	Accuracy: Diagnostic accuracy of the conversational agent for diagnosis of achluophobia was 85%, 86.64%, and 87.2%, for three different age groups (ages 18-28). Accuracy of autism disorder diagnosis was 88%, 87.6%, and 87.53% for another 3 patient groups (ages 1-7); Objective.
Nadarzynski <i>et al.</i> (2018); Case study; UK; HIC	Sexual Health Education	Unspecified/ Unspecified	EDU; general population (n=148)	To employ conversational agents to increase public awareness regarding good practice for sexual health in terms of education and healthcare services	Acceptability: Participants' willingness to use the conversational agent was quite low (40%) and they cited privacy and wanting to interact with physical healthcare professionals as reasons for their reservations; Self-reported.
Ni <i>et al.</i> (2017)	Automating patient intake	Free Text/ Text	HS, DG; general population (n=11)	To propose a CA which can assist healthcare staff by automating the patient intake process	Accuracy: The CA was assessed for the question and prediction accuracy. Question accuracy = matched questions/expected questions. Prediction accuracy = matched hypotheses from our system/diagnostic hypotheses in the guidebook. Question accuracy was 100% for respiratory issues, chest pain, and headache but only 66.7% for dizziness. Prediction accuracy was 100%, 64%, 25%, and 14% for respiratory issues, chest pain, headache, and dizziness respectively; Objective.
Razzaki <i>et al.</i> (2018); Diagnostic accuracy study; UK; HIC	Diagnosis and Healthcare Support	Free Text/ Text	DG, HS; general population (n= 7 doctors, 100 vignettes)	To test the safety and efficacy of an online symptom checker as compared to a human doctor using clinical vignettes	Accuracy: the conversational agent was able to produce differential diagnoses with precision (44.4% vs 43.6) and recall comparable to that of doctors (80 % vs 83.9%). Safety: Conversational agent provided safer triage recommendation than doctors on average (97.0% versus 93.1%), at the expense of marginally lower appropriateness (90.0% versus 90.5%); Objective.
Rhee <i>et al.</i> (2014), Case study, USA, HIC	Asthma management	Free Text/ Text	HS, T&M; adolescents with asthma (n= 15 adolescent-parent dyads)	To describe the development and evaluation of a mobile phone-based asthma self-management aid for adolescents (mASMAA) designed to	Effectiveness: mASMAA help raise awareness of symptoms and triggers, promoted treatment adherence, a sense of

				monitor symptoms, treatment adherence, and nurture the adolescent-parent partnership.	control, and facilitated adolescent-parent partnerships; Self-reported. Acceptability: average response rates to questions asked by the bot were high (81%-97%). Generally, response rates in the morning were higher than at bedtime. Each adolescent on average submitted 19 self-initiated text messages (range 3-38) and symptom-related messages were the most common topic of teen-initiated messages, followed by activity-related and medication-related messages; Objective.
Shaikh <i>et al.</i> (2018), Before and after study, India, LMIC	Diabetes management	Free Text/ Text	T&M; diabetic patients (n=102)	To report results that aim to validate Wellthy Diabetes as an agent for diabetes management and risk reduction	Effectiveness: 102 participants reported a mean post-intervention reduction of -0.49% A1c (95% CI:-0.73 - -0.25, p=0.01) with the use of conversational agent. Sixty percent (n=61) of participants achieved a greater reduction in their A1c with a mean reduction of -1.17% A1c (95% CI: 0.75-1.60); Objective.
Stasinaki <i>et al.</i> (2016), Pilot study, Switzerland, HIC	Obesity	Fixed Text/ Text	BC; obese adolescents (n=21)	To test the feasibility of a novel health application game as a therapeutic option for adolescents with obesity.	Effectiveness: Physical performance improved significantly in 5 of 8 tests in the intervention group (agility, flexibility, strength) and only in 2 tests in the control group (agility, endurance). There was no difference between the groups for BMI; Objective.
Stein <i>et al.</i> (2017), Longitudinal observation study, USA, HIC	Diabetes prevention	Free Text/ Text, images	BC; overweight and obese adults (n=70)	To evaluate weight loss, changes in meal quality, and app acceptability by users of the weight loss and health coach app – Lark.	Effectiveness: The participants reported 2.38% weight loss compared to baseline weight and the percentage of healthy meals increased by 31%; Objective. Acceptability: The average duration of app use was 15 (SD 1.0) weeks, and users averaged 103 sessions each. In-app user trust survey had a 100% response rate and positive results, with a satisfaction score of 87 out of 100; Self-reported.
Van Heerden <i>et al.</i> (2017), Case study, South Africa, LMIC	HIV counseling	Free Text/ Text	EDU; adults (n=10)	To describe a CA that guides the user through the counseling phase before an HIV test	Acceptability: the participants reported a high acceptability rate with the use of a conversational agent and the majority (6/10) thought that the familiarity with a chat interface, avatar image, and

					appropriate responses made the interaction feel like talking with a human.
Vita et al. (2018), before and after study, Italy, HIC	ART adherence and HIV management	Free Text/ Text	HS, T&M; patients (n=34)	To design, develop and evaluate a digital health healthcare personnel-patient interface for better long term management of HIV infection	Acceptability: The conversational agent was used for a variety of functions: 24.6% to book appt., 22% to get results of exams, 20% for therapy, 10.2% for upcoming event updates (appts), and 13.6% for other activities e.g. start, free interaction, and cancel. The “Medication reminder” function was used in less than 10% of days; Objective.
Wang et al. (2018); RCT; Hong Kong; HIC	Smoking Cessation	Free Text/ Text	BC, EDU; adult smokers (n=401)	To assess the performance of a social-media-driven conversational agent employed for a smoking cessation program	Effectiveness: Weekly questionnaire indicated that 38.3 percent reported not having smoked in the past week, while 69.4% reported that the frequency of smoking had been reduced during the past week; Self-reported. Acceptability: the average number of messages per day for each group (I/C) was 5.24 when the CA was not activated and 8.45 when the CA was activated. More than half of the conversations were initiated by the CA. Content-specific interactions were more effective than scheduled plans in stimulating user activities; Objective.
Wilson et al. (2017); Case study; New Zealand; HIC	Sexual Health Education	Free Speech/ speech, text, image, video	EDU; general population (n=50 questions)	To assess the quality of the information about 50 questions concerning sexual health provided by Siri and Google assistant compared to a Google search	Accuracy: Google search performed better than the two digital assistants, by offering 72% (36/50) of the best (or equal best) responses for the questions on sexual health. Google Assistant exceeded Siri in its performance with 50% of best (or equal best responses) versus 32% by Siri; P=0.036). Google searches also had the lowest rate of failure, whereby they offered no useful response for only 8% (4/50) of the questions compared with 12% and 36% for Google Assistant and Siri, respectively; Objective.

Outcome measures of focus were categorized as Effectiveness, Accuracy, and Acceptability.

^aHIC: high-income country.
^bTDV: technical development.
^cHS: health care support.
^dUS: Unites States.
^eEDU: education.
^fIQR: interquartile rate.
^gBC: behavior change.
^hN/A: not applicable.
ⁱLMIC: low- and middle-income country.
^jT&M, treatment and monitoring.
^kDG: diagnosis.
^lRCT: randomized controlled trial.
^mCBT: cognitive behavioral therapy.
ⁿUK: United Kingdom.
^oMOL: methods of levels.
^pPHQ-9: patient health questionnaire-9.
^qSTI: sexually transmitted infection.
^rmASMAA: mobile phone–based asthma self-management aid for adolescents
^sA1c: glycated haemoglobin. Indication of blood sugar levels.

4.4 Discussion

4.4.1 Key findings

This scoping review identified 45 studies and two ongoing clinical trials. As much of the literature was published in recent years, it reiterates the point that the use of CAs in healthcare is still in its infancy. Most conversational agents used text-input and were machine learning-based and mobile-application delivered. The three most frequently targeted healthcare themes were treatment and monitoring, health services support, and patient education. Results on effectiveness, accuracy, and acceptability were generally positive. However, it was noted that there is currently a dearth of robust evaluations and a predominance of small case studies.

4.4.2 Comparisons with literature & Implications for the future

Most of the existing healthcare CAs reported in the literature used machine learning and were long-term goal-oriented. This suggests that CAs are evolving from conducting simple transactional tasks towards more involved tasks such as long term-disease management [34] and behavior change [134]. The main target audience identified in this review were patients, with only a handful aimed at healthcare professionals, for processes like automating patient intake or aiding in patient triage and diagnosis. In addition, future research can look into extrapolating the use of CAs to support both formal and informal caregivers given that previous systematic reviews into the use of digital technology for caregivers of patients with psychosis [182] or dementia [183] have shown positive outcomes.

These findings show a predominance of text as an input and output modality for CAs with only a few applications using speech. Yet certain populations, such as older people, maybe more comfortable interacting via speech as some individuals may find the dexterity involved with typing on small keypads on smartphones challenging. Furthermore, most conversational agents included in this review were app-based. Research shows that the use of apps (which need to be downloaded and regularly updated) is often associated with high dropout rates and low utilization [184]. The adoption barrier may be lower with messaging apps such as Facebook Messenger, iMessage, Telegram, WeChat, or WhatsApp, already commonly used in the general population. Future research could explore methods to overcome this limitation brought on by smartphone applications by embedding healthcare CAs in platforms with which the target population is already closely acquainted. The multiple options of delivery channels to choose from are an advantage for CAs.

A recent systematic review outlined the lack of an established method for evaluating healthcare CAs and a dearth of data on adverse effects [41]. This corresponds to our findings with most

studies being case studies and lacking information on potential adverse effects. Side effects may relate to the content of the CAs which may not be accurate, evidence-based, or suitable for the specific circumstance. For example, if a mental health CA user has suicidal tendencies, the CA may not be best equipped to handle such a situation and may provide inappropriate advice, leaving the user at fatal risk. Additional unwanted effects could arise from the black box effect associated with the use of machine learning-based conversational agents whereby the source of their suggestions are sometimes untraceable and hence, cannot be reasonably explained [185]. Furthermore, CAs allowing users to type free text may lead to privacy concerns as individuals can overshare private and sensitive data in conversations [186]. Compliance with mHON code principles could potentially address this issue [100].

The healthcare themes for CA application identified were very broad with references to only a few specialties including mental health [187], neurodegeneration [158], metabolic medicine (obesity [151], and diabetes [188]), and sexual health [75]. Future applications could expand towards other healthcare fields where evidence has suggested the potential for digital health interventions such as dermatology [184], primary care [41], geriatrics [185], and oncology [189, 190].

There is also a need for more geographically diverse research. In this review, 12 articles were based in Asia, the evidence from middle-income countries was scarce and there were no studies from a low-income country. Yet, digital health initiatives are becoming more common in developing countries, through the use of social media, for example [189]. To ensure the safe and effective use of solutions developed in high-income country settings, it may be beneficial to conduct more research in developing low-income and middle-income countries.

Furthermore, it is important to explore a “hybrid system”, where digital technology supplements, existing healthcare services for example where CA users also have frequent meetings and phone calls with the physicians. [190]. In most studies, CAs were developed and presented independently, unsupported by humans, and separate from the existing healthcare delivery models which may prove unsustainable in the long run.

While the studies reported accuracy, efficacy, effectiveness, and acceptability as outcomes, there were no measurements of cost, efficiency, or how the solution led to improved productivity when used instead of or to support the work of a health professional. Hence, cost-effectiveness could not be ascertained.

4.4.3 Strengths and Limitations

In this scoping review, a comprehensive literature search was conducted of multiple databases including grey literature sources prioritizing sensitivity over specificity to capture a holistic representation of CA usage in healthcare. However, considering the novelty of the field and the evolving terminology, some unpublished studies discussed at niche conferences or scientific seminars may have been omitted. Furthermore, while the CA healthcare themes here are based on thorough analysis, team discussions, and consensus, they might not be all-inclusive and may require further development with the advent of new CAs. In addition, while some CAs belong to more than one theme, they were mostly classified based on the dominant mode of application for the sake of clarity. Finally, articles with poorly reported evaluation data were excluded so certain applications may have been missed. For example, articles with no quantitative data or results presented in one or two sentences were excluded because they did not appear to contribute much additional value to the review. The general notion towards these articles was that more research and more rigorous evaluations with properly defined outcomes were needed. The personality traits presented were guided by a reference paper on chatbot personality assignment [127] and also through condensation of descriptive terms from several articles. The lack of detail in the description of the content and development of many conversational agents, lead to the organic development of a framework for this paper. This personality framework is, therefore, still exploratory, suited for this review but needs to be explored and further refined with more in-depth analysis such as previously published frameworks [191].

4.4.4 Conclusion

CAs are an up-and-coming form of technology to be used in healthcare that has yet to be robustly assessed. Most CAs reported in the literature to date are text-based, employ machine learning, and are delivered through smartphone apps. Future research should focus on assessing the feasibility, acceptability, safety, and effectiveness of diverse CA formats aligned with the target population's needs and preferences. There is also a need to explore the role of CAs within existing health systems and clearer guidance on healthcare-related conversational agents' development and evaluation.

Chapter 5 – Designing and implementing a mobile phone-delivered, rule-based conversational agent: a conceptual framework

This chapter achieves objective 2 which is to, “propose a systematic list of steps on how to develop a rule-based text healthcare conversational agent and present it in a conceptual framework, using evidence from the literature and a validation study.” Section 5.1 provides some background information and rationale for this project. Section 5.2 elaborates on the methodology encompassing two literature searches and a study to validate the framework. The results of the literature reviews, the actual conceptual framework, and the validation study are presented in section 5.3. Key findings, comparisons with existing literature, and implications for the future are described in the discussion in section 5.4.

5.1 Introduction

The scoping review in chapter 4 noted the heterogeneity in the design, development and evaluation of CAs. Similarly, a systematic review by Laranjo et al. reported that majority of studies reporting AI CAs were quasi-experimental studies (as opposed to robust clinical testing) and sometimes omitted efficacy and safety in their evaluations [41]. The literature indicates a clear inconsistency in the CA development and testing process and this poses as a caveat for comparisons between CAs in various studies to identify strengths and areas for improvement.

In a systematic review assessing the quality of mHealth apps, it was noted that despite the vast number of apps entering the market ($n > 325,000$), about a quarter of them do not get installed [192]. Reasons cited were low quality, lack of rigorous supporting evidence and app creation in the absence of careful considerations of the needs and characteristics of the target population [192]. Potential hazards and risks associated with the apps were also sometimes overlooked. This stimulated discussions around the safety and quality of upcoming apps and a potential insistence on regulatory oversight [192]. This emphasizes the consequence of a lack of substantiated guidelines for mHealth intervention development. An evidence based conceptual framework for CA development could provide sufficient guidance on design, development and evaluation to overcome the limitations associated with mHealth interventions developed without a proper, comprehensive framework.

Currently, multiple mHealth (mobile health) frameworks exist, for various general mHealth applications [110, 193-195]. However, CAs are a very specific type of mHealth intervention and thus require an individualized framework. Such a framework would guide the design and development of these agents to fully harness their capabilities which are centered around

conversing, CA personality development and rapport building, encouraging behaviour change, accessing healthcare resources etc. The increasing number of publications in this research area indicates the growing interest in healthcare CA development (Chapters 1, 4). Consequently, the demand for evidence backed, step-by-step guidance on the development, implementation, and evaluation of CAs for health-related purposes is escalating and thus, such a conceptual framework is justified at this point. While CAs can be delivered via various platforms, there is a growing interest in mobile phone based CAs, given their high level of accessibility, convenience, and scalability.

5.1.1 Objective

As our knowledge evolves in this young research area, investigators are urged to record and publish their evidence-based methods for developing interventions to help advance the field [196]. As such, this project aimed to inform the design and development of smartphone-delivered, rule-based healthcare CAs which are goal-oriented and use text as the mode of communication through a conceptual framework

5.2 Methods

The eight steps for conceptual framework development as outlined by Jabareen et al. were employed [102] (Appendix 5.1).

5.2.1 Mapping the data sources, categorizing data, and identifying concepts

Two literature searches were conducted: (1) a systematic search and analysis of frameworks for mHealth and CAs and (2) an updated systematic search for conversational agents in healthcare using the search strategy from a previous scoping review [42].

5.2.2 Literature review of mHealth frameworks for the development of mobile health interventions

A systematic search in Medline (Ovid) was conducted in April 2019 and updated in December 2020 using the search terms: “mHealth”, “texting”, “text messaging”, “short message service OR SMS”, “chatbot”, “conversational agent” AND “framework”. Backward and forward citation searches were also performed. Eligible studies were to be in English, be published in a peer-review journal, and outline a set of principles for the development and implementation of mHealth interventions. MHealth was defined as those interventions which employ mobile phones or tablets to support healthcare delivery [197]. Studies were included if they discussed the development of a new framework from scratch or if they adapted an existing framework in a novel manner. This could involve the combination of more than one framework or application of an existing framework for more than one of the following: behavior change, personalization,

or a user-centered approach as these are of importance in CA interventions. Data was extracted on the name of the framework, aims, steps, and the healthcare application. The search, screening, and data extraction were performed in parallel by two review authors. Any discrepancies were resolved through discussion and consensus.

5.2.3 Scoping review of conversational agents

An updated search using the search strategy comprising 63 search terms from a previous scoping review of CAs (Chapter 4) was conducted [42]. These search terms included “conversational agents,” “conversational AI,” “chatbots,” and associated synonyms. Screening and data extraction were performed in parallel by 2 reviewers. As the original literature search was conducted in April 2019, the search was updated in Dec 2020 in PubMed.

Studies were included if they reported on the development and design of rule-based, text-based CAs delivered via a smartphone. Studies were excluded if they described speech-based or embodied conversational agents, used AI or machine learning programs with predesigned and optimized frameworks (e.g. IBM Watson or Watson assistant, Siri, Google Assistant, etc.), included a diagnostic or symptom checker, questionnaire, or online form that didn’t engage in conversation with the user, and papers discussing technical development of CAs with no information on the development of content. The extracted data was compartmentalized using thematic analysis.

5.2.4 Deconstructing, integrating, and synthesizing concepts

The method of meta-synthesis, described in Jabereen’s recommendations, was employed to synthesize concepts and steps for the conceptual framework [102]. This involved grouping the concepts extracted from both literature reviews into overarching domains. These domains were then linked and presented diagrammatically. Following discussion within the research team, the concepts and domains were further amended and presented at scientific seminars and conferences for feedback from colleagues. This led to the refinement of concepts and domain labels, definitions, order, and grouping.

5.2.5 Validating the framework - Formative research

The resulting framework was used to develop a rule-based, text-based, smartphone-delivered conversational agent prototype, designed to support healthy lifestyle change and educate participants about healthy living and diabetes prevention. The conversational agent (*Precilla*) delivered advice on healthy living over four weeks with a focus on healthy eating, physical activity, stress management, and healthy sleeping habits. The feasibility and acceptability of *Precilla* were tested in an online, single-arm feasibility study [198].

5.3 Results

5.3.1 Literature review of mHealth frameworks

The initial search for mHealth frameworks retrieved 749 citations, of which 23 were relevant and included. The updated search in Dec 2020 (including snowball searching) resulted in an additional 6 frameworks. In total, data was from 29 studies to inform this conceptual framework (Figure 5.1). The list of excluded studies and reasons for exclusion are presented in Appendix 5.2.

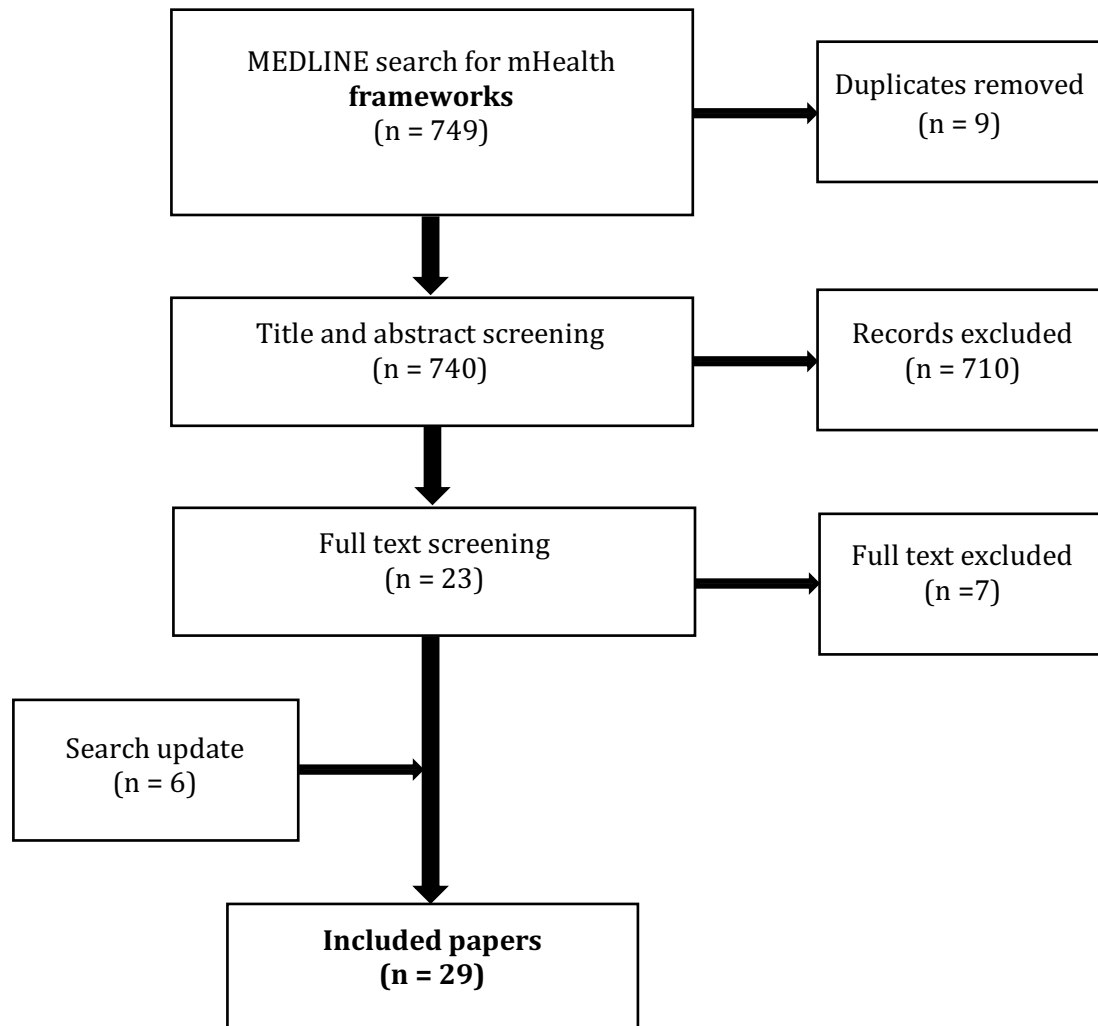


Figure 5.1 PRISMA flowchart of mHealth development frameworks search

The identified frameworks covered these digital interventions: (1) mHealth (n=10), (2) text messaging (n=3), (3) smartphone applications (apps) (n=10), and (4) conversational agents (n=6). Tables 5.1 and 5.2 present a summary of the included frameworks.

Table 5.1 mHealth, text messaging, and smartphone app frameworks and their key features.

Study ID Country of origin Study methods /Framework adaptation	Healthcare topic(s)	Name of framework	Digital technology (app, SMS, etc)	The main conceptual framework steps or features
Depp 2016; USA; NA [199]	Bipolar disorder	NA (New framework)	Smartphone apps; text messaging; home-based telehealth; wearable sensors	The five main steps were: (1) selecting a platform for data collection (2) defining inputs and outcomes patient-reported or sensors; (3) setting up predictive analytics ; (4) outlining decision rules that determine the thresholds and actions and (5) providing feedback .
Direito 2018; New Zealand; a combination of existing frameworks [200]	Increase levels of physical activity	Intervention Mapping Framework (IM); Behavioral Intervention Technology Framework (BIT)	Smartphone app	The four steps addressed in this paper were: (1) conducting a needs assessment, (2) determining goals and objectives, (3) identifying theory-based and practical methods to target behavior, (4) producing program components (app features).
Fedele 2019; USA; based on an existing framework [193]	Pediatric psychology	IDEAS framework	Smartphone app	The framework indicated a need to gather an interdisciplinary team before starting the design process using IDEAS (Integrate, Design, Assess, and Share).
Horvath 2015; USA; based on existing framework [201]	Alcohol policy, adolescent health, medication adherence, and HIV prevention with men who have sex with men (MSM) and transgender persons.	User-Centered Design Process Map	Smartphone app	The framework suggested six main steps: (1) Hiring a strong research team, (2) planning before designing, (3) bridging the gap between researchers and users, (4) developing a detailed contract, (5) documenting the whole process, (6) having a content management system, (7) budgeting sufficient and extra time for testing and debugging of any issues.
Jennings 2019; Bangladesh; a combination of existing frameworks [195]	Diabetes	COM-B (Capability, Opportunity, Motivation, Behavior) model of behavior change; Transtheoretical Domains Framework (TDF)	mHealth (Voice messaging)	The six steps outlined in the framework were: (1) Defining the context of the intervention, (2) Define outcomes, (3) List the enablers and barriers to behavior change, (4) Categorise the barriers and enablers according to COM-B and TDF, (5) Suggest behavior change approaches for each enabler and barrier, (6) Tabulate message content based on the intended outcomes, barriers and enablers, and BCTs.
Mohr 2014; USA; literature review [202]	Behavior change	Behavioral intervention technologies framework (BIT-Tech)	Smartphone app	The five main steps mentioned were: (1) Why – defining the aims of the intervention (2) How – behavior change strategies to achieve aims (3) What – defining elements of the intervention eg. information delivery, messaging, reports, (4)

				Outlining technical characteristics, (5) When – timing workflow
Mummah 2016; USA; literature review & collaboration among subject matter experts [203]	Behavior change (promote vegetable consumption in overweight adults)	IDEAS framework	Smartphone app	The framework had these 10 steps: (1) Empathise with target users (through qualitative insights), (2) Indicate highly specific target behavior, (3) Ground in behavioral theory, (4) Brainstorm creative implementation strategies, (5) Prototype potential products, (6) Gather user feedback (on the prototypes), (7) Build a minimum viable product, (8) Pilot efficacy and usability, (9) Evaluate efficacy in a randomized controlled trial (RCT), (10) Share intervention & findings through publication.
Whittaker 2012; New Zealand; literature review [196]	smoking cessation; depression	mHealth development and evaluation framework	Video messaging, Multimedia messaging	The framework outlined these 6 steps: (1) Performing formative research, (2) Pretesting: online surveys, focus groups, individual interviews, (3) Performing a pilot study (small and non-randomized), (4) RCT, (5) Performing qualitative research (for further improvements), (6) Performing an evaluation of impact.
Wittenberg 2019; USA; NA [204]	Support for informal caregivers (cancer)	NA (New framework)	Smartphone app	The six steps mentioned were: (1) Reviewing existing print resources, (2) Selecting a relevant theoretical framework for content development, (3) Integrating stakeholder feedback and literacy assessment of target population, (4) Reviewing existing mHealth resources, (5) Developing a prototype, (6) Evaluating caregiver acceptability and quality/perceived impact.
Yardley 2015; United Kingdom; based on existing framework [205]	Behavior change (online weight loss intervention)	Person-based approach	Smartphone app	The framework described 4 steps: (1) Planning (months 0-6) – identify necessary behavior changes, needs, and challenges, (2) Designing (months 3-9) – identify suitable theoretical models to help developers create features for the intervention which address the objectives, (3) Developing and pilot testing (months 6-18), (4) Implementing and trial testing (months 12-18) – evaluating the intervention in a real-life context, and identifying points of improvement for future iterations.
Duff 2018; Ireland; based on existing framework [206]	Patient Self-Management (via exercise-based rehabilitation) of cardiovascular disease (CVD)	mHealth development evaluation framework	Smartphone app	The framework proposed these 2 phases - Phase 1: a systematic review (review literature & BCTs) and consultation with advisory panel (multidisciplinary research team of experts in relevant areas). Phase 2: usability and acceptability testing.
Pham 2018; Canada; based on existing framework [207]	Self-management of chronic pain	Analytics Platform to Evaluate Effective Engagement (APEEE) guided by the UCD (user-	Smartphone app	Indicators to be considered in the analysis were: (1) Health status (how users are managing with pain-related outcomes), (2) Check-ins (number of daily check-ins), (3) Goals (are users completing set goals), (4) Community (are users engaged with

		centered design) framework		community features?), (5) library (most popular library articles), (6) history (what is being reviewed in the history feature), (7) others.
Ravn Jakobsen 2018; Denmark; based on existing framework [208]	Self-management of osteoporosis	Participatory design process	Smartphone app	The paper described three main steps: (1) Identifying users' needs, (2) Designing and developing prototypes, (3) Testing the mHealth app (evaluation and re-testing). This design requires the involvement and collaboration of users and health IT designers. Participatory design is usually an iterative process.
Schellong 2019; Germany; based on existing framework [209]	Post-traumatic stress disorder (PTSD)	Information system research framework and requirements of international standards (most importantly IEC 82304-1:2016)	Smartphone app	The framework comprised these 6 steps: (1) Development (technical norms, IT structure, Healthcare goals, specific for PTSD, Medical norms), (2) Beta test and pilot test with patients, (3) Clinical trial/RCT studies, (4) Evaluation assessment, (5) Release mHealth app, (6) Postmarket communication and localization.
Schnall 2016; USA; based on existing framework [111]	HIV prevention for high-risk MSM (men who have sex with men)	Information Systems Research (ISR) framework	Smartphone app	The framework comprises three cycles: (1) relevance cycle - understanding the environment of the end-user and their requirements; (2) design cycle - producing prototypes; (3) rigor cycle - evaluating them.
Tay 2017; Australia; evidence-based research/guidelines [210]	Prevention of osteoporosis by monitoring calcium intake in young women	NA	Smartphone app	Four steps were proposed: (1) Conceptualising, (2) Developing and pretesting, (3) Pilot testing, and (4) Performing mixed methods evaluation (eg. questionnaires, focus groups).
Van Agteren 2018; Australia; a combination of existing frameworks [211]	Smoking cessation	Intervention Mapping framework; Behavioural Change Taxonomy v1; Theoretical Domains Framework; and Persuasive System Design framework	Smartphone app	Five phases were outlined: (1) Needs assessment of health behavior, (2) Defining program objectives, (3) Selecting evidence-based methods (BCTs) and practical applications to address each objective, (4) Developing the mHealth app which provides support and feedback, (5) Adoption, implementation, and evaluating the app.
Vilardaga 2018; USA; based on existing framework [212]	Smoking cessation (for people with serious mental illness)	User-centered design framework	Smartphone app	Four phases were shared in the framework: (1) Understanding and defining the context of use, (2) Specifying user and organizational requirements and matching them to BCTs, (3) Producing design solutions, (4) Evaluating the design.
Wilhide 2016; USA; evidence-based	Chronic disease (diabetes, epilepsy, asthma, chronic	NA	Smartphone app	The framework had five steps:

research & collaboration with subject matter experts [213]	obstructive pulmonary disease, lupus, HER2+ breast cancer, and low back pain)			Identifying value drivers (entities that increase the value of a product or service), (2) Defining outcome measures of interest, (3) Identifying those key stakeholders the system is designed to support, (4) Intervention planning – BCTs and personalization of user's experience through the delivery of tailored message content, ensure intervention is evidence-based, (5) translating the intervention into a format which suits app content.
Woods 2018; Australia; based on existing framework [214]	Heart failure self-management	Stanford University's Design Thinking process	Smartphone app	The proposed framework had 5 steps: (1) Empathising (needs assessment with patients and their carers), (2) defining (the healthcare problem and list potential opportunities), (3) Ideating (strategies for heart failure management), (4) Developing a prototype (select the best ideas and make a visual prototype), (5) Testing (using a validated tool to assess usability).
Fjeldsoe 2012; Australia; based on existing framework [215]	Increase levels of physical activity (in a population of mums with young children)	mHealth development evaluation framework	SMS (short message service)	The five steps in this framework were: (1) Conceptualising – scoping for relevant literature and theory, (2) Conducting formative research (eg. focus groups), (3) pre-testing, (4) Pilot testing, (5) Performing qualitative evaluations of the refined intervention.
Good 2019; United Kingdom; a combination of existing frameworks [110]	Improving healthcare delivery to rural areas; provide advice to pregnant women via SMS	Activity theory and user-centered design (UCD) framework	SMS	The framework included five phases: Identifying the need for human-centered design, (2) Understanding and specifying the context of use, (3) Defining the user and organizational requirements, (4) Proposing design solutions, (5) Evaluating designs based on requirements.
Skrabal Ross 2019; Australia; a combination of existing frameworks [216]	Chemotherapy and medication adherence for people with cancer (adolescents and adults)	(BIT) model and mHealth development & evaluation framework	SMS	The framework proposed three main stages: (1) conceptualizing (literature/scoping review and identification of relevant behavioral change strategies); 2) defining features and structure of the program based on a needs assessment with end-users; and 3) selecting an appropriate program delivery technology, depending on availability and relevance to the target population.

The mHealth frameworks discussed these healthcare topics: behavior change, self-management of chronic conditions, HIV prevention, smoking cessation, mental health, and support for caregivers. Thirteen papers reported the development of a new framework, while 11 papers applied a pre-existing framework, and five studies combined two or more existing frameworks in a new intervention.

The methods used to develop these frameworks were: an expansion of existing framework(s) (n=16) [110, 111, 193, 195, 200, 201, 205-209, 211, 212, 214-216]; literature review (n=6) [196, 202, 217-220], systematic review [221], collaboration among researchers, subject matter experts, engineers and designers [213]; evidence-based guidelines and recommendations [210]; and empirical observation [222]. One study employed a combination of literature review and collaboration with subject matter experts [203]. The methods for framework development were not disclosed in two papers [199, 204] (Table 5.1).

In studies which reported on the methodology for framework development, a popular first step involved conducting an initial (whether a literature or systematic) review. This served a dual purpose of (1) defining the goal by identifying gaps in the evidence and (2) forming a database of information available to inform the content of the intervention [196, 220]. Next, some form of interaction with the target users (qualitative study or co-design step) was implemented to help define the target population and identify their needs so that content could be appropriately developed to suit their requirements [203, 207]. Having the users' target profile could also indicate which delivery platform should be employed (eg. standalone app or existing messaging app) [199]. Further, co-design was pivotal in making the intervention relevant and useful to the user and could also reveal which health behaviors should be targeted for change [214, 216]. As most of the included frameworks (n=15, 65%) discussed design, development and evaluation, a pivotal, final phase entailed evaluating a prototype through, pilot testing of usability, feasibility, and acceptability. This helped identify any bugs or necessary changes that needed to be made to improve the conversational agent before testing it in a larger randomized trial [206, 215].

As such, the key features identified collectively from these studies were: conducting a literature review, defining the target population, selecting a delivery platform, co-design, behavioral change, developing an initial prototype, and subsequently conducting a larger-scale evaluation (Figure 5.2).

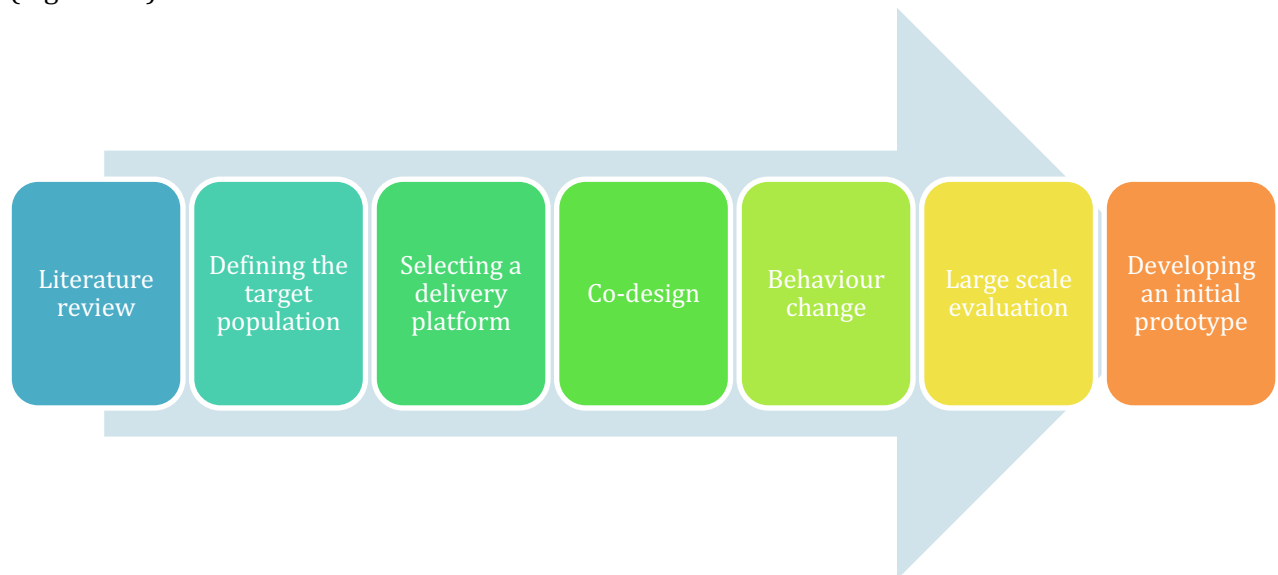


Figure 5.2 Key steps prioritized in the collated mHealth frameworks

The methodology used in CA framework development was sometimes similar to the mHealth frameworks, regarding suggestions to conduct a review of existing agents, build up the profile of the ideal user, and conduct evaluations of the agent where possible (Table 5.2). Unlike mHealth frameworks, the CA frameworks included additional pointers such as building a social bond between the user and CA (eg. rapport), placing importance on user experience (eg. auditory, visual considerations, invisible measures like timing and personality), and considered technical details (eg. chat script).

Table 5.2 Conversational agent frameworks and their key features.

Study ID Country of origin Study methods /Framework adaptation	Healthcare topic(s)	Name of framework	The main framework steps or domains
Diederich 2019; Germany; empirical observation [222]	NA (not restricted to healthcare)	The Nickerson Method - combining existing conceptual knowledge with empirical observation	The framework considers the following domains in the design: (1) Communication mode (text, speech, both); (2) Context (general purpose/domain-specific); (3) Language (single/multiple languages); (4) Intelligence (Rule-based/self-learning); (5) Implementation (programming, modeling, supervised learning, hybrid). Hosting (on-premise/cloud/both). Reporting (without reporting, with reporting); (6) Pricing model (usage-based, user-based, instance-based, free); (7) Sentiment detection (with/without sentiment) (8) Enterprise integration; (9) Platform integration (single/cross-platform)
Hussain 2019; Australia; literature review [218]	NA (not restricted to healthcare)	NA	The framework proposes the following classification and design techniques: 1. Rule-Based Approaches, 2. Retrieval-Based Approaches, 3. Generative-Based Approaches. With each of the approaches using any of these techniques: 1. Parsing, 2. Pattern Matching, 3. AIML, 4. Chat script, 5. Ontologies, 6. Markov Chain Model, 7. Artificial Neural Networks Models.
Stal 2020; Netherlands; literature review [219]	General Health	NA	The framework outlines the following main design features: (1) Speech/text, facial & gaze features, hand and body gestures (emotion, relational/empathetic behavior, rap, etc.) (2) Looks (human-like vs cartoon), clothing, body shape, gender, age, cultural tailoring, role - friend/professional, etc.)
Zhang 2020; USA; literature review [220]	Promoting physical activity and a healthy diet	The AI Chatbot Behavior Change Model	(1) Designing chatbot characteristics and understanding user background; (2) Building relational capacity; (3) Building persuasive conversational capacity; and (4) Evaluating mechanisms and outcomes.
Zierau 2020; Switzerland; systematic review [221]	NA (not restricted to healthcare)	NA	The framework outlines five major variables to consider: (1) rapport (social bond), (2) social presence (perceived humanness), (3) trust (in the AI agent), and (4) utility (usefulness) as well as (5) behavioral intention and outcomes (behavior change/interaction behaviors). <u>Additional variables:</u> auditory, visual, verbal, interaction (chat/voice interface), invisible measures (eg. timing, response times, intelligence, personality, tactile sensations eg. vibration).

5.3.2 Literature review on mobile phone delivered, rule-based conversational agent development

The original scoping review yielded 11,401 search results of which 47 studies were included [42]. Of the 47, ten fit the inclusion criteria for this conceptual framework, and the updated search retrieved 7 additional articles. In total, 17 papers discussing CA design and development were included here (Figure 5.3). A list of excluded studies and their reasons for exclusion are presented in Appendix 5.3.

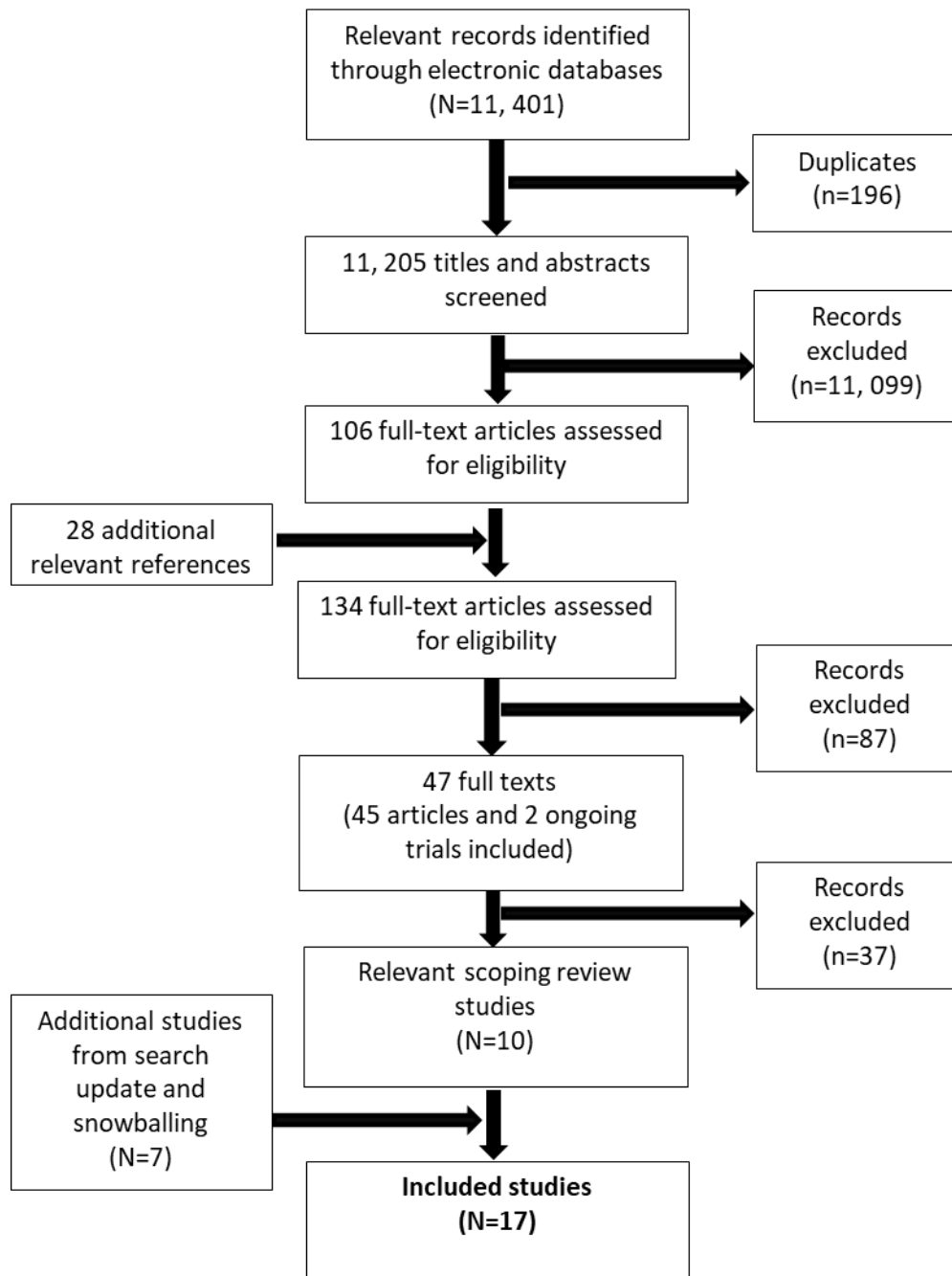


Figure 5.3 PRISMA flowchart for literature on rule-based healthcare CA development

Sixteen studies reported on rule-based agents only [128, 131, 133, 135, 149, 152, 156, 223-231](conversation restricted to multiple choice options or one/two-word answers by user), and one had a combination of rule-based and AI modalities (n=17) [232]. Methods of communication included text alone (including emoticons) (n=8) [131, 149, 152, 223, 228-230, 232], text with visual aids like images, videos or graphics (n=6) [128, 156, 224-227, 231] or text with sensor integration to monitor physical activity (n=2) [133, 135]. These healthcare areas were addressed by the CAs: smoking cessation [156], healthy eating [131, 230], asthma [228], obesity therapy [128, 133, 135, 229, 231], mental health treatment [149, 152, 225-227] or administrative tasks such as taking a patient history [232], and personal health assistance for self-management of chronic conditions [223, 224]. The studies were all conducted in high income countries, particularly Europe – Switzerland [128, 131, 133, 135, 227-229, 231, 232], Sweden [149], Spain [224], Italy [225, 230], and France [223], USA [226], and Asia - Hong Kong [156] and Japan [152]. The included study designs were randomized trials/randomized controlled trials (RCTs) [128, 133, 135, 149, 152, 156, 224, 227, 230, 231], pilot/feasibility studies [225, 226, 228], case studies [131, 232], a controlled before and after [229] and prospective study [223] (Table 5.3).

Table 5.3 Characteristics of the studies on rule-based conversational agents

Study ID Country of origin Study design	CA intervention; Evaluation	Total number of participants and duration of the study	CA personality	Delivery channel; Comm channels	Engagement features	CA Development (platform, coding, development team, coding)
Carfora 2019; Italy; RCT [230]	Healthy eating, aimed at reducing red and processed meat consumption.	N=180, 2 weeks	Not mentioned	Facebook Messenger; Text	None	Not mentioned
Casas 2018; Switzerland ; Pilot study [131]	Healthy eating, aimed at decreasing meat or increasing fruit and vegetable consumption.	n=36, 4 weeks	Friendly, positive, and stimulating	Facebook Messenger; Text, emoticons, images, videos, etc.	Food diary for goal-setting; Notifications; Weekly summaries	Development platform: Chatfuel
Chaix 2019; France; Prospective observational study [223]	Personal health assistant that educates patients by answering questions on breast cancer .	N=958, 1 year	Not mentioned	Facebook Messenger; Text	Regular reminders (medication adherence); Sending a question of the day.	Development platform: Wefight Inc.
Denecke 2018; Switzerland ; Case study [232]	Self-Anamnesis; Taking a patient history.	N=22, 30-45 mins	Human-like, funny, friendly	Standalone app; Text	Use of motivational statements by the chatbot and opportunity for users to ask queries.	Sources of information: a literature review. Coding: AIML.
Echeazarra 2021; Spain; RCT [224]	Improve BP and self-management behaviors;	N=112, 2 years	Not mentioned	Telegram messaging app; Text, image, video	Reminder prompt to use the tensiometer and take BP readings; and reminder for the next medical appointment.	Sources of information: a review of mobile-based solutions. Coding: Telegram chatbot development API.
Gabrielli 2020; Italy; Pilot study [225]	Mental health; Life skills training and well-being promotion among adolescents using positive technology approach to trigger self-reflection.	N=21, 4 weeks	Like a nice, smart old friend, trustworthy with important issues. Occasionally missing empathy.	Website; Text, video	NA	Sources of information: co-design (to collect the needs and preferences from 20 secondary school students) and positive technology approach which employs technology to elevate personal experiences (eg. emotional quality, engagement, and connectedness).

Greer 2019; USA; Pilot RCT [226]	Mental health; Consider the psychosocial needs of young individuals treated for cancer. Techniques used: stress and coping theory and the broaden-and-build theory of positive emotion.	N=45, 4 weeks	Non-judgmental, offered space for the user to “vent”	Facebook Messenger; Text + video	Group support	Sources of information: interviews and focus groups with patients treated for cancer (adolescents & young adults) for refined content.
Hauser-Ulrich 2020; Switzerland ; RCT [227]	Mental health; To promote Self-Management of Chronic Pain using Cognitive Behavioural Therapy (CBT).	N=102, 2 months	Warm greetings, empathic, a sense of humor. Use of emojis.	Standalone app; Text, audio, video clips	Personalized text messages are sent by CA SELMA daily or once in two days. Every message was also a notification that served as a reminder on the dashboard.	Coding: Mobile coach - an open-source software platform for the design and evaluation of mobile TBHCs.
Heldt 2018; Switzerland ; RCT [133]	Obesity therapy; Help overweight adolescents lose weight.	N=22, 5.5 months	Culture-specific	Standalone app; Text, images, sensor (physical activity measurement)	Daily encouragement Patients could earn virtual rewards in case of 24 weeks intensive phase of intervention.	Coding: Mobile coach - an open-source software platform for the design and evaluation of mobile TBHCs.
Kamita 2019; Japan; RCT [152]	Mental Health (Counselling method); Boost user motivation and stress reduction using the SAT (structured association technique) method for emotion stabilization therapy.	N=27, 85 mins	Not mentioned	LINE messaging platform; Text	Frequent stimulation of smiley face images to improve mental health and the practice of course repeating was important.	Development platform: a commercial HEROKU (Heroku, Inc.) server. Coding: Messaging API program provided by LINE for servers' linkage.
Kowatsch 2017 (a); Switzerland ; RCT [128]	Obesity therapy; Implement Obesity therapy for children.	N=11, 1 year	Friendly. Gender-neutral, empathetic.	Standalone app; Text, images	Daily challenges. Linguistic and visual TBHC to create interpersonal closeness which is positively related to the attachment bond between patient and conversational agent.	Development team: computer scientists, physicians, a psychotherapist, diet and sports experts for a technology-supported intervention targeting childhood obesity. Coding: Mobile coach - an open-source software platform for the design and evaluation of mobile TBHCs.

Kowatsch 2017 (b); Switzerland ; Controlled before & after study [229]	Obesity therapy; For obesity therapy in adolescents.	N=14, 2 months	Friendly. Gender-neutral, empathetic.	Standalone app; Text, emoticons	Various chat-based photos, physical activity, and quiz interactions.	Coding: Mobile coach - an open-source software platform for the design and evaluation of mobile TBHCs. Development team: computer scientists.
Kowatsch 2021; Switzerland ; Feasibility study [228]	Asthma; To deliver an asthma intervention to children-parent teams per BCTs - action planning and plan social support/social change.	N=49, 60 days	Empathetic digital assistant. A linguistic style that evokes interpersonal closeness.	Standalone app; Text; on-demand option, a phone call, or face-to-face interaction when required or triggered by parents, the patient, or the health care professional.	Gamification was implemented with a point and lottery system. This impacted enjoyment and intention to use the app. App usage reminders were also triggered.	Sources of information: data on asthma management in children, research on information systems, technology acceptance, conversational agents, behavior change, and experiential learning theory. Coding: Mobile coach - an open-source software platform for the design and evaluation of mobile TBHCs. Development team: physicians, a psychotherapist, diet and sports experts.
L'Allemand 2018; Switzerland ; RCT [135]	Obesity therapy; Motivate participants to make lifestyle changes that promote relaxation and physical activity.	N=22, 4 months	Game character	Standalone app; Text, sensor integration for activity measurement	Daily encouragement. The mobile chat app had a friendly, game character. The gamified reward system, and integration in smartphones for easy accessibility	Coding: an open-source platform with a text-based healthcare chatbot (TBHC), validated in public health interventions.
Ly 2017; Sweden; RCT [149]	Mental Health; Increase levels of subjective, social, and psychological well-being, and decrease levels of perceived stress using positive psychology and CBT.	N=28, 2 weeks	Empathetic	Mobile app Text	Empathic responses, daily check-ins, tailored content, summaries	Not mentioned
Stasinaki 2021; Switzerland ; RCT study [231]	Obesity therapy; Provide obesity therapy to adolescents using daily conversational agent counseling and relaxation exercises.	N=41, 5.5 months	Not mentioned	Mobile app Text-based with chatbot and direct verbal communication with HCP	Gamification (virtual rewards and competition with peers), daily encouragement by the chatbot.	Not mentioned

Wang 2018; Hong Kong; RCT [156]	<i>Smoking cessation;</i> using information and tips as well as social support through WeChat groups.	N=401, 2 months	Not mentioned	WeChat Text, images, video	Announcements of the group rules, Reminders.	Coding and development platform: WeChat's public Web API services.

The studies focused on defining the goal, creating an identity, and developing the content (Table 5.3). In several studies, conversational agents were delivered through well-known messaging apps like Facebook Messenger, Telegram, WeChat, and LINE. Engagement strategies included notifications, reminders, personalized information, gamified elements, and daily check-ins to keep the users regularly in touch and prevent dropouts. The studies did not refer to any existing CA frameworks for their development but did mention behavior change and relevant techniques that were implemented based on the healthcare topic and application [227]. Conversational agent personality was considered in most studies, and this was mostly human-like, friendly, and/or culture-specific [131, 133]. In some studies, information on the CA development process was not mentioned but the final framework was presented in the paper from which valuable inferences could be made and thus, were included [149, 230, 231].

5.3.3 Framework for rule-based conversational agents in healthcare

Following the analysis of existing frameworks, for mHealth and CAs and development information in the published studies evaluating mobile phone-delivered, rule-based CAs, five steps were derived for healthcare CA development. The five sequential steps include: (1) Defining the goal, (2) Creating an identity, (3) Developing the content, (4) Building the conversation, (5) Piloting & Evaluation. Co-design and privacy & security are two additional concepts presented as fluid components, spanning all five phases (Figure 5.4, Appendix 5.4).

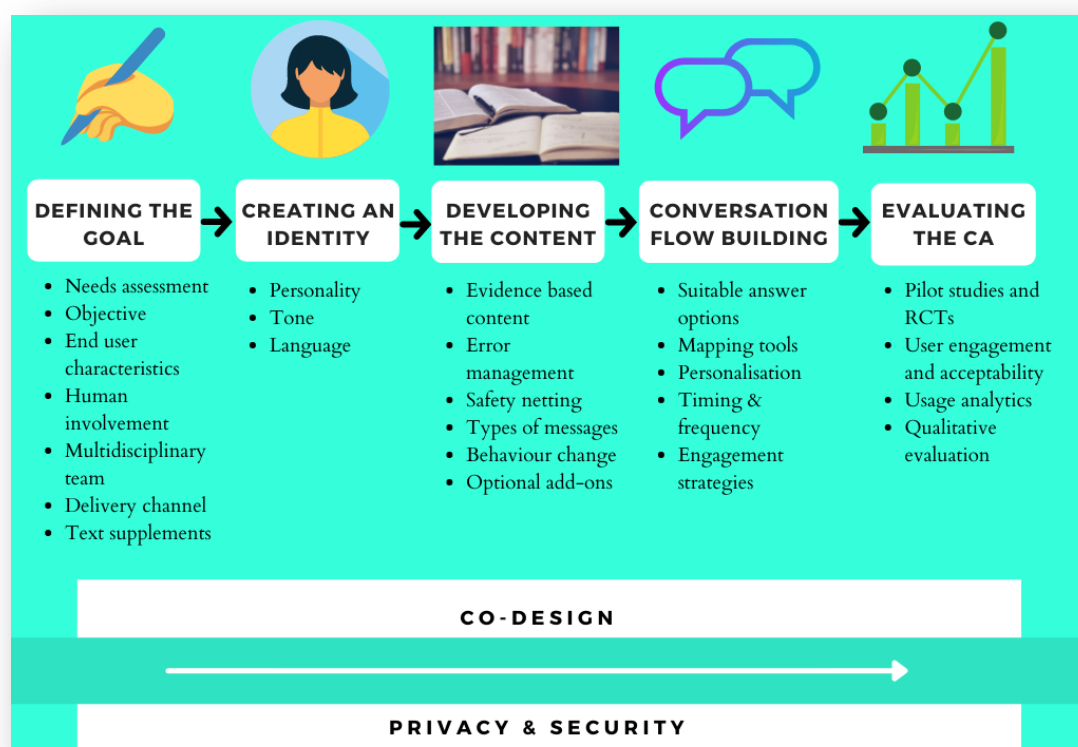


Figure 5.4 A five-step conceptual framework for rule-based healthcare CA development with two overarching themes.

(1) Defining the goal

Conversational agent development should begin with a clear definition of the agent's goals. This can be general or specific. The goal determines the primary and secondary outcomes which need to be tested and reported. Thirty-five papers described "defining the goal" as the first step of mHealth or conversational agent development [110, 111, 128, 149, 152, 156, 193, 195, 196, 199-214, 216, 217, 220, 222, 224-226, 228, 229, 232].

i. Needs assessment

Several mHealth frameworks emphasized the importance of conducting a needs assessment before starting the design and development process. This could help researchers to understand existing gaps which could be filled by the novel CA, such as the needs and challenges faced by patients and caregivers [110, 111, 200, 205, 208, 211, 214, 216, 217]. The needs assessment could be conducted as a literature review [216], scoping review [200], interviews, online surveys, or focus group discussions [226].

ii. The objective

The goal set must be clear, singular, and refined. A further point to consider when deciding on the topic of focus is whether the goal to be achieved is short-term or long-term. A short-term goal would be something that is achieved as soon as the interaction with the conversational agent is complete, such as a medication reminder [223]. A long-term goal would involve multiple interactions between conversational agent and user over a (fixed) period to eventually succeed in the goal that has been set, such as smoking cessation [156]. Some examples of objectives of previous CAs include prevention, diagnosis (e.g. symptom checker), decision support, self-management, education, lifestyle behavior change. For example, Kowatsch et al. employed prompt and reminder texts as these can be useful for instilling discipline and routine, which are both favorable traits for better self-management of health [228].

iii. The end-user

The next step involves defining the target population. An initial assessment should establish if the CA will be offered to healthy users, or individuals affected by a specific medical condition, as well as their gender and age group. Assessments should also include, smartphone penetration rate, digital and health literacy, cultural beliefs, social concerns, and perceived risk, particularly for highly stigmatizing conditions [233]. If the intervention is educational, a knowledge test should be implemented as this will inform the level of complexity the information should target [228]. Users can suggest the content features (e.g. what the app can do etc.) and design elements (visual aids etc.) they expect from the agent [225, 226]. Greer et al. conducted focus group discussions with potential end-users who revealed the type of text and videos regarding

psychosocial well-being adolescents treated for cancer (prospective users) would appreciate in a CA.

iv. Human involvement

If applicable to the CA's aim, it may be beneficial to consider a "hybrid" intervention where the interactions with the CA are complemented with regular interactions with a healthcare provider. In this way, the HCP can provide timely feedback on a self-management technique or provide regular support and motivation for the patient [128, 135, 228].

Alternatively, as presented in Stasinaki et al., instead of an in-person component, the conversational agent may be fitted with multiple channels, where the user can converse with the conversational agent in one and directly with an HCP in another [231].

A further point of human involvement can be via group support. This could take the form of online/physical peer support groups as an additional channel to the main chat channel with the CA. Adding a social or peer support component to the conversational agent intervention could positively impact participants' health outcomes [156, 234, 235].

v. Delivery channel

CAs can be delivered through a variety of channels such as a standalone smartphone application, or incorporated into existing conversational agent publishing platforms such as Facebook Messenger, Telegram, WeChat, WhatsApp, etc. Each target population has its usage patterns and preferences for messaging apps and/or standalone apps. For research teams comprising non-developers, plenty of online CA creation software exist such as Instabot or Collect.chat [236]. These can be employed to create a CA that is embedded in a website, for example. Alternatively, a CA could be delivered through a standalone app. This would involve hiring a programmer, collaborating with an app development company, or using an app builder such as "Appy Pie" or "App Sheet" [237]. The method for CA publishing would depend on the amount of funding and resources available, as expert manpower and any subscription fee charged by app builders will be an added expense [237].

The target population could help to determine the most suitable delivery channel and operating system (Android and/or Apple's iOS). In Kamita et al. the researchers chose the delivery channel "LINE" for their CA because it was the most popular social network service in Japan, where the conversational agent was to be implemented [152]. For a similar reason, WeChat was used in Wang et al. for target users in Hong Kong [156].

vi. Multidisciplinary team

Based on the objectives of the intervention, members with the necessary expertise need to be recruited into the team. For example, Kowatsch et al., who developed a CA targeting childhood obesity included computer scientists, physicians, a psychotherapist, diet, and sports experts [128]. The end users can also be consulted and would be part of the research team as in the example of a CA for pediatric asthma management where along with asthma experts and pediatric pneumologists, young asthma patients and their parents were also consulted on the type of asthma knowledge and behavioral skills that should be included in the conversational agent [228]. Whilst not all papers mentioned the members of their multidisciplinary teams, computer scientists and physicians were most frequently mentioned [128, 156, 228, 229]. Other allied health professionals like physiotherapists could also be included, to either contribute to CA content or to help with the recruitment of participants for CA testing [227].

vii. Text supplementations

In this conceptual framework, the text is the main mode of input and output considered for the CAs. The text for each message should fit the mobile screen without the need for scrolling [225]. Visual aids are also useful as demonstrated in Greer et al. where cancer patients reported on the types of text and videos which can help improve their psychosocial well-being [226]. Similarly, the participants in Denecke et al. who employed the self-anamnesis CA thought there was too much text on the page and would have preferred it if some videos or animations were included as well [232]. When using multimedia content, it is important to use high-resolution pictures to avoid pixelated or blurred images. This would lead to a more satisfactory user experience. Furthermore, as pictures obtained from the web abide by different copyright regulations and might require payment to be used, arrangements should be made to budget for the cost of the picture rights or the research team to take their pictures.

(2) Creating an identity

Identity creation is necessary to make the conversation more engaging and was discussed in three papers on CA frameworks and 10 papers describing CA interventions [128, 131, 133, 156, 219, 221, 222, 225-229, 232]. Creating an identity involves determining the conversational agent's tone of communication, the type of language they use, and other characteristics that will contribute to their personality.

i. Personality

The personality of the conversational agent should be designed to suit the target audience. Being able to convey a personality is important for a conversational agent as this enables the user to form a connection and make the interaction more personal and engaging [238, 239]. In healthcare, CAs often have a coach-like personality, particularly for motivational behavior change interventions, or adopt a healthcare professional-like persona, to come across as

legitimate and trustworthy [42]. For example, conversational agent “*Ruuh*”, has a female persona and adopts a very feminine tone of speech to connect authentically with women to discuss their health issues [240]. Empathy is also a personality trait highly regarded by users [227]. This shows the conversational agent’s ability to act with compassion, be human-like, caring, and interact, sans judgment [226]. The importance of empathy is reiterated by the fact that its absence in a CA, is quickly noticed and flagged by users as a limitation [225].

Enhancing the anthropomorphism of the conversational agent tends to increase user satisfaction as it gives the impression of conversing with another human being. Using the “typing” function to suggest the CA is preparing its next message is one way this can be achieved. Acceptability may be further enhanced if the CA acknowledges specific cultural or demographic traits of the target population or offers options to personalize the interface (e.g. offering a male and female persona) [133]. Finally, giving the conversational agent a name and profile picture can help with bonding and relationship building with the user [128]. On the other hand, conversational agents may explicitly disclose their identity [226], so the user doesn’t mistake them for an actual human, and this can help to reduce user expectations.

ii. Tone and Language

In line with the suggestion to be “culture-specific”, the conversational agent could be translated into the native language of the target population [156]. The language used for text-based interventions has generally been recommended to be encouraging, positive, friendly, polite, light-hearted, potentially with some light humor, nevertheless, formal so that the conversational agent maintains a respectable image [241]. It has been suggested that capital letters shouldn’t be used successively as this gives the impression of shouting at the user [241]. With regards to the format of the script, sociolinguist John McWhorter suggested that texting is best when communicated as “*fingered speech*” [242] whereby the text should be conveyed more in the form of casual speech as opposed to written language [242].

Emojis can sometimes communicate non-verbal expressions better than text [226]. However, emojis are vulnerable to varied interpretations across cultures and contexts, and so should be used mindfully. Their efficacy may also vary depending on the target group. Fadhil et al. noted that the context-specific nature of emojis increased efficacy in a mental health intervention but did not help promote physical well-being, where a text-only condition was better received [243].

There are some additional points to consider when discussing highly sensitive topics such as HIV/AIDS, STIs, or mental health disorders. For such applications, it may be worth employing some simple coding in case someone else close to the user inadvertently reads the message [244]. This may be more relevant in some low- and middle-income countries where family

members sometimes share a single smartphone between themselves. Reminders that usage of the conversational agent app or service was "*free*" and "*confidential*" were positively received by participants [244]. Also, the emphasis these words placed on how easy it is to reach the desired outcome tended to make the messages more attractive to the user.

Scientific jargon should be avoided and instead clear, short sentences with simple language should be employed to reach a diverse range of individuals with varying levels of literacy [245]. The National Institutes of Health recommend that patient education materials in web-based or paper form should be written at a reading level at or below the sixth grade (11 and 12 years old) [246]. The readability of the text can be assessed using a scale such as the Flesch Kincaid grade level [247]. Furthermore, whilst the conversation should resemble that which happens over text, simple but formal text has been indicated to be preferred over "textese" (a type of slang associated with texting whereby unconventional spelling and grammar is used e.g. "tonite" instead of "tonight"), abbreviations and contractions, such as "pls" or "wanna" for "please" and "want to," respectively [248].

(3) Developing the content

Twenty-six papers mentioned content development and what this entails [110, 111, 131, 156, 195, 200, 202-206, 210-213, 215, 216, 218, 220, 223-226, 228, 230, 231]. Content development involves determining the sources of information, mapping out user intents, and deciding on error management and safety netting strategies. In addition, different message types can be employed to fulfill different purposes. Finally, relevant theories such as behavior change techniques and persuasive strategies also feed into the process of developing the content [211, 220].

i. Information

The information provided in the healthcare CA must be evidence-based [249]. Sources of evidence-based information include comprehensive literature reviews, clinical practice guidelines, Cochrane database, and reputable organizations like the World Health Organization, USA's Centre of Disease Control (CDC), UK's National Health Service (NHS), etc. [250]. [217]. Kowatsch et al. for example used evidence from multiple sources for a CA on asthma management [228]. The sources included published literature on the improvement of asthma management in children [251], technology acceptance research [252], and information on user-conversational agents working alliance [253].

The collated evidence then needs to be converted into a conversational format conducive for a CA intervention and the target population. This would involve converting the scientific evidence

into a conversational format taking into account the target population's beliefs, determinants, and context [200].

ii. Managing errors

Additionally, to ensure that the conversation proceeds uneventfully, contingencies need to be put in place to address misinterpretations of user inputs. For example, the conversational agent could request clarification when receiving an unfamiliar input or direct the user to further support, such as an HCP or a human administrator [254, 255]. These strategies were included in TensioBot, an intervention to promote self-management of blood pressure (BP), where the CA asked for confirmation if the patient entered a BP value outside the normal range, and alerted the doctor if the abnormal BP measurement was not due to a typing error and indeed an accurate reading [224]. Other strategies to manage unintended errors include validated data entry fields, limiting the data input to a pre-determined number range, certain words, or characters.

iii. Safety netting

For high-risk applications, such as mental health disorders, or emergencies, a more robust management strategy needs to be implemented to protect users' safety. "*Wysa*," a popular mental health conversational agent includes a disclaimer stating that "*Wysa cannot and will not offer medical advice*" but instead can refer the user towards its team of mental health professionals [256]. Self-management apps should include clear messages to provide decision support to users aligned with the best evidence [257]. Additionally, suicide prevention apps should include emergency contact information or direct access to crisis helplines [97].

iv. Types of messages

Conversational agent messages differ according to their aims. For example, messages could be educational [230], motivational [131], or reminders [223]. In applications where the conversational agent is required to follow up on doctor's appointments, it would also be beneficial to send the patient a status report or summary a few days after a consultation [255]. This could act as a reminder to stick to a medication regime or revise educational content discussed during a therapy session with a psychiatrist.

Some coaching style messages might prioritize sympathy, empathy and emphasize the participants' achievements so far in the intervention [227]. Carfora et al. attempted to reduce users' consumption of red meat [230]. In addition to an informational message condition, they also had an emotional message condition where they received information on the damaging effects of meat consumption on the environment and on health, aiming to evoke a sense of regret in the user and subsequently reduce meat consumption [230]. Wang et al. outlined a CA which employed four types of messages for smoking cessation: group announcements, sharing

of health-related information, reminders to share positive results and progress as well as fixed answers to frequently asked questions or requests [156].

v. Behavior change theories

Conversational agents have been effective when used for healthcare behavior change [258]. In the literature collated for this framework, nine conversational agents described either a behavioral change theory or concept used in the design of the intervention. Two studies [152, 227] reported the use of behavioral change theories in the conversational agent design, including Health Action Process (HAPA), and the Technology Acceptance Model (TAM), while seven studies [128, 131, 133, 156, 228, 230] reported the use of specific BCTs including goal setting, anticipated regret, use of rewards, tracking and feedback, self-monitoring, and social support. [229] .

The information should be presented in a simple enough format that the advice is actionable. For example, in an intervention to help users improve their food lifestyle, a motivational and coaching approach was adopted to show users empathy and encourage them to change their eating behaviors, without any judgment, by reducing their meat consumption and increasing vegetable intake [131]. Another study used multi-component behavior-changing interventions such as goal-setting, self-monitoring, stimulus control, and behavioral contracting to support a healthy lifestyle behavior change in obese adolescents [231].

Furthermore, including group chats where peers can provide informational and emotional support for one another can also effectively promote positive behavior change such as smoking cessation [156].

vi. Optional add-ons

In some circumstances, integrating the CA with external devices may be appropriate like a glucometer, blood pressure cuff, or activity tracker [133, 135, 224]. Employing wearable sensors and tying these readings to the CA could allow for readings on physical activity or heart rate monitoring to be relayed in real-time to HCPs as used in a CA intervention to promote physical activity in obese adolescents [134]. Other types of sensors would be beneficial if used to monitor older people who live alone and are at risk of falling or have had multiple hospital admissions.

(4) Conversation flow building

A good CA is eloquent and knowledgeable and thus, requires a meticulously crafted script. Conversation flow building was discussed in 19 of the papers included in this framework [110, 111, 133, 149, 195, 200, 202, 205, 210-213, 217, 220, 221, 223, 227, 228, 231]. These considered generating multiple, relevant answer options, using tools to organize and build

conversations, making considerations for personalization of the CA for the user, using suitable message timings, and employing engagement strategies.

i. Providing suitable answer options

For good conversation flow, the pre-defined answer options should be sufficiently aligned with possible user intents, defined as responses the user may use in a typical conversation turn. Constructing a mind map outlining the different facets associated with a topic (e.g. medication adherence) and the likely influencers (lifestyle components or emotional state) on user responses would help to predict the most relevant answer options to present the user with [259].

ii. Selecting a mapping tool

Using mapping or organization tools can help to streamline the conversation flow-building process. Specific tools to consider include Recast.ai, Bot Builder, and Motion AI [260]. Conversation flows can also be built via other tools such as mind mapping software Xmind [261]. Mind mapping may assist in recording the flow of conversations between different topics or different user interactions. A well-constructed conversation flow leads the conversation, guides the user, and can address multiple relevant queries about its purpose. Furthermore, interactivity, personalization as well as consistent messaging were noted to be valued qualities [262].

iii. Personalizing content and delivery

As compared to generic CAs, context, situational, and/or individually-aware agents support a more positive user experience [249]. A simple example of a personalized CA is addressing the user by their name or nickname [263]. Taking personalization further, interventions should be tailored to individual participant needs [241], for example, in the case of promoting self-management of chronic pain, providing personalized content based on the user's duration and type of pain as opposed to general advice would be appreciated [227]. Notifications for unread messages and prompts to take medications or reminders for doctor's appointments add a degree of personalization to the CA too [223, 227].

iv. Selecting appropriate message timing and frequency

The timing and frequency of messages are important considerations in conversation delivery. The literature on text messaging interventions generally indicated a preference for weekly messaging [248]. But specific applications may require a higher or lower frequency. For example, for smoking cessation programs, a higher number of messages were delivered on the day of quitting and additional messages could be requested by the user when they had a craving or needed some motivation to prevent relapsing [264]. Similarly, for high-risk behaviors such as

binge drinking or inappropriate sexual behaviors, messages could be delivered promptly when the risk behavior is expected to occur eg. a Friday night [244, 263]. Furthermore, Finitis et al. observed that researchers should refrain from sending daily messages as the excessive messaging resulted in no reply from the participants [265].

Timing and frequency may also be determined by user preference during the co-design phase or included as programmable options in the app settings.

v. *Employing engagement strategies*

Six of the papers discussed strategies to keep the users engaged for the intended duration of the interaction and/or to get users to keep returning to the conversational agent [133, 149, 223, 227, 228, 231]. These engagement strategies were notifications, weekly summaries, reminders, motivational statements, persuasive techniques, high frequency of messages, daily encouragement (if suitable). Other strategies were directed towards the CA's personality. In some cases, the conversational agent was designed to build a rapport and attachment with the user [128, 228], or a comic profile and gamified component were included to introduce some competition or include the possibility for rewards and points [228, 231].

(5) Evaluation

Conversational agents should be evaluated in a stepwise, comprehensive, and methodologically robust manner. Below are several important aspects of CA evaluation to consider namely pilot studies and randomized controlled trials, user engagement and acceptability, usage analytics, and qualitative evaluation.

i. *Pilot studies and randomized controlled trials*

Once a prototype of the CA is prepared, it should first be tested in a pilot study to assess its functionality before testing it in an RCT with a large sample size [229]. CA evaluation was discussed in 20 papers [110, 111, 135, 152, 196, 201, 203-212, 214, 215, 220, 228]. Primary and secondary outcomes of interest should be defined in line with the CA's aim. Some examples of outcome measures include knowledge (for an educational intervention), skills (e.g. self-management), patient satisfaction, etc. Trial design and pre-and post-intervention assessment should be designed to suit these outcomes. Casas et al. aimed to cultivate healthier eating habits and this was implemented with short daily commitments, daily check-ins, and close monitoring of users to avoid premature abandonment of goals [131]. Greer et al. studied wellbeing outcomes whereby anxiety and depression symptoms were measured using the patient-reported outcomes measurement information system (PROMIS) initiative [226]. Similarly, Kamita et al. employed the SAT psychological scale to measure reductions in users' stress levels

following the CA intervention [152]. Denecke et al. also suggested assessing technical competency to understand the digital literacy and needs of the target population [232].

For effectiveness, assessments could target improvements in behaviors (e.g. medication adherence or depression), or specific changes in knowledge and skills directly relevant to the healthcare topic eg. BP checking best practices [224]. It is necessary to use validated outcome measurement tools for the relevant application such as the PHQ-9 (patient health questionnaire – 9) for depression, flourishing scale (FS) for wellbeing, perceived stress scale (PSS-10) for measures of global perceived stress [149].

Novel methods of evaluation can be employed such as micro-RCTs, where randomization happens every time that intervention component is delivered. This method has been implemented for CA interventions to promote physical activity [266]. Another relatively niche method is ecological momentary assessments (EMAs) which study an individual's behavior (using surveys, for example) at a specific moment in their natural environment at regular intervals [267]. These have been used in mHealth interventions to reduce alcohol intake in adolescents and have the advantage of overcoming recall bias which can happen in conventional methods of surveying [267]. EMAs have also been implemented to assess stress, hunger, cravings, and stress eating behaviors in an intervention delivering mindfulness training to type 2 diabetes patients [268].

ii. User engagement and acceptability

Engagement with conversational agents can be measured using an open app ratio (ie. how long the user keeps that app open) [149]. Acceptability can be assessed by the average duration spent on the app/conversational agent and how many active sessions were completed [226]. Participants can also be questioned about their willingness to recommend the CA to friends and family as this can be a good indicator of satisfaction and acceptability [226].

Other items indicating acceptability include interest in specific conversational agent domains or themes or narrative feedback. Chaix et al. measured usage duration, interest in various educational content, and level of interactivity as indicators of acceptability [223]. Kowatsch et al. studied acceptance using a 7-point Likert scale (strongly agree-strongly disagree) for perceived usefulness, ease of use, enjoyment, and usage intention [228]. Similarly, Denecke et al. used a usability test for performance, humanity, affect, and acceptability to outline the usefulness and appropriateness of the CA [232]. Echeazarra et al. employed a survey with questions on ease of use, preference for the CA over existing methods if the CA was helpful for its intended purpose, and whether the user had ceased usage as measures of acceptability and satisfaction [224].

iii. Usage analytics

Conversational agent developers would need to analyze and track the CA's usage to see which aspects of it are under-utilized, missing, or not working as anticipated [260]. Such usage analytics are often embedded on host platforms. For example, *Manychat* offers some built-in analytics tools such as the number of people who opened the chats, the number of times the conversational agent was accessed, the duration of participants' interaction, "*click-throughs*" to continue in conversation, [269] or how quickly users responded to CA inputs, which could indicate the user's level of interest. Other measures to consider including: duration of interactions, free text typed by users (which sometimes may be questions), number of complete, incomplete, or absent interactions [198]. Third-party analytic tools that can easily be integrated into platforms like Facebook Messenger, Kik, and even Alexa include Dashbot.io and Botanalytics [270-272].

iv. Qualitative evaluation

User-specific feedback can also be obtained through qualitative methods such as free responses in questionnaires or interviews. Yan et al. described a focus group discussion as part of the evaluation process for an mHealth intervention to promote physical activity [273]. In the discussion, each text message by the CA was displayed and participants were required to respond either with "*Yes, I like it*" or "*No, let's change it to make it better.*" This voting was then followed by a conversation where sub-optimal messages were improved, and the strengths of effective messages were noted. Gabrielli et al. facilitated a participatory design workshop where suggestions for improvement were provided via open-ended questions [225]. In Ly et al. semi-structured interviews were conducted whereby participants commented on the benefits, opportunities, and challenges associated with the CA for mental health [149]. Some examples of the qualitative feedback were where participants commented on the content (some liked the learning part whilst for others, the reflection evoked by the content was more effective), constant availability of the CA, and technical functionalities of the CA [149].

The necessity for thorough evaluation was accentuated by Laranjo et al. [41] who noted that user experiences and technical measures were inconsistently reported and that only one study in their systematic review employed robust methodology (RCT) and validated questionnaires to assess CA efficacy.

(6) Overarching themes

i. Co-design

Co-design in digital interventions for healthcare can be defined as the collaboration between the developers (i.e. researchers or healthcare and technical staff) and the users of the intervention (patients, caregivers, HCPs, stakeholders) at any stage in the development process. The aim is to take full advantage of each group's expertise to procure the best outcome for the intervention. Including this element as a floating member of the conceptual framework may have a positive impact on health and health-related behaviors as the intervention should emerge to be beneficial and desirable to the user [234]. Obtaining the views and opinions of end-users early in the development is also constructive in designing culturally sensitive content for the intervention [234].

At the stage of defining the goal, conducting a needs assessment to figure out what they need could help direct the intervention in the right direction. When creating an identity, interactions with the target population could help inform the kind of agent they want to interact with; whether that's friendly, professional, funny, serious, culturally specific, male/female, etc. In the process of converting evidence-based information into conversational content, having participants' input on their tone of speech and language used could help inform the types of terms, phrases, or slang that they can most relate to. For example, in Greer et al. they conducted interviews and focus groups with young adults treated for cancer to refine informational content for a conversational agent format based on their preferences, vernacular, and shared experiences [226].

Co-design is also relevant at the stage of piloting and evaluating. Usability testing may, for example, employ thinking out loud testing. Thinking out loud testing is an in-person testing method where the participant is monitored as they use the conversational agent, and they verbally express their thoughts and opinions of the agent. [274].

This was cited in several mHealth frameworks, emphasizing the necessity for user-centered design in several applications [110, 111, 199, 207].

ii. Privacy and security

It is necessary to ensure the privacy and security of participant data. User authentication through a login component is a good start [275]. Having a username and password or through biometric authentication prior to having access to the conversation exchange heightens privacy [275].

A framework for governing responsible use of conversational agents in healthcare posited that all the *"data and the history of interactions including intended and unintended revelations of private data"* must be securely safeguarded and appropriately disposed of in line with privacy and data protection laws [276]. Users of the conversational agents maintain the right and access

to all the personally identifiable information exchanged in the conversations. Furthermore, the data collected must not be used as a means of surveillance or to discriminate users against healthcare privileges or coverages [276].

It can also be helpful to minimize the amount of personal and sensitive information collected from users. This can be achieved by asking specific questions, so users do not end up over-sharing or simply providing predetermined options for users to choose from to minimize the collection of unnecessary personal information.

All data must be encrypted during transit (when the message is being sent) and at rest (when the message has been delivered) [277]. Denecke et al. discuss this issue where some healthcare conversational agents run on social media platforms such as Facebook messenger where the collected data is tied up in data policies that are lengthy and difficult to understand and assess [232]. The user resultantly has no ownership over the data which means it could be sold, traded, or marketed by the distributor of the conversational agent or social media platform. Hence, they recommend a data security feature in a health conversational design framework [232].

(a) Compliance with data privacy laws

When designing healthcare conversational agents which collect sensitive patient data, there is a need to ensure its compliance with country relevant data privacy laws, such as the Health Insurance Portability and Accountability Act (HIPAA) in the US, [277] or the General Data Protection Regulation (GDPR) in the EU [278]. This means that information must be directly delivered from patient to provider and vice versa. Privacy assessment remains to be an issue with mHealth applications, as 600 of the most-used health-related apps don't currently include a privacy policy [279]. The GDPR was updated in 2018 to instill further stringent measures on such data privacy, also in the interest of protecting research participants [278]. The law considers anonymization, consent, and data sharing and makes clear that even pseudonymized data still counts as personal data [278]. Individual countries have specific requirements and adopt their own data protection laws and regulations. In Singapore, for example, the Personal Data Protection Act (PDPA), is a baseline regulatory framework informing the collection, distribution, and usage of personal data [280]. PDPA compliant telehealth tools have been developed and tested on a trial basis. One such example is an app for the education of medical students on clinical case discussions [281].

5.3.4 Validation study

The conceptual framework was used to develop a CA for healthy lifestyle behavior change and to provide educational information on diabetes and prediabetes (chapter 7). The design and development processes were integrated into each of the five steps outlined in the conceptual framework. Formative research was done in the form of a needs assessment comprising semi-

structured interviews (chapter 6) and a subsequent pilot study to test the feasibility and acceptability of a conversational agent to deliver advice on healthy lifestyle change to the general public in Singapore (chapter 7) [198, 282].

Some preliminary research revealed that Singapore has the second-highest prevalence of diabetes amongst developed countries [54]. Furthermore, a needs assessment served as co-design in the early stages of CA design and development by highlighting the barriers and facilitators to healthy living in Singapore. This informed the goal of our conversational agent which was to educate the general public on topics relevant to a healthy lifestyle change and diabetes prevention. The process of developing and testing the CA is elaborated on in Chapters 6 and 7.

Considering the avid use of messaging apps by Singaporeans and the ease of accessibility to Facebook Messenger (2nd most frequently used messaging app), this was chosen as an appropriate delivery channel [283]. Information about diet, exercise, sleep, and stress were included. Text was the main mode of input and output, along with free-to-use stock images and pictures taken by the research team to demonstrate food items or exercises and sometimes links to videos for physical activity and stress management exercises. Our research team comprised a faculty member with medical background and a Ph.D. in public health, a research associate with a Master's in public health, and a Ph.D. student with a degree in Biomedical sciences. We also consulted a post-doctoral research fellow in digital health with a medical background and other colleagues with expertise in digital health.

The personality of the CA was carefully crafted. As it was to deliver healthcare-related advice, the tone was formal, but friendly so as not to come across as daunting to the user. The CA was given a name (*Precilla*) and a profile picture to enhance anthropomorphism. The language used was to be easily understood by the layman and fully in English.

We conducted a pilot feasibility trial including 75 participants who used the conversational agent for 4 weeks [198]. The pilot study is described in more detail in Chapter 7.

5.4 Discussion

5.4.1 Key findings

This chapter presents a new conceptual framework for the development and assessment of mobile phone-delivered, rule-based CAs. The framework comprises 5 steps: (1) Defining the goal, (2) Creating an identity, (3) Developing the content, (4) Building the conversation, (5) Piloting & Evaluation. Co-design and privacy & security are overarching concepts, relevant at each stage of development.

This framework was based on the analysis of 23 mHealth frameworks (general mHealth, smartphone apps, text messaging), 6 CA frameworks, 17 primary studies reporting on the design and development of rule-based healthcare CAs. A pilot study was also conducted to validate the framework. The existing mHealth frameworks provided general guidelines on the development of mHealth interventions for healthcare. Conversely, the CA frameworks provided more granular advice about AI-based and non-healthcare-focused CAs. This framework combined the relevant suggestions to present a guide for the step-by-step development of healthcare, rule-based, text-based conversational agents.

Considering the multifaceted nature of ECAs and the exploratory nature of the field, this conceptual framework only addressed non-embodied conversational agents. In other words, those which used text as the main mode of communication, without an avatar.

5.4.2 Rule-based and AI CAs

Research has indicated that some of the common reasons behind the failure of mHealth and health apps are a lack of evidence-based, transparent, and step wise approaches in their design [13]. Many health interventions are driven by “silver bulletism” which is the tendency to assume that the newest forms of technology can contribute to panacea-like breakthroughs [284].

Contrastingly, in reality, some of the most successful mHealth applications employ mobile phones for simple purposes such as text messaging or voice calls for smoking cessation reiterating that healthcare researchers developing complex interventions with advanced AI may be over-engineering for the intended purpose [13-15]. Rule-based CAs can address some of the limitations of text messaging-based interventions (cost, user preferences, and usage as well as streamlined and more functionalities such as bidirectional communication) whilst also avoiding the severe challenges associated with AI technology for healthcare.

5.4.3 Comparisons with existing literature

Existing mHealth frameworks provide systematic suggestions on mHealth intervention development, and this process often commences with a review of existing literature or formative research to understand the needs and the profile of the end-users [218].

Furthermore, there is also frequently a need for patient and public involvement to make the intervention as relevant to the target population as possible [214, 216]. The mHealth frameworks also described the importance of conducting preliminary evaluations to identify limitations before testing the mHealth intervention in a larger scale trial [205, 214, 220]. CAs are a niche type of mHealth intervention which requires their own framework to guide design and development. Hence, this conceptual framework includes the above-mentioned items as well as some additional elements directly relevant and necessary for CAs such as personality development, converting evidence-based content into conversations, and suggestions on up-to-date methods of evaluating CA prototypes.

In comparison to the points proposed in Diederich's taxonomy (a framework for the design of non-healthcare-based CAs), technical aspects are not emphasized in this framework [222]. Payment options and "Pricing models" were also excluded as payment models may be very different in various settings and for different CA applications.

Hussain et al.'s framework was focused on the technical development aspects of CAs [218]. This conceptual framework instead homes in on the development of the content for the CA and how to follow a step-by-step workflow that is evidence-based and directed towards suitable healthcare-related outcomes [218].

Zhang et al. looked at a model for AI conversational agents for physical activity [220]. They considered some relational components which were also relevant and similar to this framework such as establishing chatbot personality, empathy, and disclosing its identity. Some of the evaluation measures were also adopted in this conceptual framework e.g. user experience, usage patterns (login times, length of usage episodes), conversational quality (e.g. fluency), user perception of relational capacity (likeability, empathy).

Finally, Zierau et al.'s framework which stemmed from a review of the literature on health and non-health-based conversational agents shared some similarities and differences with our framework [221]. Rapport and trust were not topics of great focus because they are not entirely relevant or achievable with rule-based agents. Lack of AI leads to a lack of involved discussion and hence, not much opportunity/chit-chat to establish rapport. As opposed to chit-chat agents, rule-based CAs are more goal-oriented and can direct the user through a series of conversations to acquire knowledge or achieve a specific outcome.

5.4.4 Implications for future research

This conceptual framework is grounded in advice and recommendations from existing research and draws evidence from a broad range of areas ranging from clinical trials to systematic reviews and empirical research. This makes it beneficial for physicians, healthcare app designers, young

researchers etc. It's first implication would be for healthcare researchers to apply the chronological recommendations to build a similar healthcare CA which is relevant, necessary and likely to be accepted by the target population in a time and cost-efficient manner. Furthermore, no conceptual framework is static and definitive. The understanding of the basis behind the framework, as meticulously outlined in this chapter, could lead to further refinement of this framework for its practical use. Thus, this conceptual framework provides a basis for meaningful future research, as other mHealth conceptual frameworks have, for example, for augmented reality apps in healthcare [285].

Future research should consider the application of this framework for CA interventions in different healthcare specialties, settings (hospital/out-patient), target groups, and cultures. It remains unclear which behavior change theories and techniques are best suited for CAs. Future work on this can explore and ascertain which BCTs best suit which CA interventions.

Whilst concepts of identity creation, conversation flow development, and delivery are important, the preferences of each facet for different target populations are still unknown and need to be crystallized so interventions can be tailored accordingly. Additionally, more research on the evaluation of healthcare CAs can help to inform the ideal data analytics required. Information on how to best assess conversational agent interventions is still lacking. Finally, although this framework is comprehensive and many components apply to AI CAs, a separate framework is needed to integrate AI-specific concepts such as natural language processing, voice vs text parsing, etc.

5.4.5 Strengths and limitations

This is, to our knowledge, the first attempt at creating a conceptual framework outlining the components which need to be considered when designing a healthcare-specific rule-based CA. This conceptual framework is both comprehensive and simplified in its descriptions to cater to healthcare researchers without expertise in computer science and no previous experience working with CAs. This framework builds on an analysis of existing mHealth frameworks and is also based on a stringent and up-to-date search and analysis of rule-based CA literature. The validity and applicability of the framework have also been demonstrated by the study team in their development and testing of a rule-based CA for healthy lifestyle changes and diabetes prevention.

A large portion of the information provided is anecdotal or derived from research conducted into text messaging and other mHealth interventions because of the scarcity of research published on rule-based CA development for healthcare. This framework provides a high-level overview of the main steps required in a CA development framework and most steps can and should be explored

in more depth. Furthermore, this framework is restricted to rule-based, non-embodied CAs and does guide the development of AI CAs. Additionally, economic, social, and behavioral differences in different populations may limit the generalizability of this framework.

5.4.6 Conclusions

The interest and potential for conversational agents in healthcare are growing but guidelines on how to design, develop and implement these interventions are currently lacking. Drawing on published evidence, this conceptual framework provides a consolidated summary of how healthcare researchers can create their own CA. The process is divided into five essential steps, with two overarching themes. Future research should explore the applicability of various behavior change theories, privacy and safety concerns as well as the relevance of this framework in other healthcare areas and geographical settings.

Chapter 6 – Public’s views and needs regarding Diabetes, Healthy Living and Conversational Agents in Singapore: Needs Assessment

This chapter addresses objective 3 of this thesis, to “understand the public’s perceptions and needs with regards to conversational agents for healthy living and diabetes prevention in Singapore, using semi-structured interviews and subsequent qualitative analysis”. Section 6.1 introduces the project with some background information on the applicability of diabetes and prediabetes and healthy lifestyle changes and digital health solutions in Singapore. The methods for the needs assessment interviews are elaborated on in Section 6.2. The results of the qualitative analysis are described in Section 6.3 where the participant demographics are disclosed along with the specifics of the thematic analysis. In the discussion (Section 6.4) the key findings are presented along with comparisons of these findings with existing literature. The implications of this qualitative analysis on future work are discussed together with the strengths and limitations of this study (Sections 6.4.3, 6.4.4).

6.1 Introduction

The growing burden of diabetes is a matter of global concern, especially in Singapore. Singapore’s life expectancy increased from 83.2 in 2010 to 84.8 in 2017; however, when adjusted for time in perfect health (health-adjusted life expectancy), Singapore’s life expectancy dropped to 74.2 [286]. The main contributors to this drop are chronic conditions such as diabetes, hypertension, and high cholesterol levels, which are largely due to unhealthy lifestyles and habits [286]. Prediabetes, a precursor of type 2 diabetes is also becoming increasingly common. Fortunately, lifestyle interventions, including promotion of physical activity and a healthy diet, delivered by trained health professionals can reduce the incidence of type 2 diabetes in those with prediabetes [60, 62, 287].

CAs can be employed to deliver such lifestyle interventions and will be beneficial in reaching large groups of at-risk individuals especially in technologically astute settings such as Singapore given the ubiquity of Wi-Fi and smartphones. Singapore’s Ministry of Health has highlighted the increasing use of conversational agents in health care to tackle issues such as the rising chronic disease burden within an aging population [67]. This situation makes Singapore an ideal candidate for the implementation and testing of novel mHealth interventions, such as conversational agents [288]. CAs can help to reduce the burden on healthcare professionals whilst maintaining integrity in their

delivery of healthcare support through personalization, efficiency, bidirectional interactivity, and a chance to build up a working alliance, highlighting the potential to improve patient care [253].

Thus far, they have been employed in health care service provision, chronic disease management, patient education and can be delivered via existing messaging apps, individual apps, or even standalone devices [42, 289]. People's knowledge of diabetes, prediabetes and the role and impact of healthy living can affect their lifestyle choices and ultimately health and this can be nurtured through a CA intervention [290]. However, before introducing a novel method of digital health delivery to a study population, a needs assessment study is required especially in this case where the evidence on conversational agent use for healthy lifestyle promotion in different settings, including Singapore, is currently limited. A needs assessment will identify the challenges faced by the population, their knowledge gaps, the essential components they wish to see and gauge their likely acceptability of a new digital initiative for their health [291].

6.1.1 Objectives

To this end, this study aimed to conduct a needs assessment to support the design and evaluation of a CA to promote healthy lifestyle behavior changes for the general population in Singapore. To inform the development of such a conversational agent, members of the public were invited to share their experiences, knowledge, and opinions on diabetes, prediabetes, healthy living, digital health interventions, and CAs which are reported in this study.

6.2 Methods

This qualitative study was conducted between August and September 2019. 20 volunteers from the public participated in the semi-structured interviews. The study was approved by the Nanyang Technological University Ethics Committee (IRB-2018-11-032). All participants read the study information sheet and provided informed consent remotely via a digital form. Consent was also obtained to record the interviews. This study followed the consolidated criteria for reporting qualitative research (COREQ) guidelines [292] (Appendix 6.1).

6.2.1 Participants and Recruitment

This study was part of a larger trial (Chapter 7) testing the feasibility of a healthy lifestyle change CA where participants were recruited using a study poster on relevant healthy living Facebook (Facebook, Inc) pages and groups. A purposive sampling strategy was used to capture variations in ethnicity and age groups from a wider range of perspectives. Participants taking part in this study provided consent, completed a series of questionnaires: eligibility, baseline, and follow-up

questionnaires, and conversed with a Facebook Messenger CA over the 4-week study period (Figure 6.1). Included individuals were adults aged ≥ 21 years who were fluent in English with a Facebook account. The following exclusion criteria were applied:

- individuals with a history of major illness, such as cancer, heart disease, stroke, chronic liver disease, chronic kidney disease, neurodegenerative condition, and hypertension
- individuals with a physical disability that would prevent regular physical activity
- pregnant women
- participants younger than 21 years of age, illiterate or nonwriting individuals (as all questions were asked in English)



Figure 6.1 The order of events participants underwent, from (1) recruitment, (2) interviews for needs assessment, and subsequent participation in a 4-week feasibility study.

Individuals with a history of major illness, those with physical disabilities, or pregnant women were excluded as advice provided by the CA may not have been appropriate for these vulnerable populations especially with regards to their dietary and physical activity requirements. Additionally, fluency in English was needed as the content in questionnaires, informed consent, interviews, and conversational agents were all in English. As such, illiterate, and nonwriting individuals were also excluded.

6.2.2 Data Collection

Interviews were conducted by a female Ph.D. student (DAD) in private meeting rooms at the Lee Kong Chian School of Medicine, NTU, Singapore. An interview guide was used, and interviews were conducted either in person or over the telephone, depending on the volunteer's preference. DAD was equipped with sufficient details, resources, and exposure to web-based courses on qualitative research and the proper conduct of telephone interviews before study commencement. Field notes were taken during the interviews. The initial interview guide was informed by the literature on the development of digital and conversational agent-delivered health interventions (Appendix 6.2) [293-295]. The interview was adapted for optimization after the first few interviews, to better

accommodate emerging themes, and each interview ended when the saturation of novel topics (eg, for barriers and facilitators, respectively) was reached.

Basic demographic data were obtained from all participants, including gender, age, ethnicity, marital status, monthly household income, number and age of children, and occupation. Their previous experiences with CAs and feedback on these interactions and suggestions for future health applications (eg, tone, media, and direction) were also explored. These insights were collected to direct the future design and development of CAs for this population. Participants were also asked to share the types of messaging and health apps they had previously interacted with and their opinions on them. Information on their use of messaging apps was of interest, as many CAs were delivered via these platforms. Public preferences can direct an appropriate delivery platform for these agents. The patterns of use of health apps could also be an important indicator of the comfort level the population is at regarding the use of technology to monitor their health and what applications they find mHealth solutions appropriate for.

Questions were asked about participants' knowledge of diabetes and prediabetes as well as their family history of these conditions. Identifying their knowledge gaps and awareness of being at risk could help in preventing and reducing the risk of developing diabetes and prediabetes. Finally, questions were asked about what they thought were the barriers and facilitators of a healthy lifestyle. The focus was on eating habits, physical activity, stress, and sleep. Participants were asked to elaborate on their current habits and how they would advise change for healthier habits. This would inform the content of future conversational agents educating the public on healthy lifestyle changes.

Both short-and long-answer questions were asked. An example of a short answer question was *"Do you use messaging apps such as Facebook Messenger?"* An example of a question that required a more detailed response was *"Sometimes we do not always eat the way we would like to. Let's talk about some difficulties you experience with eating healthy."*

All in-person interviews were conducted using a portable audio recorder. Telephone interviews were conducted via a speaker phone using the departmental iPhone 6 (Apple, Inc) and recorded using an audio recorder. The interviews lasted approximately 30 minutes. The audio files were transcribed verbatim. All transcripts were checked with the corresponding sound files for accuracy, and subsequently corrected for errors and cleaned.

6.2.3 Data Analysis

The data were analyzed using Burnard's method, a structured approach for thematic content analysis first established in 1991 and still avidly used today [296]. Two researchers worked to familiarize themselves with the transcripts by reviewing them several times. Next, initial codes were proposed from which themes were derived. Individual coding was piloted on five transcripts after which the two coders discussed and collated their themes for comparison. Finally, they reached a consensus on the themes to be used and how to define them. This information was presented in a comprehensive codebook. This codebook was then used to code the remaining transcripts. The coding of transcripts was performed using this codebook and the "review" and "comments" functions on a word processor. The final codebook is presented in Appendix 6.3.

6.3 Results

The 60 participants enrolled in the conversational agent feasibility study (via email), were approached to participate in a telephone interview (Figure 6.1). A total of 20 participants responded and agreed to be interviewed. This sample size was chosen based on the recommendation from existing evidence that qualitative analysis should be based on data from 1-30 respondents [297]. Seven interviews were conducted in person, and 13 over the telephone.

6.3.1 Participant Demographics

Eighty percent of the participants were female, median age was 30 (in complete years), age range was 23-60 years, and 50% (10/20) were married. Of the 20 participants, 65% (13/20) were working, 25% (5/20) were full-time students (undergraduate or postgraduate), and 10% (2/20) were unemployed. More than half of the participants (11/20, 55%) were Chinese, 25% (5/20) were Indian, 15% (3/20) were Malay, and 5% (1/20) was Caucasian. The monthly household income was SGD \$6000–S\$8000 (US \$4500-6000) for 25% (5/20) of the participants, SGD \$8000–S\$10,000 (US \$6000-7500) for 20% (4/20), SGD \$10,000–15,000 (US \$7500- 11,000) for 20% (4/20), SGD \$4000–S\$6000 (US \$3000-4500) for 15% (3/20), less than SGD \$4000 (US \$3000) for 15% (3/20), and more than SGD \$15,000 (US \$11,000) for 5% (1/20). The median monthly household income in Singapore in 2019 was SGD \$9425 (US \$7000) [298]. Only one participant had diabetes [P049]. More details on the basic demographics of the participants are outlined in Table 6.1.

Table 6.1 Basic demographic information of interview participants

Participant ID	Gender	Age	Ethnicity	Marital status	Monthly household income	Occupation
P01	Female	60	Indian	Married	\$20,000	Retired dental surgeon

P02	Female	25	Caucasian	Married	\$7000	Post graduate student
P03	Female	29	Chinese	Single	\$1890	Post graduate student
P04	Female	24	Malay	Single	\$5000	Post graduate student
P05	Female	23	Chinese	Single	\$13,000	Teacher
P06	Female	38	Chinese	Married	\$10,000	Engineer
P07	Female	23	Chinese	Single	\$10-15,000	Unemployed (recent graduate)
P08	Female	39	Chinese	Married	\$15,000	Human Resources (HR) Manager
P09	Female	53	Indian	Married	\$4000	Teacher
P10	Female	36	Malay	Married	\$6000-\$8000	Teacher
P11	Male	25	Chinese	Single	\$8000-10,000	Graduate student
P12	Male	50	Chinese	Married	\$8000-10,000	Logistics firm
P13	Female	30	Indian	Married	<\$4000	Researcher
P14	Female	54	Chinese	Single	\$6000-\$8000	Admin assistant
P15	Female	38	Chinese	Single	\$8000-10,000	Entrepreneur
P16	Male	44	Chinese	Single	\$6000-\$8000	IT
P17	Female	26	Chinese	Married	\$8000-\$10,000	Assistant Manager at Tariff (LKC Medicine)
P18	Female	25	Malay	Single	<\$4000	Childcare
P19	Female	28	Indian	Married	\$4000-\$6000	Customer service
P20	Male	24	Indian	Single	\$7000	Medical student

6.3.2 Thematic Analysis

In line with the methodology associated with thematic coding the data was segmented and then condensed into categories for thematic analysis [299]. Some themes were anticipated as they were explicitly highlighted during data collection. For example, the concept of barriers and facilitators to healthy eating, physically activity, good sleep habits and stress management were all points of discussion in the interview guide. Other promising ideas were assigned as coding categories through analytic induction which involved comparisons within and between cases [299]. For example, when participants were asked to describe their experiences interacting with conversational agents in the past, some referred to them as a “virtual call centre,” or “AI”. The diversity in names used resulted in a category “defining conversational agents”. Others discussed varying barriers and facilitators to their uptake. The definitions, barriers and facilitators were then drawn together and condensed under an umbrella theme of “use of conversational agents,” whilst each subordinate concept was assigned as a sub theme.

The transcripts were initially coded line by line by both researchers in parallel. As mentioned by Renata Tesch, the codes needed to be decontextualised from each interview and recontextualised into a new theme [300]. In this way, each transcript no longer contributed data in isolation but became better integrated with the other transcripts. Resultantly, all the points of interest from all the interviews were reduced and condensed into a meaningful list of themes. Both coders were always

aware of considering the suitability of each theme to the broader research focus of this needs assessment to keep the evolving analysis integrated and relevant [299]. As this needs assessment was concerned with identifying the barriers and facilitators to smartphone CA usage to promote healthy lifestyle changes for diabetes prevention in Singapore, the following four themes emerged from the data: (1) use of CAs, (2) ubiquity of smartphone apps, (3) understanding of diabetes, and (4) barriers and facilitators to healthy living in Singapore (Figure 6.2).

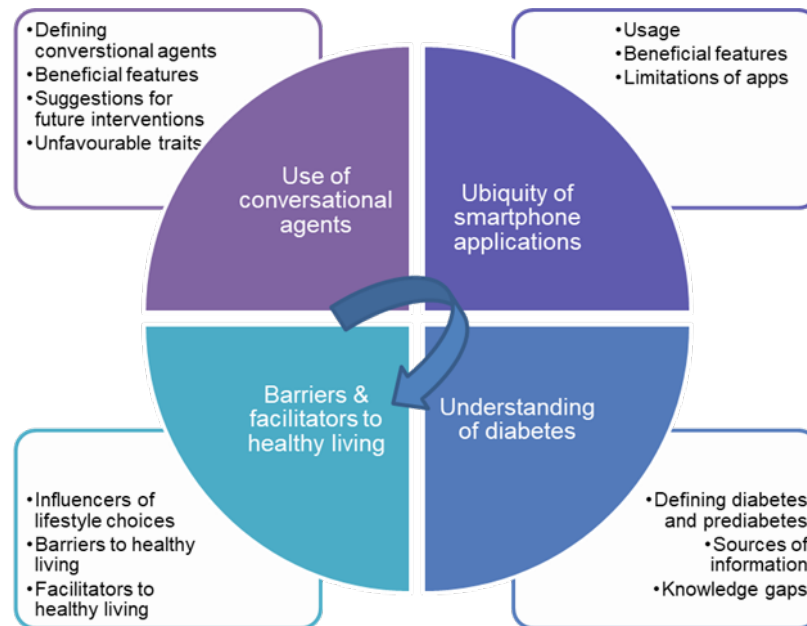


Figure 6.2 The four main themes and subthemes derived from the needs assessment interviews with 20 participants.

Each theme is presented below with subthemes (and some sub-sub-themes) and excerpts from interviews to highlight different facets of the theme.

i. Theme 1: Use of Conversational Agents

(a) Definition of Conversational Agents

Many participants managed to define CAs and mainly used the keywords “*AI program*” or a “*virtual call center*” with which one can communicate. Participants also noted that CAs were readily available at all times of the day, “*That’s 24-7, right?*” [P18]. Participants had encountered CAs when dealing with car insurance, public transport, web-based shopping, banking, and airline services. A couple of

participants were unfamiliar with this term and unable to provide a definition, *“No, never heard of that.”* [P19]

(b) Beneficial Features and Suggestions for Future Interventions

They offered their views on what makes a CA appealing. Their preferences included accurate sharing of information, interactivity, visual aids, text, predefined options, a group component, novelty, good reach, incentives, appropriate tone, frequency, duration, and bidirectional communication. Participants indicated a preference for visual aids (ie, images and videos) along with the text when communicating with the CA.

Predefined options were preferred over AI if the CA was not sufficiently programmed to deal with more complex questions and requirements. A few participants liked the concept of peer support in a CA to support their behavior change. Some participants wanted novelty to be introduced in the CA’s method of delivery to avoid the routine from becoming too monotonous and boring.

“But that one was not very good because they cannot... It becomes a pattern. Once they get used to it, then people will try to run away from the pattern.” [P15]

Novelty was proposed in the timing of message delivery and message format. For example, some may provide educational information or reminders for healthy behaviors or a quick tip for the day.

CAs that had a good reach (ie, were relevant to a wide range of age groups) were most appreciated. Participants mentioned that web-based or digital health platforms such as apps and CAs would be well received by the younger population. They pointed out a need to present all the relevant information, advice, and reminders on healthy living to older populations too who may be less technologically astute. They suggested this could be achieved by employing simple technology or supplementing with non-digital means. CAs that offered incentives in the form of rewards or discounts were also of interest. A couple of participants explicitly indicated greater interest in apps or CAs that provided discounts at grocery stores or offered coupons for redemption.

Participants had varied views on a suitable frequency and duration of interaction, ranging from daily to weekly. The average suggestion from all 20 participants was a messaging frequency of 2-3 days per week, where each interaction lasted for less than 20 minutes. Participants did not have much to say regarding the preferred tone of a CA besides a friendly and informal tone. They also mentioned that layman terms should be adopted for easy digestion of information. Bidirectional communication (where the user can respond and talk to the conversational agent) was preferred over unidirectional

communication (where only the conversational agent talks and the user consumes information). Some participants shared that bidirectional communication would give them the option to ask questions to the conversational agent, clarify doubts, or ask for more information on a topic of interest.

"I think two way should be an option. In case like there are any questions or anything you might want to follow up with. Or as well as like... you're interested in something, you can like ask for a contact or they could get you to a website for more information or somewhere that kind of thing." [P05]

(c) Unfavorable Traits

Participants shared instances in which CA use was unfulfilling. A few participants had negative experiences where the CAs they encountered were unable to address their queries or provided irrelevant responses to their input.

"When I had questions, they were not answered appropriately. So, I had to resort to calling them." [P09]

Others indicated that sometimes they were unsure of the conversational agent's level of intelligence and realized only later, through the interaction, that the agent was not sufficiently intelligent to address their more complex questions. In such cases, they would then prefer a simpler agent (with predefined options) that could fulfil its tasks more seamlessly.

ii. Theme 2: Ubiquity of Smartphone Applications (Specifically Messaging and Health Apps)

(a) Use of Mobile Apps

As conversational agents are frequently embedded in existing messaging apps or implemented as standalone apps, participants' use and opinions of messaging and health apps were explored. All participants used messaging apps, and 17 (85%) used health apps. The messaging apps mentioned were WhatsApp, Facebook Messenger, Telegram, WeChat, and LINE. Their uses included communicating, socializing in large groups, or sharing information in educational and professional settings.

A variety of health apps were mentioned with *Healthy 365* and step trackers being the most frequently used. Other apps were *"MyFitnessPal, AIA vitality, Fitbit, Samsung Health, Health Promotion Board app, Nike Run Club, Chronometer* (monitor nutrition), and *Stride* (habit tracker)".

(b) Beneficial Features

Participants attributed the beneficial features of apps to fulfilling a necessity, ease of use, engagement (promoted by the availability of stickers), and their multi functionalities, such as texting, calling, and video calling.

Something else participants shared which promoted smartphone app use, was the provision of incentives.

“And then for the Healthy 365 app by HPB (Health Promotion Board is a Singapore Government organization), I use it to redeem the rewards and participate in their promotional activities and programs.” [P17]

(c) Limitations of Apps

App accuracy was sometimes questionable, with different apps reporting different numbers for the same measure. Other mentioned that sometimes the large volumes of message content and high frequency of messages can be too overwhelming to deal with.

“It can be useful. But it also can be it can also be a harassment (laughs) turns out how you look at it” – [P06]

iii. Theme 3: Understanding of Diabetes

We explored participants’ understanding and knowledge of diabetes, prediabetes, and the sources of information that they used to obtain relevant information. Understanding their existing knowledge and identifying knowledge gaps, would provide valuable insights into the content required for a healthy lifestyle CA to prevent diabetes and prediabetes in the general population.

(a) Diabetes

All participants were able to define diabetes, with varying levels of detail. Certain individuals mentioned insulin or the inability of the body to control their blood sugar level resulting in high blood glucose levels. Others stated that diabetes requires one to reduce their consumption of sugar.

They were also aware that there may be a need to take medication and that there is no cure for diabetes. Participants shared the impact that diabetes can have on one’s health. A few were aware that diabetes can affect other parts of the body and lead to co-morbidities like kidney failure, stroke, high blood pressure, and limb amputations.

Besides its effects on other parts of the body, participants acknowledged the effects diabetes can have on an individual’s family, their ability to exercise and breathe with ease, and how it can culminate in high medical expenses.

"Having good health to be drained away. And then, how does this affect families concerned...I was told that people with diabetes may not be able to exercise actively. Have to change the type of exercise that they do. Sometimes breathing will be affected, also." [P14]

(b) Prediabetes

Majority of the participants defined prediabetes as a condition where someone is at borderline risk of developing type 2 diabetes. Some were also aware that prediabetes was reversible.

"Basically, you are at risk of diabetes if you don't change the way you live your life or your diet." [P20]

Other participants were very unfamiliar with prediabetes and acknowledged that diabetes was more commonly talked about.

"I always saw this like pre-type or something number 2 or something like that so I don't really understand it as well." [P08]

(c) Sources of Information

Participants obtained their information on diabetes from a variety of sources. Some shared that their personal experiences with diabetic family members educated them on diabetes and its implications.

"And you know when she (mother) was around, there's a diabetes society or something we used to go and visit, attend talks and yea, somewhere in (...) it's quite near my place actually but after she passed away uh I'm not involved in anything anymore." [P01]

Others mentioned the government's role in educating the public on diabetes, prediabetes, and healthy living and how the government can further influence the public soon.

"But from what I used to experience, the government does have certain policies acting through MOE (Ministry of Education). For example, last time we used to have no fried food in the canteen. I suppose they must be doing something with school canteens because of diabetes." [P17]

In addition, participants mentioned that a dearth of sufficient information was readily available to them, causing a barrier to awareness on healthy living. Some participants felt that the older generation may have less information because of lower levels of technological competency in searching for information on the web.

Participants proposed advertisements on the television and internet, as well as a display of the number of calories for each food item at hawker centers (food courts with affordable local meals) to improve information availability and population awareness.

"There could be more awareness like you know when we go to a restaurant or maybe hawkers more information could share like the calories. Because yea I know some hawker centers, they do show but it didn't exactly show the number of calories for all the food." [P08]

(d) Knowledge Gaps

Participants identified areas where they thought societal understanding of healthy living was lacking and could be enhanced in the future with mHealth. One such example was more information on physical and psychological changes for a healthier lifestyle.

"What the individual lifestyle changes that they can make or consider to make." [P12]

"Yes. Definitely. I think when you are more stressful, then your sugar goes up too." [P20]

"Mental health." [P15]

They made specific references to having a solid understanding of the importance of healthy eating and physical activity. Some proclaimed that instead of imparting more knowledge, what is needed are strategies on how to apply the knowledge concerning diabetes prevention and healthier lifestyle habits. More information on symptoms, prevention, risk factors, and awareness of being at risk of diabetes was of interest. In addition, there was an interest in more information on the implications of diabetes, its management, and sources of support.

"Maybe even things like prognosis if you are diagnosed, how are you going to manage it. If they need help, who can they look for." [P11]

iv. Theme 4: Barriers and Facilitators to a Healthy Living in Singapore

(a) Influencers of One's Lifestyle Choices

Participants shared their existing lifestyle habits and the factors that support or hinder healthy living. These barriers help inform what a future CA can aim to address when administering advice, education, and support for healthy lifestyle changes.

They first discussed the relationship between their existing lifestyle choices and how this related to their perceptions of current body weight. Some participants reported dissatisfaction with their current weight. They perceived themselves as either too heavy or underweight.

“Every time, women want to lose weight right? And for me its constantly trying to lose weight.” – [P08]

“Yes. I’m underweight. Yes. Not being overweight is not a problem. Yes. But, yes, I try to eat more to get to the acceptable weight. But over the years I’ve given up on trying to gain weight.” – [P20]

Participants also highlighted the existing determinants of their lifestyle choices.

i) Convenience

Convenience was a big factor when it came to diet and exercise. Food available in the close vicinity, easily accessible gym, or group exercise classes were more likely to be adopted.

“Maybe I’d join Zumba. I like that kind of thing but I’m thinking about affordability and the convenience.” [P18]

ii) Cost

Cheaper food was preferred, regardless of lower nutritional value, over healthier options if they were more expensive. Participants also indicated a desire to engage in group exercise sessions, but only if they were affordable.

“The biggest problem we have here is the price of healthy food. It’s just infinitely more costly. For example, some salad can cost you six, SGD 7 (US \$5), whereas chicken rice can cost you SGD 3 (US \$2). You cannot possibly eat salad every day even if you are earning okay. But if you are one of those that are low income, definitely you will go for three-dollar chicken rice every day. And in the long run, it just causes a lot of problems.” [P17]

iii) Personal Preferences

Some participants chose to eat food based on their personal preferences. Similarly, they chose to engage in physical activities that they enjoyed or that suited their schedules.

“Because I know that I’m not really hectic with the exercising, so I would tend to think that my house is three bus stops away from the interchange, so I will tell myself let’s just walk from the MRT back home, so at least it’s some walking activity for me.” [P18]

iv) Social Activities

Some participants said their social life influenced the amount of exercise they could do as more time spent socializing left less time for physical activity. Likewise, their choice of food was sometimes based on the social setting they were in and what was easily available there.

"I would say my social activities, like when I go out with my friends it's a bit hard to tell them let's eat healthy, let's just eat salad, so probably I'd say my social circles, so friends." – [P19]

v) **Advertising and Marketing**

Marketing of food was a factor influencing the kind of food they decided to eat in a day.

"It just depends on what I feel like eating or maybe a bit of marketing as well. For example, if McDonald's has a new menu, then I might just go for it." [P17]

vi) **Nutritional Value**

Sometimes, the nutritional value was a contributory factor in participants' decisions on what to eat.

"If I look at the menu item and then it looks like this is 600 calories, which is top of the list out of the whole list of things that I'm looking at, I would think twice about eating it. It does have some sort of impact knowing that you are intentionally consuming so much. You will trigger some automatic response to self-regulate." [P17]

vii) **State of Mind**

Some participants mentioned that their mood or state of mind largely determined the food they decided to consume. For example, when upset, stressed, or in a bad mood, they gravitated toward food that was more comforting than healthy. Likewise, their state of mind also influenced how much physical activity they would get. For example, low energy levels, stress, or fatigue after work meant they would choose to relax over exercising.

"like after a long day of work, I'm not like oh I'm going to go to the gym you know? Like nah I just want to stay at home and relax." [P05]

(b) Barriers to Healthy Living

Participants shared some barriers to healthy living.

i) **Dining Out**

Eating out regularly meant that participants had no control over the nutritional value of their food.

"Sometimes if you like to go to a restaurant and you hope that there's not so much MSG (monosodium glutamate – flavor enhancer) it's really beyond your control to tell the chef hey I don't want more... I

don't want the MSG or I want less salt. I mean you can tell them less salt. But pre-prepared stuff usually it's hard." [P06]

ii) Lack of Appeal

Some participants mentioned that healthy foods have an unappealing taste and texture, which dissuaded them from consuming those food items.

"Eating healthy, sometimes, they say is very bland, very tasteless. No fried is something without aroma. So, eating healthy sometimes lack of the aroma. And then they say it's either very too liquid or so tasteless, that is the thought they have." [P14]

iii) Health Condition

One participant shared that a health condition brought about compulsive eating habits, which caused her to deviate from a healthier diet.

"For me, in the past, because of my brain tumour, I have compulsive eating problems" – [P15]

iv) Fear of Injury

Some individuals mentioned a fear of injury, which may be caused by physical activity, prevented them from taking up some exercise. Hence, this fear became a barrier to a more active lifestyle.

"but there is also a fear that at a certain age there is some wear and tear in the body." [P14]

v) Lack of Free Time

Long working hours were associated with less free time in schedules for exercise. This lack of work-life balance also exacerbated stress.

"Whether I have time to dock off for a run or not. It really depends because most of the time, I'm actually working quite late.... I think for myself would definitely be work stress.... So myself sometimes I work 15 to 16 hours a day." [P08]

vi) Specific Events

Specific events, such as unexpected incidents, can disrupt an individual's equilibrium and cause stress. Changes in the weather, such as a hazy or wet period in Singapore, can prevent individuals from being physically active if there is a preference to exercise outdoors. Other events such as exams can negatively impact sleep schedules.

"I would say that previously when I was still studying it was a bit difficult to maintain eight hours of sleep because of exams and all that. But now that I'm working and I understand the importance of sleep"
– [P19]

vii) Lack of Knowledge

Lack of knowledge was mentioned by several participants as a barrier to healthy living. For example, some participants had trouble going to bed early and did not know how to fix this problem. Others were unsure about how to manage their stress, and hence, the problem persisted as the knowledge needed to overcome the barrier remained unknown.

"I'm having a bit of difficulty here because I'm experiencing a new sort of stress after coming back from my maternity leave. So, that's why I don't really know what to tell you because I think I haven't really learned how to manage the stress." [P13]

(c) Facilitators of Healthy Living

Participants shared some factors which can help to promote healthy living.

i) Early Intervention and Knowledge Building

Building knowledge on healthy living from a young age (in school) was stated as a promoter. Building more knowledge as an adult on the importance of healthy behaviors could also positively influence individuals.

"It's best to not just educate the older people, because we tend to think that the older people that they will be diabetic easier, so it's best to teach younger ones." – [P05]

ii) Reminders

Although the knowledge may already be there, reminders were cited to reinforce healthy eating behaviors. Participants also requested education on calorie counting, which would promote healthier eating.

"So, maybe when it comes to before lunchtime, two hours, tell them this is a suggested meal, then it's a guideline. Because everybody like to hold their phone, and then the message pops up, and they say, oh, yes, why not, I go and try." [P14]

iii) Incorporate Activity into Existing Routines

Sourcing out alternative methods of reaching the recommended weekly level of physical activity and how to be more active in general was helpful. These came in the form of efficient exercising (more vigorous activity for shorter durations) or reducing the amount of time spent sitting after a meal.

"So, yes, I only exercise 15 minutes each time. But I try to keep it more vigorous. So, yes, I do HIIT kind of training. So I do that about four or five times a week maybe. I think that there's more yield in doing vigorous exercise over moderate exercise. So in terms of the efficiency, the amount of time you need to spend, so I try to do more of HIIT." [P20]

iv) Empowering Oneself

Empowering individuals to support themselves and take responsibility for their health by making homemade meals, for example, was suggested as an effective way of eating healthier. Engaging in exercise at home and setting up their own workout routine was also mentioned. Similarly, healthy self-management of stress in the form of therapeutic activities was cited as a good way to maintain a healthier lifestyle.

"So, I have a treadmill at home and I use that. Like I walk as I'm doing some work on my laptop. So I try to incorporate that. Or I might do maybe like 10/15 minutes of yoga or cardio or something just around the house yea wherever I can." [P05]

v) Group Support

It was proposed that group support, could encourage more physical activity because of the collaborative spirit it evokes.

"If I did it with other people, like with friends and stuff yea, you could make it like a group effort. Then it'll be easier because then there are other people to monitor you, just for yourself." [P05]

vi) Internal Motivators

Internal motivators, such as wanting to be healthy in the present and future, encouraged individuals to take ownership and eat healthily too.

"I just think of wanting to be healthy, so tend to discipline myself on my food intake, so that's how I do it." – [P18]

vii) Discipline

Enforcing some discipline with regards to eating on time, drinking enough water, and planning meals for the week were all suggestions for promoting a healthier diet.

"But personally, I do my own meal prep actually. So, I cook for three days, so three of the lunches is covered for me." [P11]

Having good discipline and planning slots for physical activity were mentioned as effective methods for developing a more active lifestyle. Discipline in developing and sticking to a bedtime routine that involves winding down, relaxing, and reducing screen time as well as cutting down caffeine intake was proposed by participants to improve sleep habits. Disciplining food intake could be achieved through moderating portion sizes (overall quantity and food pyramid recommended portions), balancing healthy and unhealthy foods, controlling, and restricting unhealthy food items. These choices were cited as effective promoters for healthy living.

“For example, bubble tea right, everybody likes bubble tea, so to me, I will limit myself to buy that maybe once every two months, take that as a reward for myself, yes, that’s what I’d do.” [P18]

viii) Moderation

Participants recommended establishing moderation at work to have a more work-life balance to reduce stress levels.

“And I think having a good balance, life, social life, helps as well, that it’s not just work.” [P11]

Others also mentioned catching up on sleep over the weekends to moderate overall sleep quantity.

6.3.3 Trustworthiness in qualitative research

The quality and trustworthiness of this study was assessed using the credibility, transferability, dependability and confirmability criteria [301, 302]. Please refer to Chapter 10.9 for a detailed account of the quality assessment and the reflexivity account for the qualitative studies presented in this thesis (Chapter 6 and 8).

6.4 Discussion

6.4.1 Key Findings

To our knowledge, this is the first qualitative study to investigate the public's perception of conversational agents for a healthy lifestyle change intervention in Singapore.

Participants shared their perceptions of CAs, where they provided definitions, usage experiences, and preferential features. These are pivotal in informing the design elements for a relevant and well-accepted health CA for the Singaporean population. Participants also talked about the ubiquity of smartphone apps, whereby messaging apps were a commodity in their daily lives, used for communication, social, educational, and professional purposes. They also highlighted the most frequently used and preferred messaging app platforms. This knowledge is beneficial when considering CA delivery channels that can be embedded in messaging platforms most familiar to the target population for easy adoption.

Health apps were used for tracking of steps, health (nutrition, water intake, and habit tracking), or pairing with a Fitbit watch. The diverse patterns of use indicated an inclination to use mHealth for numerous purposes. Furthermore, participants' views on the beneficial and limiting features of apps and CAs were important insights into developing a health CA for this population. They shared their existing knowledge on diabetes, prediabetes, and existing sources of information. In addition, they offered their views on what constitutes healthy living with a focus on diet, exercise, sleep, and stress as well as thoughts on the barriers and facilitators of healthy living in Singapore. The culture of eating out in Singapore is encouraged by affordable hawker centers, fast food, and fast-paced life, which does not allow much time for grocery shopping and cooking of food at home. Frequent dining out, high-stress levels, lack of work-life balance, and lack of free time to exercise were some of the barriers cited for healthy living. In addition, work-related stress contributed to poor sleep quality and quantity. Participants proposed solutions for some of these barriers such as self-discipline, preplanning, and sticking to a routine for healthier living patterns. Given the ubiquity of smartphones, the avid use of messaging apps and an inclination to use mHealth initiatives, and awareness of these barriers and facilitators for healthy living has the potential to be communicated to the population via a CA.

6.4.2 Comparisons with Existing Literature

In 2019 Lim et al. explored the barriers and facilitators of healthy eating in Singapore [303]. It was noted that items that encouraged healthy eating included self-discipline, fear of disease complications, education by a health professional, mass media, and health promotion campaigns [303]. Similarly, participants in this needs assessment also mentioned self-discipline as a strong facilitator and acknowledged the presence of health promotion campaigns without explicitly citing these as influencers of their lifestyle choices. Furthermore, fear of disease complications and education by health professionals was more relevant in Lim et al. than this study [303], as participants were already diagnosed with prediabetes and therefore at higher risk of complications. This also points out the need for a more hybrid approach when it comes to more high-risk populations (such as those with prediabetes), as HCP involvement could be beneficial in addition to a CA. Furthermore, the lack of knowledge and skills to prepare and select healthy foods was shared as a difficulty in healthy eating [303]. This barrier to healthy eating was also noted in this needs assessment. Additionally, participants added that a lack of time to prepare home cooked meals was due to their hectic schedules, making dining out a more convenient, but often less healthy, compromise.

Another study noted that healthy eating and physical activity were of prime importance for health education and communication in patients with prediabetes in Singapore [304]. This needs assessment yielded a similar outcome however, the components of sleep and stress were also considered for a more holistic view of healthy living. Other necessities for health communication were risk factors and diabetes prevention [304]. These were also identified in the qualitative analysis. In addition, participants shared an interest in information on where to go for help if they were at risk or diagnosed with prediabetes/diabetes.

A scoping review of conversational agents showed that qualitative data were presented in some studies to show the acceptability and satisfaction of these interventions, namely, participants' opinions of an already developed intervention [42]. These parameters were often reported with Likert scale ratings, but not comprehensively or thematically [134, 150]. As such, an extensive qualitative analysis, such as the one presented here, for a CA intervention has not been performed so far in the Singaporean population.

In a study on behavior change in overweight adolescents, high compliance to CA interventions was attributed to a rewarding game system [135]. This aligns with a suggestion from the participants in this study who mentioned that incentives will be well received and will contribute to their

willingness to adopt and continue to use a CA. Other studies have examined the use of conversational agents for chronic conditions, including diabetes, and reported a preference for personalization features [163]. Personalization was also noted as an advantageous trait by this study sample.

Ta et al. [305] discussed a popular CA available on app stores, *Replika*, with generic capabilities such as stress management and social support. Based on the analysis of public user reviews and a survey conducted with participants recruited via social media in the United States, companionship was the greatest form of social support users identified. The positive feedback was attributed to Replika's ability to use a range of message types (eg, text and images) and a high level of anthropomorphism. Similarly, interviewees in this needs assessment supported the use of visual aids along with text. Although they did not indicate specific personality preferences or the need for the CA to be human-like, a general inclination toward a friendly and informal tone was identified.

Studies on physicians' perceptions of conversational agents have indicated their apprehension toward agents that are unable to comprehend the emotional state of vulnerable patients and have dissuaded their use in cases where expert medical knowledge and skills are required [45]. However, they have supported the use of CAs to support, motivate, and coach patients. In line with this, the CA to be developed will support healthy lifestyle changes by understanding the needs and preferences of prospective users via this needs assessment.

6.4.3 Implications for Future Work

Future digital interventions aimed at encouraging diabetes prevention in the general population to promote and prolong healthy living can incorporate the content-and delivery-specific options highlighted in Table 6.2.

These suggestions include providing information directly relevant to diabetes prevention, building up an awareness of risk factors, symptoms, and prevention methods, as well as the complications and consequences of diabetes and prediabetes to build up users' knowledge. Future applications should also incorporate actionable advice so that participants can adequately implement the advice provided to them via the CA.

Prospective users could benefit from advice that nurtures skills related to food choices, calorie counting, and stress management. Advice on healthy eating should be mindful of the Asian diet that most Singaporeans are familiar with and have easy access to. Highlighting the number of calories in frequently consumed food items in Singapore hawker centers could be a useful start to support calorie counting. Stress management techniques relevant to establishing a work-life balance would

be beneficial, as this is a common complaint in Singapore, as disclosed by our participants. In addition, teaching participants time management skills may help to reduce overall stress, freeing up time for physical activity, meal preparation, and a bedtime routine.

Future work into CAs can also explore options to reach the weekly recommendation for physical activity, such as methods to increase efficiency, make suggestions for safe exercise options, or provide options to seek out a healthcare professional for more personalized advice. Group exercise is a concept of interest to make and should also be acknowledged. Cheaper options or even web-based options may be listed. Recommendations on alternative methods of exercising, which do not require participants to be outdoors, could be suggested to overcome barriers related to the weather, such as home workouts. A CA could also reinforce the existing promoters of healthy behaviors, for example, reminding them of the feelings of contentment they will experience following physical activity.

Participants in this study also provided recommendations based on their preferences and what they believe others will enjoy, for the delivery of future CA interventions, for healthy lifestyle changes.

These included: personalizing the conversational agent to the user's preferences (timings, advice, etc.), making the interface easy to navigate, making the exchange engaging, and introducing some novelty to boost user experience. The tone, personality, and language used by the conversational agent should also be adapted to suit the target population (eg, being friendly and using simple). Some additional considerations, which could be included, are incentives, such as rewards or discounts for groceries, and group support, as participants indicated a preference for these too.

Table 6.2 Recommendations on content and delivery for future conversational agent interventions for healthy lifestyle behavior change

<u>CONTENT</u>	
<u>Furthering knowledge on:</u>	Implications of diabetes and prediabetes
	Awareness of being at risk
	Symptoms and prevention methods
<u>Actionable advice on:</u>	How to apply the knowledge gained on healthy living
<u>Further development of skills relating to:</u>	Eating healthily
	Calorie counting

	Options available for meeting the weekly recommendation for physical activity
	Stress management
	Time management
<u>Type of content:</u>	Novel – up to date
	Trustworthy – evidence-based
<u>DELIVERY</u>	
Personalization of content	Advice on how to incorporate exercise into one's specific schedule, availability of facilities, and preferences. Habit formation – how to discipline your healthy eating, drinking water, physical activity, and bedtime routine
Suitable frequency and duration	Ask participants for their preferences. Offer options for participants to choose from
Easy usage	Pre-defined options for better efficiency if AI is not yet maximized
Engaging	Use stickers and visual aids. Make the conversation bi-directional so the user can be part of the exchange and can ask questions if necessary
Relevant to the target population	Tone, personality, and language used by the conversational agent to be easily understood, relevant, and appropriately tailored for the target population
*Additional considerations to keep users keen and engaged in the interventions.	- Incentives eg. rewards and vouchers. • Group support eg. a chat channel for participants to communicate with others using the same conversational agent to share advice and motivate each other.

Based on this study, a rule-based CA was developed and tested in a feasibility study (Chapter 7).

6.4.4 Strengths and Limitations

The study sample included 20 participants who were ethnically diverse and covered a wide age range (23-60 years). The collected data were coded in parallel by two researchers, who worked towards a common coding framework through a series of discussions. These findings may also apply to other

high-income or technologically astute Asian countries. Although 20 participants were adequate for this analysis, a larger sample size would be preferred in future studies for greater validity and generalizability.

The focus of this study was extrapolated from healthy living to diabetes and prediabetes prevention. As such, the suggestions on healthy living would be relevant not only in the context of wellness but also as a means of preventing these chronic conditions of great burden in Singapore. The participants in this study were all from Facebook groups and pages with an interest in healthy living possibly lending towards some recruitment bias. These individuals had some knowledge and interest in healthy living and some level of technological competency to use Facebook. In addition, despite the purposive sampling, 80% of participants were female because of the interest shown by the study volunteers. Both these factors may reduce the generalizability of the findings as they may not be entirely representative of the general population in Singapore.

6.4.5 Conclusions

The participants provided valuable insights on their existing knowledge and sources of information on healthy living and diabetes. They also shared their current usage patterns of messaging and health apps, and their views on CAs, which have not yet been used for health care purposes in Singapore.

In addition, they offered opinions about the importance of healthy living and diabetes prevention. Preferences for CA content and delivery were also discussed. Finally, they shared views on barriers and facilitators of healthy living. These findings can be used to inform the development of future CA interventions and similar mHealth initiatives that target healthy lifestyle behavior changes.

Chapter 7 – Conversational Agent for Healthy Lifestyle Behavior Change: Web-Based Feasibility Study

Chapter 7 is concerned with objective four of this thesis, to “develop and test the feasibility and acceptability of a healthy lifestyle behavior change conversational agent for the general population in Singapore in an online feasibility study.” Section 7.1 provides a brief overview of the background for this study and the methods are discussed in Section 7.2. This section is broken down into methods for recruiting participants, intervention development, carrying out the study, collecting outcomes, and analyzing data (Sections 7.2.1 – 7.2.7). The results for all the primary and secondary outcomes are described in Section 7.3. Section 7.4 denotes the discussion where the key findings and the implications of these for the future are explained. The strengths and limitations of this study are also shared (Section 7.4.4) before the chapter concludes.

7.1 Introduction

CAs have the potential to deliver behavior change interventions. They have been effective and accepted for behavior change interventions related to mental health, obesity treatment, and smoking cessation. In Singapore, the chronic diseases of concern are obesity and type 2 diabetes [306] with 600,000 adults living with diabetes. Singapore ranks second among high-income countries, and obesity levels are also on the rise [307]. In addition, 34% of the men and 39% of the women in Singapore do not reach the weekly target of 150 minutes of moderate-intensity activity per week, leaving people at higher risk of developing diseases such as diabetes [55]. Research has shown that lifestyle interventions (such as increasing physical activity and eating a healthy diet) delivered by trained health care professionals can help to promote healthy weight loss [60-62]. Lifestyle change directed at high-risk populations also showed a reduction in the incidence of type 2 diabetes [60].

Administering such lifestyle changes on a large scale can be achieved using remote and automated methods such as conversational agents. Singapore is a technologically astute country, and citizens avidly use messaging apps. In addition, Singapore’s ministry of health has proposed the increasing use of conversational agents in health care in tackling issues such as the rising chronic disease burden and the aging population [308] that can lead to more primary care appointments. The ministry envisions a near future where a conversational agent can collect a patient’s history from them before their consultation, streamlining the primary care visit and thus cutting down waiting times. This makes Singapore an ideal place for the evaluation of novel mobile health interventions such as conversational agents.

7.1.1 Objective

The evidence around the use of CAs for healthy lifestyle change is limited and knowledge on the feasibility and acceptability of such novel interventions are essential to inform future trials. Correspondingly, in this study, we aimed to assess the feasibility, acceptability, and preliminary efficacy of the use of conversational agents for healthy lifestyle behavior changes in the general population in Singapore.

7.2 Methods

Considering the time and expense involved with efficacy testing, a feasibility study design was selected to test the feasibility and acceptability of such a healthcare CA to the general population in Singapore. The findings from this study would provide insight into how the CA can be optimised and the intervention amended for meaningful efficacy testing in the future.

7.2.1 Ethical considerations

This web-based single-arm feasibility study was approved by the Nanyang Technological University Ethics Committee (IRB-2018-11-032). Ethical considerations for the feasibility were limited as it was completely online categorised as a non-HBR (Human Biomedical Research) study. All interactions with participants were done remotely and informed consent was obtained digitally prior to study commencement. The informed consent form was accompanied by a study information sheet which provided the contact details of the principal investigator, the IRB-reference number, the study title and objectives, the procedures the participant will have to follow. The information sheet also highlighted their right to refuse or withdraw at any time and mentioned that this was a non-invasive study with minimal, if any, risks involved. Participants were to receive a compensation of \$25-35 upon successful completion of the whole study. Participation in interviews was completely voluntary and the consent form mentioned that these interviews would be recorded. The final section confirmed that their anonymity and confidentiality would be protected whereby only the researchers would have access to their personal details, that no identifiable information will be presented in subsequent publications and that the records would be stored for a minimum of 10 years on a secured platform with restricted access at NTU. This single information sheet and consent form was completed by all recruited participants and so applies to the research presented in Chapters 6 and 8 as well.

7.2.2 Participants

Participants were eligible if they were aged above 21 years, were Singapore citizens or permanent residents, owned a smartphone, and had a Facebook Messenger (Facebook, Inc) account. Prospective participants were excluded if they were pregnant or had any of the following conditions: cancer, chronic liver disease, chronic kidney disease, a neurodegenerative condition, heart disease, stroke, a physical disability, hypertension, or a condition that did not allow for regular physical activity. Eligibility was assessed using an eligibility questionnaire.

7.2.3 Recruitment

The participants were recruited through Facebook in August-September 2019. A digital poster listing the study aims and eligibility criteria was uploaded on relevant Facebook pages focused on healthy living, such as *Singapore fitness and health community* and *Singapore healthy cooking*. In addition, we used snowballing in our recruitment; whereby the study poster was shared by potential participants through messaging apps or word of mouth.

7.2.4 Intervention

The conversational agent was developed on a free online tool *Chatfuel* to be used on Facebook Messenger [309]. A research associate (AS) and a Ph.D. student (DAD) developed the script for the program and entered the content and scheduled the timings and intervals of messages on *Chatfuel*. The intervention included information on diabetes and prediabetes, healthy diet, physical activity, good sleep habits, and stress management. These were the key topics identified in other diabetes prevention programs targeting lifestyle change [310-312].

The content was informed by existing evidence-based sources of information, including clinical guidelines and systematic reviews. Advice on improving sleep quality was generated from published evidence reporting on techniques and successful interventions for sleep disorders [313]. The section on stress was informed by distance learning-based stress management techniques identified from a review of existing studies that described methods to reduce stress and for health improvement [314]. The nutritional advice was obtained from authenticated government health portals and validated health and nutrition webpages advising dietary modifications to reverse prediabetes [315]. Information on physical activity was informed by Singapore's Health Promotion Board where advice was provided on the recommended duration and intensity of exercise [315]. Validated fitness routines were also presented to show users how to perform the exercises and in what combinations [316]. The collated advice was compartmentalized into themes, which became a topic of focus for

each interaction between conversational agent and user (Appendix 7.1). Examples of conversational exchanges between the conversational agent and users are presented in Figure 7.1.

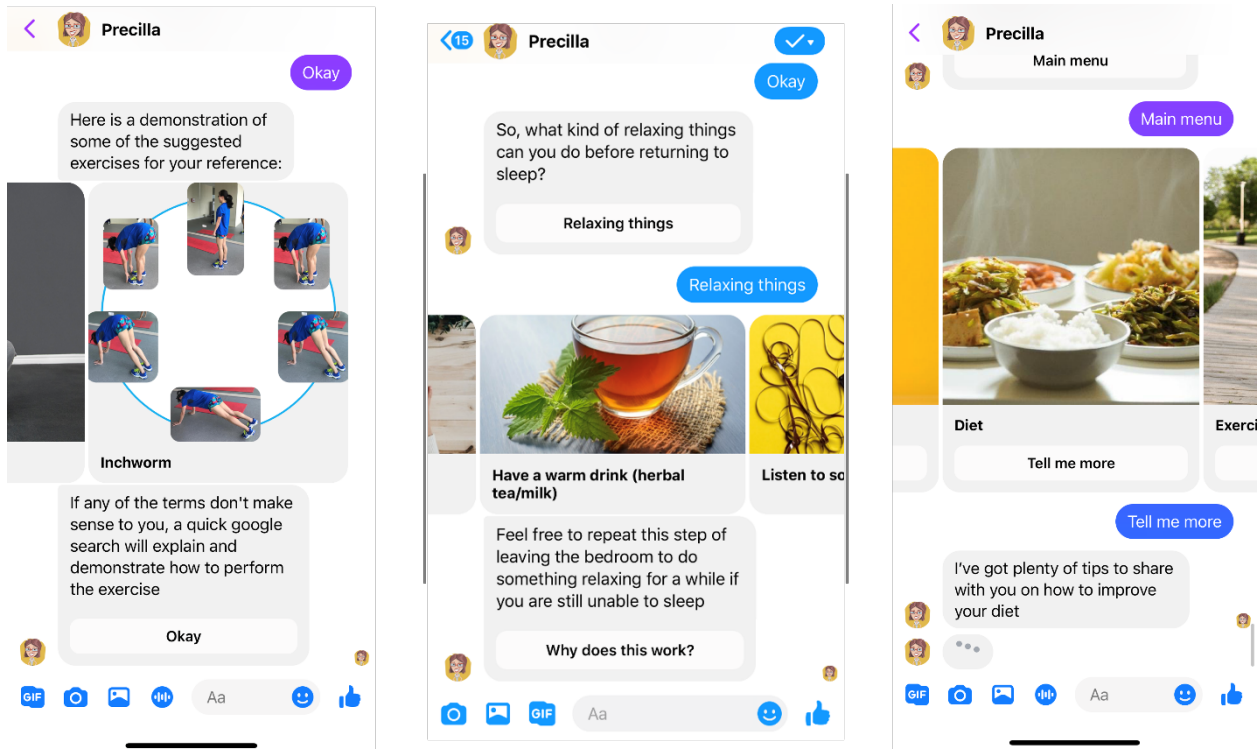


Figure 7.1 Example of conversations between the user and CA *Precilla*.

As this was a CA that would target behavior change, eventually, we also followed the Capability, Opportunity, Motivation, Behavior model of behavior to guide its development [30]. This model posits that to exhibit a particular behavior (B), the participant must be physically and psychologically capable (C), have sufficient social and physical opportunities (O) to perform the behavior, and must have the desire or need to do so—motivation (M).

The main mode of communication was text, which was supplemented with visual aids (images/videos) to make the CA more engaging to enhance the user experience. These images were obtained from free-to-use sources or photographs taken by our study team members. To further contribute to a positive user experience, the conversational agent was given a name, *Precilla*, and she had a friendly tone of speech, profile picture, and used the *typing* function for messages to make it seem like the user was interacting with another human. From here onward, “*Precilla*” will be used to refer to the conversational agent used in this pilot study.

The intervention was scheduled to last for 4 weeks, whereby participants would receive 4 messages per week, (every other day), 1 for each topic (diet, exercise, sleep, and stress). A sample of the dialog tree is presented in Appendix 7.2.

7.2.5 Procedure

An overview of the study workflow is demonstrated in Figure 7.2 The needs assessment was discussed in Chapter 6 and the follow-up interviews have been explained in more detail in chapter 8. This chapter focuses on the 4-week pilot study and the measures directly related to this phase of the study.



Figure 7.2 Overview of feasibility study workflow from screening to follow-up.

The baseline questionnaire obtained general information regarding demographics as well as data on diet, physical activity, sleep, and stress. Upon completion of the pilot study, participants completed a follow-up questionnaire. This questionnaire covered all the details outlined in the baseline questionnaire, with some additional questions on conversational agent usability and users' overall satisfaction with the study.

7.2.6 Outcomes

(a) Primary Outcomes

The primary outcomes of interest for this study were the feasibility of recruitment and retention of participants, acceptability of the intervention, and participant engagement with the CA.

(b) Feasibility of Recruitment and Retention

Feasibility of recruitment was defined as the ability of the researchers to recruit at least 75% (150/200) of the target sample using free methods on Facebook within 1 month. The feasibility of retaining participants was defined as at least 32.7% (49/150) of the recruited participants completing the study.

(c) Acceptability of the Intervention

The acceptability of Precilla was measured through questions on usability and satisfaction in the follow-up questionnaire. Questions were asked on the participants' overall satisfaction with Precilla, their likelihood of using Precilla again and recommending it to others, as well as the impact of the interactions on their health. The questions in the follow-up questionnaire were directed by assessments and measurements made in other CA evaluation studies identified in a scoping review (Chapter 4) [42].

There were 2 sections in the questionnaire; the questions in section 1 related to how participants perceived the content of Precilla's input. These included questions on ease of use, enjoyment, long-term use of Precilla, and language, as well as motivation to make healthier food choices, exercise, change sleeping habits, and better manage stress. Users were asked to provide a rating of strongly agree, agree, neutral, disagree, or strongly disagree for each statement.

The questions in section 2 were concerned with the usability of Precilla and its delivery methods. The participants were provided with the same options to choose from as in section 1 (strongly agree-strongly disagree), and they were asked to share their opinions on the mode of communication (buttons, text, images, etc), the suitability of Facebook Messenger as a channel of communication, the number of messages sent, the timing of messages, and Precilla's personality.

(d) Participants' Engagement

A further measure of acceptability was the participants' engagement with Precilla. This data was collected manually by analyzing individual conversations between Precilla and the participants on Facebook Messenger. This involved noting down the duration of interactions and counting the number of complete, incomplete, and nil interactions. Immediate responses were defined as interactions made by the user within an hour of message receipt. We also collected data directly from *Chatfuel* analytics, such as user retention, free text typed by users (such as questions they had asked Precilla or random utterances). In addition, participants were asked to rate their technological competency on a scale from 1 to 10.

(e) Secondary Outcomes

The secondary outcomes were related to the efficacy of the intervention and considered participants' quality of life (QoL), diabetes knowledge, diet, physical activity, sleep, and stress over 4 weeks.

QoL was measured using the short-form version of the QoL Enjoyment and Satisfaction Questionnaire [317]. The questionnaire had 14 items rated on a 5-point scale that indicated the

degree of enjoyment or satisfaction experienced in the past week. The total score for the 14 items ranged from 14 to 70 and covered the topics of work, social life, health, and overall well-being. A percentage of the maximum score was also reported; for example, if a participant scored 20, the percentage was 29% (20/70). The last two questions on medication adherence and overall life satisfaction were reported as percentages (Appendix 7.3).

The knowledge questionnaire had 3 separate sections. Section 1 asked about participants' knowledge on healthy living, prediabetes, and diabetes; section 2 asked for their perceptions on how healthy their lifestyle is and their likelihood of developing diabetes; section 3 tested their knowledge of risk factors related to type 2 diabetes. The questionnaire was an adaptation of questions presented by the Michigan Diabetes Research Centre [318]. The questionnaire was adapted such that, in section 1, questions regarding the definitions of diabetes and prediabetes were added to test participants' knowledge of these conditions. In section 2, the wording of some questions about participants' efforts to make healthy lifestyle behavior changes in the past year was altered to obtain more detail (in the form of a scale) compared to just 'yes' or 'no' responses. In section 3, some risk factors that did not apply to Singapore were removed (eg, being Asian American, Hispanic, or African American).

An adaptation of the Food Frequency Questionnaire (FFQ) was used to ask participants about their diet. Questions were asked on the frequency of daily intake of vegetables, fruits, fried foods, and sweetened drinks [319]. Physical activity was assessed using the International Physical Activity Questionnaire (IPAQ), adapted to suit this study setting [320]. Questions were asked on the intensity of exercise (vigorous, moderate, or light), the frequency of physical activity (in days per week), and the length of each session (in minutes). Only sessions lasting at least 10 minutes qualified as physical activity. In line with IPAQ scoring, a metabolic equivalent of task (MET) score was calculated. The MET score acted as a proxy for the amount of energy expended when carrying out physical activity. For consistency, walking was given a score of 3.3 METs, moderate physical activity 4 METs, and vigorous physical activity 8 METs. The MET minutes per week were calculated by multiplying the MET value by the number of minutes for which the activity was carried out and again by the number of days in the week that the activity was undertaken. As some participants provided a range for their responses, for consistency, a mean value was used for the calculation. For example, for the number of days, that is, 1-3 days, the mean was calculated as 2. For session length, that is, 10-20 minutes, the mean was calculated as 15.

A MET score of 600 METs per week denoted a moderately physically active person, whereas 1500 METs per week indicated a high level of physical activity. Any score that did not qualify as moderate or high was considered a low level of physical activity.

Sleep scoring was done using the Pittsburgh Sleep Quality Index (PSQI) questionnaire [321]. A score of 0 indicated excellent sleep quality, whereas 10 indicated severely poor sleep quality. Poor sleep quality was defined as participants with a global PSQI score higher than 5. The stress level of the participants was assessed using the Perceived Stress Scale (PSS), where it was possible to receive a score between 0 and 40 [322]. Scores of 0 to 13 were considered low perceived stress, 14 to 26 moderate perceived stress, and 27 to 40 high perceived stress.

7.2.7 Data Analysis

Data analysis was conducted on the data collected at baseline and follow-up separately (Table 7.1). Satisfaction and usability were presented as percentages of each option on a Likert scale (eg, strongly agree, disagree, much better than before, or neither better nor worse). The participant outcomes for QoL, knowledge, physical activity, sleep, and stress were all presented as pre-and post-scores using the individualized scoring system of each questionnaire. Percentages, means, and SDs were provided.

Table 7.1 Overall summary of results from pilot study questionnaires (N=60).

	Baseline values (n=60)	Follow-up (4 weeks) values (n=56)
Knowledge, mean (SD; range)		
Section 1	7 (1.37; 6-18)	6 (0.93; 6-18)
Section 2	28 (2.7; 12-48)	27 (3.5; 12-48)
Section 3	10 (4.1; 8-32)	9 (2.2; 8-32)
Sleep, mean (SD)	4 (2.36)	4 (2.45)
Stress, mean (SD)	17 (5.13)	16 (5.10)
Physical activity (MET ^a score), mean (SD)	1080 (816)	1075 (872)
QoL^b		
14-item score, mean (SD)	54 (6.90)	53 (6.46)

	Maximum score (%)	77	75

^aMET: metabolic equivalent of task.

^bQoL: quality of life (measured using the short-form version of the QoL Enjoyment and Satisfaction Questionnaire).

We performed descriptive analyses of the data. As this was a feasibility study, no hypothesis testing was done to assess the efficacy of the intervention [323][38]. Alternatively, mean differences between baseline and follow-up were presented, accompanied by 95% confidence intervals (CI) [323].

7.3 Results

7.3.1 Recruitment and Retention

A total of 136 individuals expressed initial interest in participating in the study; however, 9 (6.6%) were ineligible because of comorbidities or because they had not installed Facebook Messenger on their smartphone, and a further 52 (38.2%) completed the screening questionnaire but did not proceed to provide informed consent, leaving 75 (55.1%) participants eligible for participation. Of these 75 participants, 60 (80%) completed baseline assessments, and of these, 56 (93%) completed the follow-up questionnaires (Figure 7.3).

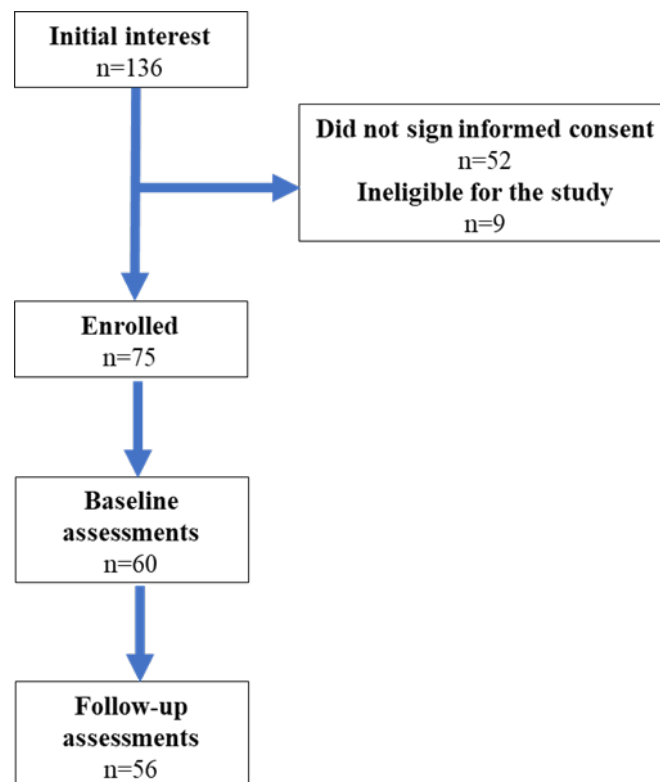


Figure 7.3 Numbers of participants who were initially recruited, completed baseline and follow-up assessments for feasibility pilot study.

7.3.2 Characteristics of Participants

Of the enrolled participants, 38% (23/60) were men. The mean age was 33.7 years (SD 9.3), the mean BMI was 22.3 kg/m² (SD 3.8; Table 7.2), and the average technical competency on a scale of 1 to 10 was 8.07.

Table 7.2 Characteristics of all the enrolled participants who completed baseline assessments for the pilot study (N=60).

Characteristics		Baseline values
Age (years), mean (SD)		33.7 (9.3)
BMI (kg/m ²), mean (SD)		22.3 (3.8)
Gender (male), n (%)		23 (38)
Ethnicity, n (%)		
	Chinese	48 (80)
	Malay	5 (8)

	Indian	3 (5)
	Others	
	White	2 (3)
	Burmese	2 (3)
Marital status, n (%)		
	Currently married	25 (42)
	Never married	33 (55)
	Separated	1 (2)
	Divorced	1 (2)
Highest level of education, n (%)		
	University and above	50 (83)
	Polytechnic diploma	1 (2)
	Other diploma and professional qualification	1 (2)
	A ^a -level or NTC ^b -1 or NTC-2 or certificate in office or business skills or its equivalent	6 (10)
	O ^a or N ^c -level or NTC-3 certificate or its equivalent	1 (2)
	Secondary school	1 (2)
Work status, n (%)		
	Employed	37 (62)
	Student (full time)	17 (28)
	Homemaker or housewife	3 (5)
	Unemployed (able to work)	2 (3)
	Retired	1 (2)
History of parents, sibling, or child with type 2 diabetes, n (%)		
	Yes	11 (18)
	No	49 (82)
History of hypertension, n (%)		
	Yes	2 (3)
	No	58 (97)

^aA-level or advanced-level examinations are taken by students at the age of 18, 2 or 3 years after completing their O-level or ordinary-level examinations, which are taken by students at the age of 16 after 4 years (or 5 years) of secondary school.

^bNTC: National Technical Certificate.

^cN-level: The Singapore-Cambridge General Certificate of Education Normal-level. Secondary students in Singapore can move between two streams based on their academic performance: 4 years of study culminating in the O-level (Ordinary level) or the N-level (Normal level) examinations. N-level students may participate in a fifth year of study to take the O-level examinations.

7.3.3 Satisfaction

Of the 56 participants who completed the whole study, 52 (92%) were *moderately or very satisfied* with Precilla, 30 (54%) thought that their likelihood of recommending Precilla to others would be *somewhat likely*, 32 (57%) felt that the likelihood of their opting to use Precilla again for personal use would also be *somewhat likely*, and 29 (51%) posited that their health was *somewhat better or much better* than before.

7.3.4 Usability

The *agree* and *strongly agree* responses exceeded 50% (28/56) for all the questions relating to Precilla's acceptability, except for question 5, "Chatting with Precilla motivated me to change unhealthy sleeping habits," where the collective response in agreement (*agree* and *strongly agree*) was only 46% (26/56) (Figure 7.4). The smallest percentage of disagreement was for questions 1 and 8. This implied that most of the participants (55/56, 98%) found the chat easy to use and found Precilla's language simple and easy to talk to. The *neutral* response was a common selection by participants for questions 2-7.

67% (37/56) of participants agreed that Facebook Messenger was an appropriate delivery channel for Precilla whilst the remaining participants were either neutral or disagreed. *Telegram* and *WhatsApp* were suggested as potential alternatives. 15% (9/56) thought that too many messages were exchanged with Precilla and they would have preferred the intervention to be more streamlined; 12 (21%) disagreed that noon was the most appropriate time to receive messages, whereas the remaining 44 (79%) either agreed, strongly agreed, or were neutral (Figure 7.4) in which case they indicated a preference for interacting with Precilla after work, in the evening or before bedtime. Specific timings were offered, such as 5 PM, 8 PM, or 10 PM, and 1 participant

suggested 9 AM or during the morning commute to work. There were no disagreements to statement 13 (“I liked images and videos in Precilla’s messages”) and minimal disagreement for statements 14 (5/56, 9%) and 15 (4/56, 7%) on the inclusion of more visual components and Precilla’s personality, respectively. Most participants did not find Precilla to be very human-like and could tell that they were communicating with a conversational agent. Of the 56 participants, 3 (5%) suggested that in the future, some degree of personalization should be incorporated, whereby Precilla’s messages would be initiated only when prompted by the user and not the other way around.

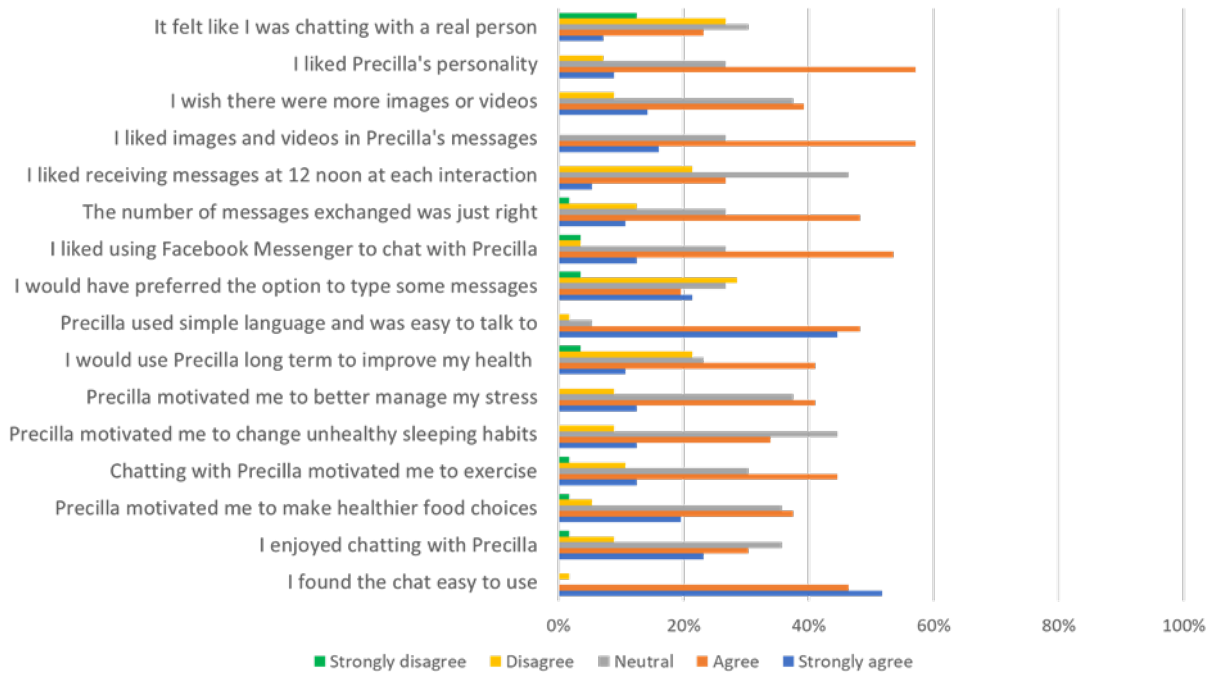


Figure 7.4 Usability scores for Precilla

7.3.5 Conversational Agent Engagement Data

All the conversations were completed by at least 50% (>30/56) of the participants consistently throughout the study (Figure 7.5). The number of participants who completed interactions from start to finish exceeded 71% (40/56) for 56% (9/16) of the interactions, denoting high engagement with the conversational agent. Some interactions were started but not completed. A frequent cause of disruption to the conversation flow was participants typing free text instead of selecting their response from the predefined options. This led them off track to the main menu, which they potentially found confusing, or they were reluctant to go through the whole conversation flow again and, hence, failed to complete that interaction. More reasons for not completing interactions are presented in Appendix 7.4. The number of immediate and delayed interactions was also counted (Figure 7.6). At least 40% (<22/56) of the participants responded immediately to the

conversational agent 75% (12/16 interactions) of the time, even in the last week of the study, denoting consistently high conversational agent engagement.

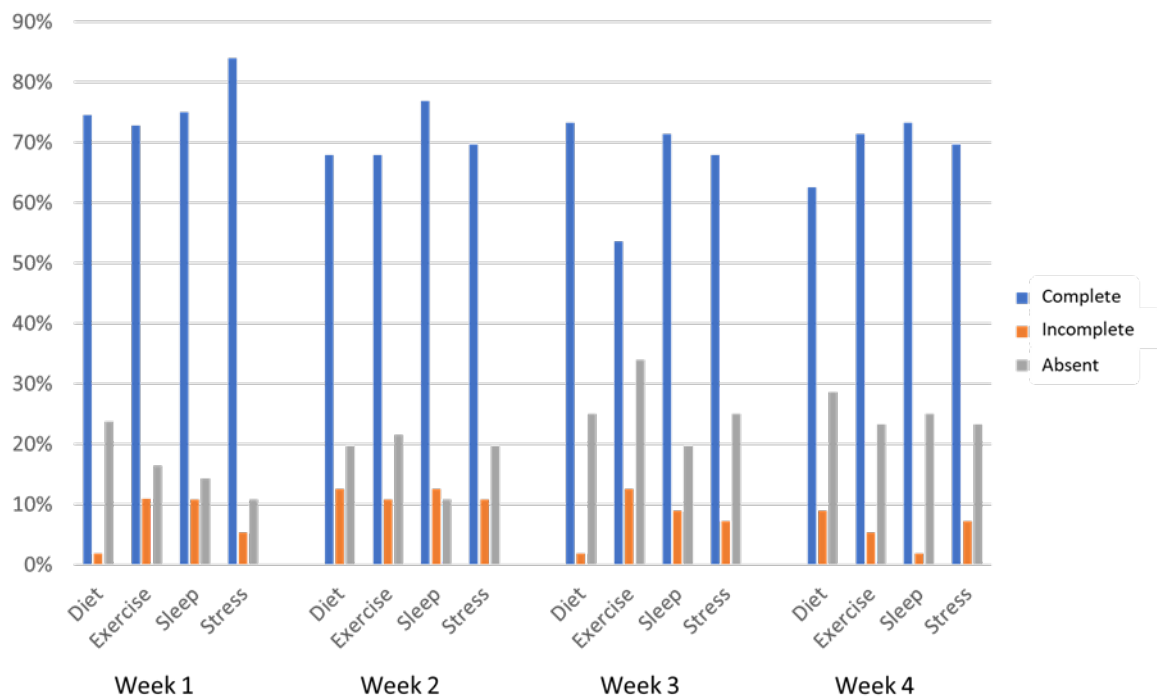


Figure 7.5 Number of complete, incomplete, and absent interactions during the 4-week feasibility study period.

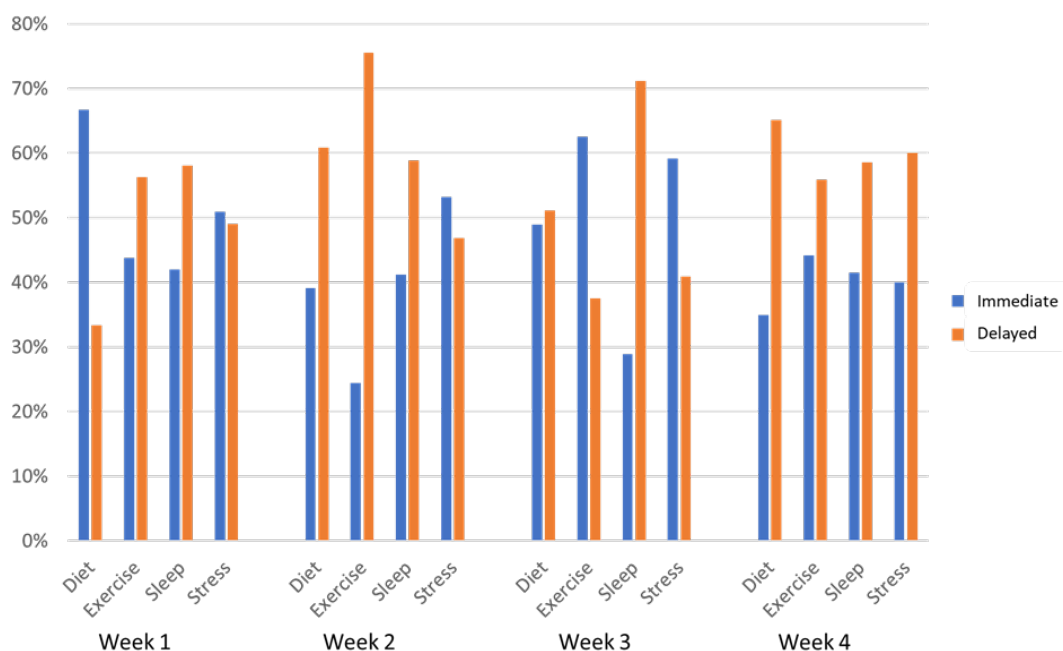


Figure 7.6 Number of immediate and delayed responses for the 16 interactions over 4 weeks with Precilla.

The average duration of each interaction was in the 2- to 5-minute range. More data on conversational agent engagement are presented in Appendix 7.4.

7.3.6 QoL Score

The 14-item QoL score was 54 (SD 6.90) at baseline and 53 (SD 6.46) at follow-up (mean difference 0.95; 95% CI -1.51 to 3.41). Of the participants taking medication (10 at baseline, 17 at follow up), participants rated adherence as fair (baseline 5/10, 50%; follow-up 6/17, 35%), good (baseline 5/10, 50%; follow-up 9/17, 53%), and very good (baseline 0/10, 0%; follow-up 2/17, 12%). The participants rated overall life satisfaction as fair (baseline 16/60, 27%; follow-up 12/56, 21%), good (baseline 35/60, 58%; follow-up 33/56, 59%), or very good (baseline 9/60, 15%; follow-up 10/56, 18%), and 2% (1/56) of participants reported poor overall life satisfaction at follow-up.

7.3.7 Knowledge

For the knowledge questionnaires used, lower scores denoted better knowledge outcomes (sections 1 and 3) and lower perceived risk of diabetes (section 2). In section 1, the scores were 7 (SD 1.37) at baseline and 6 (SD 0.93) at follow-up (mean difference 0.52, 95% CI 0.09-0.95); the possible score range was 6-18. In section 2, the scores were 28 (SD 2.7) at baseline and 27 (SD 3.5) at follow-up (mean difference 0.9, 95% CI -0.25 to 2.05); the possible score range was 12-48. In section 3, the

scores were 10 (SD 4.11) at baseline and 9 (SD 2.2) at follow-up (mean difference 1.0, 95% CI -0.23 to 2.23); the possible score range was 8-32.

7.3.8 Diet

The number of individuals consuming vegetables at least once a day was 27% (16/60) at baseline and 29% (16/56) at follow-up. Similarly, the percentage of participants having three portions of fruit per day was 3% (2/60) at baseline and 7% (4/56) at follow-up. The inclusion of unhealthy food comprised sweetened beverages and fried food or snacks in one's diet. The *almost never* option for sweetened beverages was selected 38% (23/60) of the time at baseline and 45% (25/56) at follow-up, whereas for fried food and snacks, the *almost never* consumption category was 25% (15/60) at baseline and 30% (17/56) at follow-up.

7.3.9 Physical Activity

The median METs-per-week score was 857 at baseline (IQR 902) and 765 (IQR 882) at follow-up. The average categorical score for participants remained unchanged at *moderate* intensity from baseline to follow-up. The mean time spent sitting was 439 minutes at baseline and 406 minutes at follow-up. In addition, the median amount of moderate and vigorous physical activity was the same, 30 min per week (IQR 90) at baseline and 50 min per week (IQR 90) at follow-up.

7.3.10 Sleep

According to the scoring method, a lower score indicated better sleep. The mean PSQI score was 4.38 (SD 2.36) at baseline and 4.43 (SD 2.45) at follow-up (mean difference -0.05, 95% CI -0.93 to 0.83). The prevalence of poor sleep quality was 28% (16/60) at baseline and 30% (16/56) at follow-up. The participants who reported sleeping for less than 7 hours were classified as short sleepers; 82% (49/60) were short sleepers at baseline, while 88% (49/56) were short sleepers at follow-up.

7.3.11 Stress

The scoring for stress meant that lower scores denoted lower stress levels and higher scores denoted higher stress levels. The mean score at baseline was 17 (SD 5.13), which corresponds with a moderate stress level, which was maintained at follow-up, mean 16 (SD 5.1; mean difference 0.73; 95% CI -1.15 to 2.61).

7.4 Discussion

7.4.1 Key Findings

To our knowledge, this is the first web-based pilot study to test the feasibility and acceptability of a healthy lifestyle behavior change conversational agent in the Singaporean population. In this online trial, using a Facebook Messenger CA for healthy lifestyle behavior change in Singapore, it was possible to recruit 37.5% (75/200) of the target sample size in 1 month. Retention was high at 93% (56/60); as was CA engagement, with all the conversations being completed by at least 50% (28/56) of the participants consistently. Furthermore, at least 40% (22/56) of the participants responded almost immediately 75% (12/16) of the time. Acceptability, usability, and satisfaction were also high. In general, it was possible to conduct the study with high fidelity, and each phase ensued as planned. Any definitive signs of improvement in QoL, knowledge, diet, physical activity, sleep, and stress would have to be studied in more detail in a future study of effectiveness.

7.4.2 Comparisons with Existing Literature

Conversational agents have been used for healthy behavior change; however, they have been used mainly in niche areas such as smoking cessation, alcohol misuse treatment, and the promotion of physical activity in sedentary populations [156, 157, 172]. The application presented in this study is very comprehensive, novel, and relevant to the general population in Singapore.

The outcome measures were similar to those in other studies looking at the acceptability of conversational agents in health care. Previous studies noted high response rates (ie, conversational agent engagement) and *strongly agree* or *agree* scores for user-friendliness, appropriateness, consistency, and speed of response, as in this study [150]. The other measures in prior studies that denoted acceptability were high perceived ease of use, usefulness, and intention to use, similar to the measures in this study of ease and enjoyment of use, motivating change in unhealthy habits, and intention to use Precilla again or recommend it to others [134, 135].

In another study using a health behavior change conversational agent, high compliance was attributed to a rewarding game system [135]. Considering the slight decrease in weekly compliance in this feasibility study (43/56, 77% completed interactions in week 1 vs 39/56, 69% in week 4), it may prove beneficial to introduce a gaming or reward component in the future. Personalization has previously been met with positive user reception for healthcare CA applications. Examples include providing personalized advice to maintain target blood glucose levels or personalized reminders

for taking medication [163]. Similarly, personalized content and message timing delivery should be implemented in future versions of Precilla, considering the preference indicated by users.

Past research has indicated that users prefer interacting with female conversational agents, as evidenced by Siri, Cortana, and Alexa [46]. In addition, the study by Brahnham et al. [324] explained that in the field of human-computer interaction, “the standard of believability has become inextricably linked to gender personification, especially female personification.” As such, a female character was chosen, *Precilla*, and her persona was reinforced with a profile picture.

Views on conversational agent personality have been mixed, depending on the function of the agent. For example, participants appreciated an empathetic demeanor from mental health CAs for e-therapy, in contrast to the participants’ indifference to Precilla’s humanity in this pilot study, where she played more of an impersonal, informative, or educational role [148]. Possibly, the conversational agent’s purpose determines how important the degree of humanity is.

Another aspect of this study was the effectiveness of participant recruitment through social media. A 75% (150/200) recruitment rate was selected based on other studies involving mobile health trials having achieved an application rate of 86% (70/81) for enrolment through Facebook or an 85.1% (605/711) response rate to Facebook advertisements within 45 days [325, 326]. Although this study only yielded a recruitment rate of 37.5% (75/200) in the given period, it aligns quite appropriately with the recruitment rates of 33%, 30%, or 37.7% in other web-based mobile apps or Facebook trials [327-329]. Hence, the 37.5% (75/200) recruitment rate indicates some degree of feasibility regarding the recruitment of participants for the assessment of a Facebook Messenger CA in Singapore.

A systematic review accentuated the effectiveness of Facebook as a recruitment tool [330]. It was suggested that on average, studies allowed a 3-month recruitment period, and Facebook was found to be more efficient than traditional methods (print, radio, television, email, or word of mouth) because of the reduced costs, shorter recruitment times, and ability to connect with harder-to-reach populations. In future studies, to boost recruitment, it may be worth considering advertisements, other recruitment methods, and a longer recruitment period.

7.4.3 Implications and Future Research

There was some attrition of eligible participants before the provision of informed consent. The only communication between prospective participants and the study team was through email from an

institutional email address. Perhaps, the uncertainty involved in remote participation, such as never meeting the study team in person, could have been a cause for the participants' apprehension. This lack of direct engagement may have led to some participants not feeling comfortable sharing their digital signatures because of data protection and privacy concerns. In the future, a combination of digital recruitment and face-to-face exchange (possibly in the form of recruitment or debriefing) could provide participants with an opportunity to ask questions and validate the study themselves.

Regarding preliminary efficacy, minimal improvements were observed in the participants' knowledge, stress, and diet, whereas there was a lack of improvement in QoL and sleep and no change in physical activity. The dearth of significant improvement in these areas could be attributed to the content and delivery not being adequately designed to target effectiveness and, possibly, the short study length (4 weeks), which may not have been sufficient for any noteworthy changes to have been observed. In the future, the content for the CA could be more targeted towards effectiveness and studied with an appropriate study length.

The CA engagement data revealed that there was a slight decrease in the number of interactions that the participants completed fully with each successive week. It is possible that the content and delivery methods could have come across as repetitive because a review of take-home points were shared from the previous interaction from week 2 onwards. Alternatively, perhaps the routine that Precilla followed made the intervention very regimented, not leaving much room for spontaneity in the form of varied message timings or even variations in the ways that users could respond to Precilla, for example, using free text. These factors may have contributed to the decrease in interest over time. Future iterations of Precilla could explore introducing more novelty and personalization, as recommended by the participants in the telephone interviews. The participants shared some points for improvement, including alternative modes of delivery such as *WhatsApp* and *Telegram*, a possibility of shortening the conversation lengths, and more personalized timings for message delivery.

7.4.4 Strengths and Limitations

One of the main strengths of the study is the high fidelity in the delivery of this low-cost, fully web-based feasibility study. No software malfunctions were encountered and it was possible to implement the intervention as per protocol. The content of the intervention was evidence-based and was co-designed with members of the public according to their needs as far as possible. The sample included a diverse range of age groups—from 23 to 60 years—using Facebook without the need to pay for advertising. Although the recruited sample consisted of fewer participants than the

targeted size, it is comparable with those reported in other studies that used paid Facebook advertisements within the same time frame. A further limitation is that although validated outcome measurement tools were used, all outcomes were self-reported by the participants. It should be noted that although these questionnaires have not necessarily been optimized for the Singaporean population, these well-established scales were chosen to ensure a certain level of validity and reliability in the results. Finally, this study employed a single-arm pre-post study design, which is acceptable for a feasibility study but does not allow for assessment of the effectiveness of the intervention.

Strengths and limitations of recruited sample

Using the COM-B model as a guide, participants with the capability, opportunity and motivation to perform healthy lifestyle change behaviours were recruited for this study [331, 332]. Given that they were recruited via Facebook, this population was already well-versed in the use of social media and smartphone-based apps. This could have helped to reduce any uptake barriers and provided valuable information on acceptability and satisfaction which was necessary for this pilot, feasibility trial. It may have been challenging to collect such data if the study sample was diluted with users less familiar with smartphone apps and CAs.

Furthermore, majority of the individuals were recruited from Facebook pages and groups advocating for healthier lifestyle changes. As such, the recruited participants had an inclination towards healthy living. So, in line with Michie's COM-B model, they had the capability, opportunity and motivation to change their behaviour [331, 332]. Research has shown that introducing behaviour change interventions to those willing to change the intended behaviour yielded positive outcomes at the end of the intervention and persisted beyond the period of the study [333]. Hence, the CA would likely have been better utilised and more likely to be effective with a motivated group (such as the one recruited) as opposed to an unmotivated sample.

Due to our exclusion criteria around health conditions, the older generation (>60 years of age) were excluded. The advice provided by the CA was general and applicable to the average adult aiming to adapt his/her lifestyle towards a healthier one, to avoid prediabetes or diabetes. However, the information provided was not designed to target specific populations with existing co-morbidities or those vulnerable groups who need more supervision and prescribed care. As such, this CA would not have been particularly safe or relevant to older adults, hence, their exclusion in this study was justified.

7.4.5 Conclusions

This web-based feasibility study showed that the delivery of a conversational agent for healthy lifestyle behavior change using Facebook Messenger is largely feasible in Singapore. *Precilla* was a low-cost intervention that was well received by the participants with most of them being satisfied with the intervention and prepared to recommend it to friends and family. This study demonstrated the ability to conduct a web-based trial to assess the impact of a novel intervention in a new setting. The preliminary data on the acceptability of the intervention showed the need for further enhancement of this CA, potentially through the humanization of the agent and personalization of the messaging. Such an intervention needs to be evaluated with a rigorous study design and a larger sample size.

Chapter 8 – Public’s perceptions of a healthy lifestyle change conversational agent “Precilla” in Singapore: A qualitative study.

In chapter 8, objective five of this thesis is discussed, to “understand the public’s views on a conversational agent for healthy lifestyle change and obtain their feedback on areas for improvement using semi-structured interviews in the follow-up phase of a pilot study.” The background information for this qualitative study is disclosed in Section 8.1. The methods employed in these follow-up interviews are described in Section 8.2. Under methods, participant recruitment, methods for data collection, and analysis are discussed. In a similar way to chapter 6, the participant demographics and results of the thematic analysis are presented in Section 8.3. The discussion (Section 8.4) illustrates the key findings along with the recommendations derived from this project regarding the improvement and optimization of future healthy living CAs.

8.1 Introduction

Interventions that promote healthy lifestyle changes can encourage healthier eating and increased physical activity to reverse prediabetes [312] as well as boost general health and wellbeing [334]. Additionally, research has indicated that healthy lifestyle interventions have played a role in improving mental health, which is another growing problem in Singapore with 1 in 7 having a mental health illness in their lifetime [335].

Currently, there is a dearth of evidence exploring the use of conversational agents for disease prevention or healthy lifestyle promotion in Singapore. More recent mHealth frameworks have emphasized the importance of employing a user-centered approach in intervention design. In line with this, a needs assessment was conducted (Chapter 6) to explore participants’ views on conversational agents and healthy living and used their input to design and develop the health living conversational agent Precilla which was tested in a pilot feasibility study (Chapter 7). This chapter explores the follow-up interviews conducted with participants at the end of the piloting period to obtain their views on the experience of conversing with the healthcare conversational agent, *Precilla*.

8.1.1 Objectives

The aim was to learn more about the components of the conversational agent which were well received, the methods of delivery employed which were helpful and satisfactory for the participants as well as some insights on areas for improvement.

8.2 Methods

This qualitative study was conducted in November 2019. Individuals already taking part in the feasibility study (Chapter 7) were invited to participate in semi-structured interviews and the first 20 volunteers were selected. This sample size was selected based on the recommendation that qualitative analysis should be based on data from 1-30 respondents [297]. The study was approved by the NTU (Nanyang Technological University) ethics committee (IRB-2018-11-032). All participants read a study information sheet before giving informed consent digitally. Consent was also obtained for the recording of interviews. This study is reported in line with the consolidated criteria for reporting qualitative research (COREQ) guidelines [292] (Appendix 8.1).

Details on the conversational agent *Precilla*, participants and recruitment are presented in more detail under “Methods” in Chapter 7.

8.2.1 Participants and recruitment

Participants were volunteers initially recruited for a healthy lifestyle change conversational agent feasibility study using a study poster on relevant healthy living Facebook pages and groups. They all had to complete an eligibility questionnaire. Eligible participants had to be aged 21 and above, fluent in English, and with a Facebook account. Participants went through the following events: (1) confirming eligibility, (2) baseline assessments, (3) voluntary needs assessment, (4) 4-week feasibility study, (5) follow-up questionnaires, and (6) follow-up interviews discussed in this chapter.

An email was sent by the study team to all 60 participants and interested individuals replied. The first 20 volunteers were invited to participate in interviews. Excluded individuals were those with a history of major illness, those with physical disabilities, or pregnant women as advice provided by the conversational agent was not optimized to consider diet and exercise requirements for these populations. Additionally, as the content in questionnaires, informed consent, interviews, and conversational agent were all in English, illiterate and non-writing individuals were excluded.

8.2.2 Data collection

Interviews were conducted by a female Ph.D. student (DAD) in designated private meeting rooms at Lee Kong Chian School of Medicine, NTU, Singapore. An interview guide was employed, and interviews were either in person or over the telephone depending on the volunteer's preference (Appendix 8.2). Participants met the interviewer for the first time at the time of the interview and were aware of their credentials. No relationship was established before or after the interviews. DAD was equipped with sufficient details, resources, and exposure to online courses on qualitative research and conducting telephone interviews before study commencement. Field notes were made on a physical copy of the interview guide during the interview. The initial interview guide was informed by the literature on the development of digital and conversational agent-delivered health interventions [293-295, 336]. Minor adaptations were made to the interview guide after the first five interviews, to take account of emerging themes and each interview came to an end when saturation of novel topics was reached. The interview was the final phase in a larger feasibility study in which all the participants were enrolled. This qualitative component of the study aimed to understand the components of the conversational and the methods of delivery employed which were helpful and satisfactory for the participants.

Participants were compensated with a digital voucher of \$25-35 upon feasibility study completion and there was no extra incentive for taking part in the telephone interviews. Participation was entirely voluntary.

Basic demographic data were obtained from all participants including gender, age, ethnicity, marital status, monthly household income, number and age of children, and occupation. Questions focused on their existing sources of health information, advice, and health tracking and whether this was sufficient. Participants were also asked about their experience interacting with the healthy living conversational agent "Precilla". They were prompted to comment on their impressions of Precilla (personality, ease of use, comprehensiveness, etc.), their thoughts on the delivery of the intervention (Facebook messenger, timing and frequency of messages), the effects this intervention had on their lifestyle (if any) as well as their opinions on how Precilla could be improved in the future.

Both short and long answer questions were asked. An example of a short answer question is *"at some points in the conversation you were provided with links to external webpages, did you find this useful or not so useful?"* An example of a longer answer question is *"The messages were designed to be conversational, what is your view on this? Would you have preferred it to be one-way communication, or did you like that you had some part to play?"*

All in-person interviews were recorded using a portable audio recorder for transcription. Telephone interviews were conducted via speaker phone using the departmental iPhone and also recorded using the audio recorder placed close to the speaker phone and the interviewer. The average duration of the interviews was 20 minutes. The audio files were transcribed verbatim using an external transcription service. All transcripts were checked with the corresponding sound files for accuracy, and they were subsequently corrected for errors and cleaned by DAD.

8.2.3 Data analysis

The data were analyzed using thematic analysis [296]. Firstly, two researchers familiarised themselves with the transcripts by reading them multiple times. Secondly, initial codes were proposed based on pilot coding of the first five interviews. Thirdly, themes were derived from the codes. Fourthly, the two authors discussed and combined their themes for comparison. Finally, they reached a consensus on the themes to be used and how to define them. This codebook was then used to code the remaining 15 interviews. The finalized codebook is presented in Appendix 8.3. Coding of transcripts was done using the “review” and “comments” functions on a word processor.

8.3 Results

60 participants enrolled in the conversational agent feasibility study were approached (via email) and invited 20 volunteers to participate in a telephone interview. Ten interviews were conducted in person and 10 were over the telephone at which point data saturation was reached.

8.3.1 Participant demographics

60% of participants were female, the median age of the participants was 31-35, the overall age range was 22 to 55 years and 45% were married. 65% were working, 30% were full-time students (undergraduate or postgraduate) and 1 participant was unemployed (Table 8.1). More than half of the participants (75%) were Chinese, 15% were Indian, 5% were Malay and 5% were Caucasian. The monthly household income was S\$10,000-12,000 for 30% of participants, less than S\$4000 for 25%, S\$4000-6000 for 10%, S\$6000-8000 for 10%, S\$8000-10,000 for 10%, S\$2000-3999 for 5% and more than S\$14,000 for 5%. The median monthly household income in Singapore in 2019 was S\$9425 [298].

Table 8.1 Basic demographic information of interview participants

Participant ID	Gender	Age	Ethnicity	Marital status	Monthly household income	Occupation
P01	Male	31	Chinese	Separated	<4000	Student (postgraduate)
P02	Male	25	Chinese	Single	10,000-12,000	Working
P03	Male	25	Chinese	Single	2000-3999	Working
P04	Female	49	Caucasian	Married	<4000	Student (postgraduate)
P05	Female	30	Chinese	Single	<4000	Student (postgraduate)
P06	Female	33	Indian	Married	4000-6000	Working
P07	Female	34	Chinese	Married	12,000-14,000	Working
P08	Male	51	Chinese	Married	10,000-12,000	Working
P09	Female	31	Indian	Married	6000-8000	Working
P10	Female	29	Chinese	Single	<4000	Student (postgraduate)
P11	Female	32	Chinese	Married	>14,000	Working
P12	Male	34	Chinese	Single	8000-10,000	Working
P13	Male	24	Indian	Single	6000-8000	Student (undergraduate)
P14	Female	30	Chinese	Married	10,000-12,000	Working
P15	Male	31	Chinese	Single	4000-6000	Working
P16	Female	55	Chinese	Married	10,000-12,000	Unemployed
P17	Female	24	Chinese	Single	10,000-12,000	Working
P18	Female	45	Chinese	Married	8000-10,000	Working
P19	Female	25	Malay	Single	<4000	Working
P20	Male	22	Chinese	Single	10,000-12,000	student (undergraduate)

8.3.2 Thematic Analysis

The methods of data reduction and reconstruction for thematic analysis are elaborated in more detail in Chapter 6.3.2. Five main themes were identified from thematic analysis of the interview transcripts discussing participant interactions with the conversational agent “Precilla”: (1) enjoyable and acceptable experience with Precilla, (2) sub-optimal experience(s) with Precilla, (3) comparing Precilla to other sources, (4) favorable aspects of conversational agent interventions, (5) recommendations on how to improve Precilla in future (Figure 8.1).

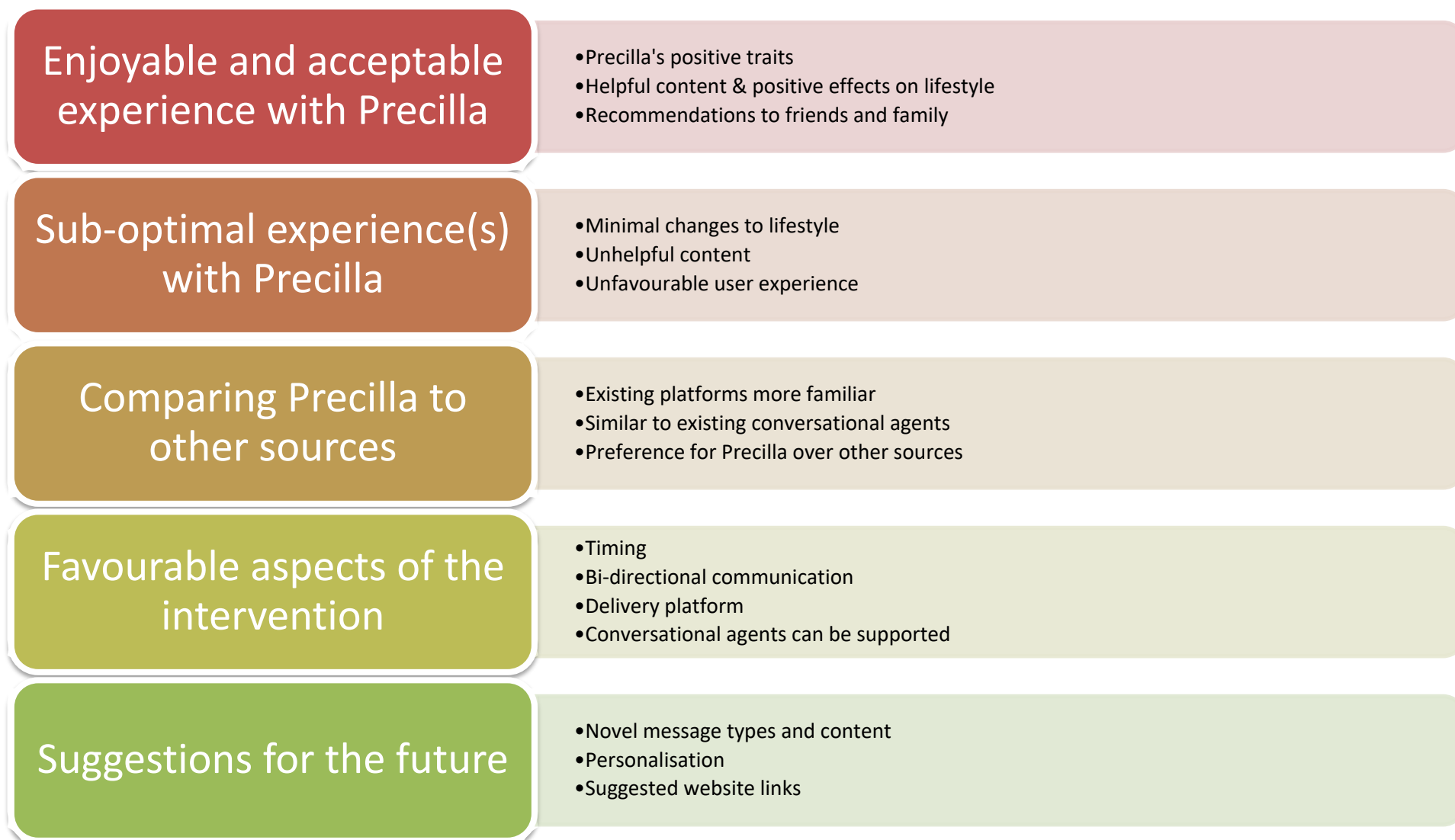


Figure 8.1 The main five themes and sub-themes identified from follow-up interview transcripts.

Each theme is presented below, along with subthemes and excerpts that highlight different facets of the theme.

(a) Theme 1: Enjoyable and acceptable experience with Precilla

(i) *Precilla's positive traits*

Many participants had a generally positive impression of Precilla. Some participants attributed this to her positive traits, such as informative content that was easy to understand.

"The content was adequate. It wasn't too much, wasn't too dense, it was in a very lay manner, so I think that bit was good." – (P01)

They also appreciated the fact that the information was always readily available anytime via the Facebook Messenger app on their smartphones.

"But what I liked about it is there is the information that she feeds, I may or may not actually read it, but I know that the information is there. If I want to, I can always refer back" – (P02)

Participants perceived Precilla and the information she shared as reliable, accurate, and informed by a team of healthcare professionals or academics.

"In terms of diabetes, I think it's a mixture of a team of people. You can have the nutritionists on the food and maybe the sport scientists about the activities, and maybe the doctors like, we need to assess how long we do this. I think it's a very multidisciplinary team, doesn't have to be a physician. I say any professionals can tackle this from different disciplines. It's like public health precisely." – (P03)

Precilla had a likable personality and used an appropriate tone when conversing.

"I liked that she was approachable. I liked the fact that the conversation was friendly." – (P04)

Participants liked the visual aids and found the user interface and images employed to be aesthetically pleasing.

"It did not become too much images, which could be a bit jarring because it's only images, and neither was it too text or too wordy, so it presented a good blend." – (P06)

Participants shared their current sources of healthcare knowledge, including health information from their field of work, advice from the Ministry of Health, and health tracking apps. Some preferred

using Precilla over these existing sources of health information because of its convenience and easily digestible content.

“Yes, I definitely preferred using Priscilla.... It was very informative, and the information is very basic and easy enough to follow without having to go through too many processes to achieve a desired result.” – (P06)

(ii) Positive lifestyle changes & helpful content

Some participants expressed their appreciation towards Precilla based on the positive effects the intervention had on their lifestyle especially in terms of sleep improvement and stress management.

“Yes, definitely. Me, personally, from all this information, especially for sleep in particular, just like the example, there were some tips and tricks that Priscilla shared. I’m actively trying to use them in my day to day because sleep, personally, is one of my issues.” – (P06)

“It did impact on me, especially the stress management. The stress management and sleep management part because I wasn’t aware of how it could better be managed.” – (P07)

The content shared was generally well-received for the four areas of healthy eating, physical activity, sleep, and stress management with some people learning new content and others having prior knowledge reinforced, like a reminder.

“I did learn some new interesting facts and information that can help with my decisions on how to improve my healthy living style.” – (P08)

“I liked the physical activity one. The sleep one was quite comprehensive also. Generally, it was good, the information was good.” – (P04)

(iii) Recommendation to friends and family

Some participants were very pleased with Precilla and indicated that they would recommend the CA to their friends and family. Their recommendations would be directed to those with sedentary lifestyles and those with a higher risk of diabetes.

“Yes, I will. For the very sedentary people. It’s a good start.” – (P07)

“I think it will serve as a good reminder for them because my parents are tech-savvy. For Precilla, maybe those who are in their late 20s to maybe their early 40s, those who are at a higher risk of getting prediabetes and diabetes serves as a good reminder for them especially.” – (P03)

"If there is different languages, yes, maybe." – (P11)

(b) Theme 2: Sub-optimal experience(s) with Precilla

(i) Minimal changes to lifestyle

A few participants reported minimal changes to their lifestyle over the intervention period. This was sometimes attributed to them having no intentions of changing their lifestyle hence, having Precilla's input had minimal effects on healthier living.

"It presented some new information, but I wasn't planning to make any changes to my lifestyle, to begin with." – (P01)

Others cited a lack of longer-term guidance by Precilla which resulted in an inability to make meaningful lifestyle changes and maintain those changes after the four-week study period.

"Not really. I don't think it's sustainable because it's just a tip for the four weeks." – (P12)

"I did try it out, I did try not drinking so much caffeine and going out of the room and then coming back. I didn't try in a consistent enough manner to actually be able to tell whether it works." – (P13)

(ii) Unhelpful content

Several participants felt like they were already well versed in healthy living and diabetes prevention or had come across similar advice in the past and found the information shared to be too simple and resultantly, did not learn anything new or valuable.

"Too simple. It differs by people. Because we are working on diabetes so we know more or less." – (P05)

"Yes, somewhat, not a lot because on the dieting and physical activity side of it it's something that I'm already doing" – (P08)

"The least useful was probably the exercise part because that information is pretty much found everywhere else, so I think wasn't useful at all." – (P12)

Furthermore, the recap sessions which were included to reinforce knowledge for better retention were often perceived as repetitive.

"I feel like sometimes I didn't really like the part where they recap because I feel like it's redundant, you're doing it again". – (P03)

Others found some of the content difficult to put into practice and indicated the need for physical instruction for some of the more unfamiliar and abstract concepts like mindfulness.

"The least helpful for me is what I mentioned, it's just informative but it doesn't really push you to put the information into real work." – (P14)

"I think mindfulness is possibly something that requires more, I guess, someone to really teach you properly. I think it's a difficult concept to grasp".

The information shared by Precilla was fairly general and this did not apply to individuals going through specific life events.

"Least helpful, I think is about sleep because I have a six-month-old so I can't sleep even if I want to. I think that wasn't useful to me." – (P09)

Some mentioned that translating Precilla into different languages relevant to Singapore or introducing more personalization, would instigate them to recommend Precilla to others.

"Not at the current stage. But when she's a bit more human I suppose and perhaps a tad more personalized, that will be great. – (P10)

"My family, no, because they don't understand English." – (P05)

(iii) Unfavorable user experience

Suboptimal interactions with Precilla were sometimes attributed to the typing delay between messages which were perceived as time-consuming. Whilst the typing function was incorporated to make Precilla seem more human-like, to the users, the delay instead seemed to contradict the concept of mHealth which is thought to be efficient and convenient.

"It comes up one line, then wait three seconds, then comes up another line, then. It's a bit annoying. Yes, or you can send multiple paragraphs, but in one chunk. I think the whole point of this space, e-learning thingy, is such that it would be convenient, like, the participants can do it at a time of their convenience. But then it takes quite a bit longer than expected." – (P01)

Others felt like there were insufficient answer options to select from and the answers provided did not always indicate the response they wished to give.

"The button design is one thing, in that there were oftentimes where the function of the button is just to review more information, instead of giving users a choice. The user intents could have been better

thought out, maybe for future iterations.” – (P15)

A handful of participants preferred using existing platforms over Precilla because of the familiarity and their ability to find what they’re for looking easily.

“In terms of finding things online, it’s something that we are habituated to doing already, so you can look for specific queries pretty quickly. With the chatbot, also because I wasn’t too familiar with the flow, but it wasn’t that easy to find that information because it felt more like I had to go through a program.” – (P15)

Others had no preference for one platform over another and perceived Precilla as equal to other conversational agents they had used in the past.

“It was fine. It’s no different from what I have seen in some standalone chatbots like Wysa or Woebot, those mental health chatbots.” – (P04)

(c) Theme 4: Favorable aspects of the intervention

(i) Timing

Timing of the messages was an important part of intervention delivery as it determined whether or not the messages would be read by participants. The consistent timing of noon was selected as a neutral lunchtime with the idea that individuals would be available to interact with Precilla at this time. However, this timing was not necessarily well-received by all as some individuals had varied lunch timings or were with colleagues and did not think it was a convenient time to be on their phones.

“No, not for me because I don’t have a set lunch break, so sometimes I didn’t get a chance to look at them until nine pm that night. – (P17)

Instead, participants suggested timings either before or after work, when they would be alone and at home where they wouldn’t have to worry about Wi-Fi access and data usage.

“Maybe morning would be a good time also when you’re starting work and you’re just about to get started, so you are trying to clear your messages on your phone or something and you can have a quick glance.” – (P09)

“And because I don’t want to waste my data, so that’s one key issue. If you do it at the timing when you’re at home, maybe it will be better, especially for those with external links.” – (P03)

Some even thought the idea of stratified timings could be interesting where messages for specific purposes were sent at specific times.

"Yes, if it's about sleep time, maybe eight or nine pm, and then we can just follow the suggestions." – (P05)

(ii) Bi-directional communication

While most participants showed a preference for more bi-directional communication when it comes to conversational agent interactions, a few participants were accepting of a more one-way-driven platform with minimal, if any, input from the user.

"I'm okay to just skim through rather than actively participate, actually. I barely reply to messages anyway even from friends and stuff. I'm okay with just a one-way conversation." "... I feel that if I were to ask a question, I can just google." – (P09)

The remaining participants shared that in the future more answer options should be provided to give the users a variety of options to select from. Some even proposed free text as a medium for users to ask questions or clarify certain doubts.

"I would have liked it if she was a bit more interactive. Because when I type stuff in, she just brings me to the main menu. That's a bit annoying. Because if that is the case, then I don't see how that's any different from me reading off a brochure or even an email." – (P10)

"Some things didn't really require a reply, because it was just like us taking in information and such, but it didn't have any outlet for if we had any questions or if we wanted to find more information for example." – (P17)

"Especially when a particular topic is of interest to me is to give me the option to find out more on that particular topic, rather than just a pre-set of the information flow and then we move on to the next topic and so on." – (P08)

Some participants did not restrict Precilla to just a health information agent but genuinely wanted to interact in a more open-ended conversation, in a more human-like manner.

"Sometimes we want to interact with it, so we don't want to follow exactly what it gives us." – (P05)

Despite these preferences, some individuals were sufficiently aware of the capabilities of various types of CAs and detected that Precilla was a rule-based conversational agent. This helped to

manage their expectations. However, they did mention a preference for future versions of Precilla to be smarter to keep users engaged and interested.

“No. I showed the chatbot to my friends and we all had the same feeling because we must follow whatever is given to us. We cannot type anything. For example, when it recommends exercise, I just ask, do you exercise, and then the chatbot said, it seems like you typed something, I don’t know whether I can understand or not, so I will direct you to the main menu. Then I will lose interest. I understand it’s rule-based, chatbot, so it’s okay.” – (P05)

(iii) Delivery platform

Participants commented on the suitability of Facebook Messenger as a delivery platform for Precilla. Some shared their acceptability of Facebook Messenger but also mentioned that when the notifications are switched off, they are unaware of when messages come in. Some individuals acknowledged that other messaging platforms may be more appropriate due to more frequent usage (especially by younger populations). For example, Instagram, Telegram, and WhatsApp were suggested along with the possibility of Precilla being a standalone app (ie. not embedded in any pre-existing platform).

“Probably Telegram or WhatsApp would have been a better option because, for instance, me, personally, if I did not actually log in or switch on my notifications for Facebook Messenger, I would miss out that Precilla has an update for me on that particular day.” – (P06)

“Well, for me, I think Facebook is fine, because I’m primarily on Facebook, but the younger people seem to be more Instagramming.” – (P01)

(iv) Blending conversational agents with other platforms

Some participants indicated that the information shared by Precilla was quite dense and may be more suited to an online course. In this way, Precilla could be administered in combination with an online page or course. Resultantly, participants could engage with the conversational agent and receive reminders whilst also learning the necessary content of interest from the course.

“If that was the case, wouldn’t it be easier to just have it as a wiki page, or in a more, quote, unquote, traditional online course format rather than someone I have to converse with.” – (P15)

Others shared a preference for a combination of Precilla with in-person consultations with healthcare professionals for more interactivity and perhaps reassurance from a human HCP.

"Because I had done another research probably a year ago, one that actually has got a dietician. A one-off, they have a bot also, but separately they also have a dietician so the chatting was more interactive."
– (P18)

(d) Theme 5: Suggestions for the future

(i) Different message types

Participants indicated that including different modes of communicating such as voice or video messages for future iterations of Precilla would introduce some novelty besides just text messages.

"Maybe a voice. We always find that listening is easier than just reading the text." – (P05)

"It's like when you try to find things on how to do certain exercises. When they have either drawings or when it's just descriptive, it's usually a little bit hard. I think what can be improved possibly is videos." – (P10)

They also pointed out a need for a better default or error message should the user type in something which the conversational agent could not recognize. This should be something that clarifies the limits of the conversational agent and points the user towards help.

"I'm not saying that she should be able to answer that, but it would be nice to at least have an answer that's at least Siri-like that says, I do not have the authority to answer that, or something like that." – (P10)

"On a more basic level, to go back on the main menu, like, because when I tried typing like main menu and stuff, there isn't a command, or at least to my knowledge, in using the chatbot. So the only way to trigger Precilla to take you back to the main menu was to type something wrong or type anything, so that was slightly annoying because you didn't know." – (P15)

(ii) Personalization

The content provided by Precilla was fairly general and could be more targeted towards specific age groups. For example, the advice to younger populations may be more towards cultivating healthier eating habits whilst the advice for older populations may be more targeted towards age-specific diabetes prevention techniques.

"Missing? I'm not sure whether there was age-related content for different age groups." "And also different age groups have different requirements." – (P09)

For those individuals already well versed in healthy eating and physical activity, they found the information provided to be too simple and requested options to gain more in-depth information and details on topics of interest to them.

"I think the app can be more specialized because it's quite a wide topic on general health. So let's say you want to delve into a topic, you can actually choose it and it will be more better for people who want to know more about certain topics, not just everything in general." – (P12)

A few participants proposed having a search function to recall specific information in the conversation flows efficiently. Additionally, follow-up notifications would also be appreciated for unread messages so that they can pick up the conversation at a more convenient time and don't completely miss out on interactions if they are unavailable to attend to Precilla immediately.

"But I'm not sure if there is actually a search function where you can look through the conversation that you had. Because she sends quite a lot of messages, sometimes you want to look back into certain things that she mentioned before. I think it's much of a hassle to really scroll up through the conversation because it can be quite lengthy. Maybe, I'm not sure, if it's possible to implement a search function in the Facebook chat." – (P02)

Finally, some participants mentioned that their likelihood of recommending Precilla to their friends and family would be elevated if the content was translated into the other languages relevant to Singapore (i.e. Tamil, Malay, Chinese).

"But the only thing I suppose that would be of concern is if let's say you are looking to broaden your target audience, then you might have to send it in other languages like Chinese and Malay. Because there are definitely, especially the older generations, they are not going to be able to read English." – (P10)

(iii) Suggestions for website links

During the intervention, participants were provided with links to some web pages where they could access further information on a topic area or links for relaxing music to play alongside their mindfulness practice. The usage of these links was not found to be very popular as evidenced by some participants not clicking the links or some unable to recall such an exchange. Participants provided suggestions on the types of links they'd be more interested in such as links to helplines, educational talks, courses to attend, and links to more in-depth information.

“For me, it’s more like links to helplines (...) or maybe there are talks or courses that are around in some medical institution, in the hospitals, poly clinics, anything like that. So, then I know if I want to learn more about stress management somebody is giving a talk somewhere.” – (P07)

(iv) Novel, additional content

The intervention lasted for four weeks, and some participants mentioned they’d be open to a longer duration of interactions with Precilla, provided that novel content was introduced. Novel content could include more educational information on prediabetes (since diabetes is more commonly talked about), signs and symptoms of prediabetes and diabetes, context or race-specific food recommendations in Singapore, and more clarity on the difference between healthy foods and “healthier options”.

“They promote like, this milk tea is a healthier choice, but it’s not healthy. This difference between healthier choice and whether something is healthier or not, it should be the emphasis for the group who are really educated”. – (P03)

“Yes, especially since I think Singapore, the bigger problem for diabetes, in particular, is, e.g., for Indians, so, yes, it would be good to... It can bring in race-specific food.” – (P13)

“I don’t recall, was there advice to look up what the symptoms are if you’re prediabetic?” The signs or symptoms to look up for.” – (P18)

8.3.3 Trustworthiness in qualitative research

The quality and trustworthiness of this study was assessed using the credibility, transferability, dependability and confirmability criteria [301, 302]. Please refer to Chapter 10.9 for a detailed account of the quality assessment and the reflexivity account for these qualitative studies presented in this thesis (Chapter 6 and 8)

8.4 Discussion

8.4.1 Key Findings

This qualitative study was a sub-study of an online feasibility trial conducted on a CA delivered healthy lifestyle change intervention. Twenty participants who volunteered to participate in interviews were recruited from the online trial. They were interviewed about their perceptions of “Precilla”, shared their experience of interacting with a Facebook Messenger delivered healthy living and diabetes prevention conversational agent, and provided recommendations for the future. Enjoyable experiences were denoted by Precilla’s friendly personality, informal tone, and helpful content which encouraged some positive changes to their lifestyle. Others were less satisfied with Precilla and found her content to be too simple and delivery sub-optimal due to the lengthy messages and insufficient predetermined answer options. When compared with existing sources of information, Precilla was comparably not as “intelligent” as users anticipated her to be. In other circumstances, she was thought to be more interactive and engaging than existing apps or sources participants had access to. Timing of messages, sufficient bi-directional interaction, and selecting an appropriate delivery platform were important considerations for the success of the intervention. Finally, for the future, participants suggested a preference for other messaging formats (video, image, speech, etc.), more personalized content and delivery, links to helplines or more in-depth information on signs and symptoms for diabetes and prediabetes, and specific healthy food recommendations relevant to and accessible in Singapore.

8.4.2 Comparisons with existing literature

A study in Singapore looking at the health education and communication of individuals with prediabetes indicated that there was a preference for messages on understanding risks associated with diabetes and how to prevent diabetes, the relationship between health and family and to avoid using the term “prediabetes” in messages [304]. In this study, participants also indicated an interest in understanding the signs and symptoms of prediabetes and diabetes and wanted to learn more about prediabetes (as diabetes is more commonly talked about) but made no comments regarding a dislike for the term “prediabetes”. Perhaps, since the individuals in Lim et al. were already diagnosed with prediabetes, their aversion to the disease may be fuelled by the term reinstating their diagnosis [304]. Since participants in this study were healthy, no such aversion was noted. The study also concluded that healthcare professionals at polyclinics could provide preliminary advice before directing patients to community-based programs [304]. A similar proposition was suggested by some of the participants during the interviews where they expressed an interest to receive updates and

details on where to attend community-based programs or educational talks on diabetes prevention and healthy living from Precilla.

Zhang et al. presented a timely review on artificially intelligent conversational agents for healthy behavior change, focussing chiefly on healthy eating and physical activity [220]. This reiterates the point highlighted by participants in this study that plenty of existing interventions and sources of information focus on diet and physical activity. Hence, the inclusion of additional elements namely, sleep advice and stress management were well-received for their novelty and distinction from existing health behavior change programs.

Some participants in this study faced some challenges in navigating the app in terms of returning to the main menu, asking for help, and found the default error message insufficient. This could be addressed by providing a link to a user guide as implemented in another study using a smartphone app to promote self-management of diabetes [337]. A user guide may help resolve some of the concerns raised by participants by outlining the aims and limitations of the conversational agent, additional relevant resources, or a “contact us” option to reach relevant personnel for help.

8.4.3 Main recommendations

The main recommendations on how to further improve a conversational agent such as Precilla for healthy living are outlined in Figure 8.2. These involve selecting appropriate message timings, involving participants in answer option generation, selecting an acceptable delivery platform based on target population usage, and finally, engagement strategies to keep users entertained and interested in future conversations.

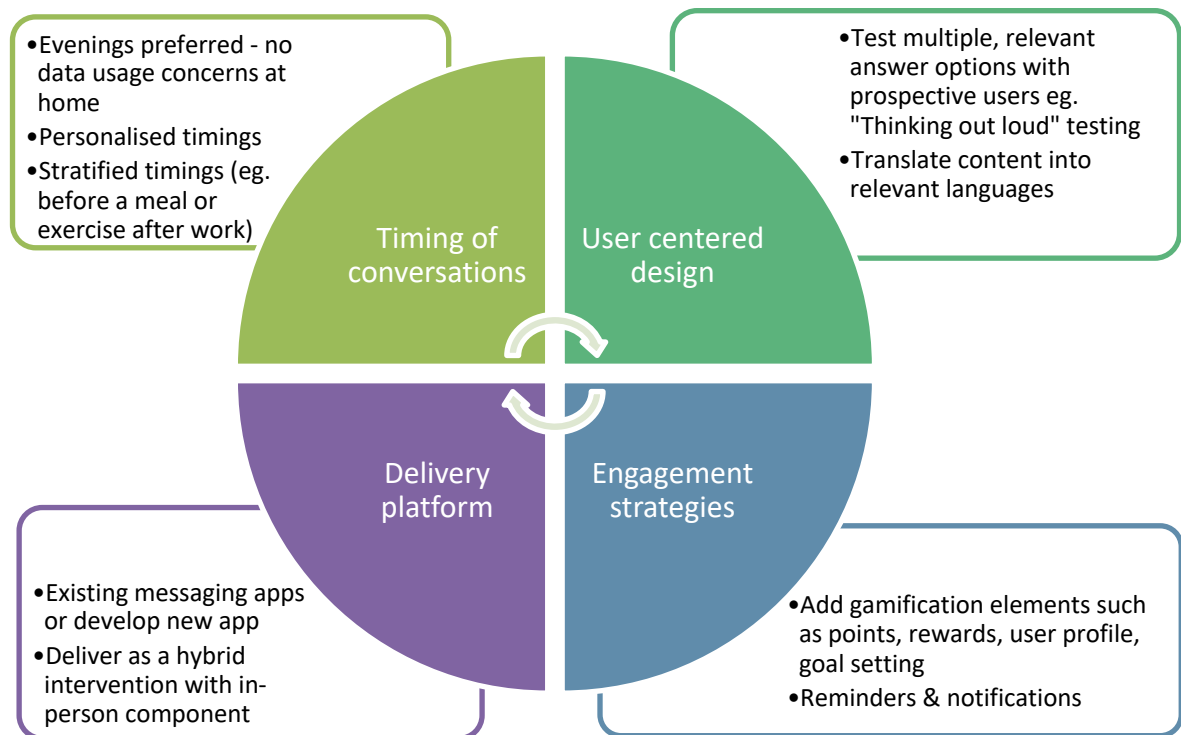


Figure 8.2 Main recommendations on how to enhance a healthy living conversational agent as suggested by study participants.

Message delivery timing is an important determinant of CA usage by participants. It would be preferable to set a timing that is in the evenings when individuals are likely to be at home (no data usage concerns) and available to interact with the CA. Message timings could be streamlined further to be personalized based on each individual's needs and preferences. For example, some individuals may wish to receive message prompts before dinner every day, to remind them to have a healthy meal. Others may prefer an interaction at the end of the workday which encourages them to use their free time to get some exercise. Having a reminder function or follow-up notification to prompt them if the message has yet to be opened can be an effective engagement strategy.

Participants want to be more involved in the conversation. Whilst free-text is a challenge for rule-based agents, providing more options for them to choose from would be helpful. Participants indicated a need to better map out more user intents to make the options more applicable and having multiple choices instills some autonomy. This is where co-design could be helpful. Different answer options could be posed to prospective users in a "thinking-aloud" testing setting where they can share their thoughts and opinions in real-time of their acceptability of the responses generated by the research group [274, 338]. Additionally, translating the conversational content into languages of

interest for the target population can be effective. This would enable a greater reach to wider age and ethnic groups who may benefit from the conversational agent but are currently restricted by a language barrier.

Conversational agents could be delivered as part of a hybrid healthcare delivery program. Having an in-person and mobile health care program may be an effective means of transitioning patients away from just traditional care and more accepting of technological initiatives. Having healthcare professionals linked to one's CA can make them more accessible to patients, can help keep track of behaviors, and potentially reduce hospital visits or consultation times.

A conversational agent, not embedded in Facebook Messenger, could be better received, especially by individuals who are not avid users of Messenger. Precilla is a rule-based conversational agent, designed and developed by non-experts to reiterate the accessibility of such digital health interventions for researchers. For such non-developers, plenty of chatbot creation software exists such as Instabot or Collect.chat [236]. These can be employed to create a conversational agent that can be embedded in a website, for example. Alternatively, a conversational agent could be delivered through a standalone app. This could involve hiring a programmer, collaborating with an app development company, or using an app builder such as "Appy Pie" or "App Sheet". Most of these builders charge a monthly subscription fee [237].

Participants indicated an interest to be "entertained" in their experience of engaging with a conversational agent. Also, allowing users to sign up with a username and password, profile image, ability to collect points may be appreciated. A similar approach was adopted by Baptista et al. who incorporated simple gamification elements, including goal setting, progress monitoring, feedback, and quizzes within their embodied conversational agent for diabetes management [337]. User authentication via username and password could also promote security and eliminate the need to disclose their phone numbers – which some participants noted as a privacy issue.

8.4.4 Strengths and Limitations

All interviews were conducted by the same interviewer to enhance reliability in terms of consistency of questioning style and quality of interactions with interviewees. Interviews were conducted in private meeting rooms within the Lee Kong Chian School of Medicine Clinical Sciences Building to ensure the confidentiality of participants. Transcripts were analyzed thematically and checked and reviewed in parallel by two reviewers (to neutralize the subjective nature of the analysis) who discussed and came to a consensus and justified the themes to be used. Although 20 participants was

an adequate number for this qualitative analysis, it would be beneficial to aim for a larger sample size in future studies to increase the generalizability of the findings.

The sample may have been biased as the individuals recruited into the online feasibility trial were from existing healthy living and lifestyle change Facebook groups and so already had a keen interest in healthy living and may have had higher levels of health and technological literacy than the general population in Singapore. The interviews were conducted exclusively in English, and thus, would have excluded the opinions of individuals from different generations who are possibly not as fluent in English but may greatly benefit from mobile health usage. Future work should consider translating conversational agent content and conducting interviews in the languages relevant to Singapore (Malay, Tamil, and Chinese).

8.4.5 Conclusions

This healthy lifestyle change conversational agent, Precilla, was generally well-received. Her content and delivery were commended with some individuals reporting positive lifestyle changes and an appreciation for her kind demeanor and lay language. Some participants who experienced less satisfaction with the interactions provided suggestions on how the scheduling of messages could be altered to maximize engagement and how a variety of answer options could promote personalization and enhance acceptability by users. Future work should consider translating Precilla into other languages applicable to Singapore to reach older generations, providing users with novel content on prediabetes, employing more messaging options (such as speech or videos), and directing users to external sources to support their interests and improve their health literacy.

Chapter 9 – Systematic assessment of healthy lifestyle change conversational agent apps

This chapter describes the study relating to objective 6, to “identify and describe the main characteristics and evaluate the reliability, transparency, and BCTs used in healthy lifestyle change conversational agents through a systematic assessment of apps.” Section 9.1 provides some background information and highlights the objective of the project. Section 9.2 outlines the methods and the various characteristics that were assessed. Section 9.3 presents the results and Section 9.4 is the discussion where a summary of key findings and implications of the results are mentioned together with strengths and limitations of the project and methodology.

9.1 Introduction

CAs are being implemented to deliver interventions for disease treatment, management, and prevention [339]. An important component in such interventions nurturing prevention over treatment is lifestyle changes. This was noted as an effective method of preventing the onset of diabetes in the Finnish Diabetes Prevention Study where intense lifestyle modifications had positive effects on diet, physical activity, and diabetes risk [312]. Such lifestyle modifications have been implemented using digital techniques such as text messaging and have shown success in general preventative health behaviors and specific situations like smoking cessation [14, 340]. CAs can provide more to the patient and healthcare provider with its added functionalities such as low/no-cost, easy accessibility from the smartphone and widespread Wi-Fi access, bidirectional communication to boost engagement and user involvement. These factors make CAs a more attractive option compared to existing remote methods of digital health interventions for healthy lifestyle behaviour change.

Behavioral change interventions can be defined as “coordinated sets of activities designed to change specified behavior patterns” [341]. The majority of behavior change interventions are complex and include a combination of one or several behavioral change techniques (BCTs) [342]. A BCT is “an observable and replicable component designed to change behavior” [342]. Numerous BCT taxonomies currently exist, however, the BCT Taxonomy v1 (BCTTv1) [101] is widely used to provide standardized terminology to indicate the various facets to support behavior change in health care interventions. BCTTv1 consists of 93 distinct BCTs divided into 16 major groups such as “goals and planning” and “feedback and monitoring” [101].

A scoping review previously conducted (Chapter 4) revealed that the majority of CAs are delivered through smartphone apps [42]. Furthermore, the qualitative analysis conducted with participants in Chapters 6 and 8 also revealed that CA users who do not avidly use social media apps, like Facebook Messenger, would prefer to use health CAs which were standalone apps. Furthermore, the scoping review was conducted at the beginning of the PhD project and since then, healthcare CA development and implementation has accelerated at lightning speed. Although several reviews have been published to scope existing health CAs in the literature [42], and dissect their applications and evaluations [41, 116, 119], an up-to-date assessment of CA apps freely available for the general public to download for healthcare behavior change, is currently lacking.

Moreover, exploration of the BCTs implemented for healthy behaviour change in existing CA apps could help inform the techniques which can be integrated into Precilla's platform (Chapter 7) to maximize its potential for effectiveness in the future.

9.1.1 Objectives

This project aimed to identify and describe the main characteristics of healthy lifestyle change CA apps in the app stores (Apple's App Store and Google Play) and evaluate them based on reliability, transparency, and use of BCTs. A further aim was to make comparisons between the identified healthy living CA apps and CA Precilla to indicate areas for improvement in subsequent iterations.

9.2 Methods

This cross-sectional study explored all CA-based apps targeting healthy lifestyle behavior change. The study emulated a systematic review methodology adapted for app assessments as presented in previous systematic assessments of healthcare apps [96, 97, 99]. The process involved a comprehensive search for relevant apps, screening, and selection based on pre-determined limits and assessment based on relevant criteria. The apps were searched for using the 42 matters database in December 2021, under the categories health and fitness, lifestyle and medical. Apps were identified using search terms which were various synonyms of "conversational agent" and "chatbot" identified from a scoping review (Chapter 4) with additional terms identified from keywords used by relevant apps. A comprehensive list of the search terms is presented in Appendix 9.1. Apps were first screened for inclusion based on app store descriptions. Suitable apps were downloaded and tested to enable more extensive data extraction. Apps that appeared in both app stores were identified, noted, and assessed in the iOS store.

9.2.1 Inclusion and exclusion criteria

Only apps available in English were included. Standalone CAs, as well as multifunctional healthcare apps with a CA component, were assessed. Included apps demonstrated some degree of behavior change, defined as “coordinated sets of activities designed to change specified behavior patterns” [341]. The behavior change interventions needed to be targeted at a healthier lifestyle change. The target users were individuals interested in changing their lifestyle either for general wellbeing or for disease prevention. These included patients, healthcare professionals, caregivers, etc. Both embodied (with avatar) and non-embodied agents were considered. Embodied conversational agents (ECAs) were cartoon-like avatars that presented similarly to humans in face-to-face conversations [343]. They were able to emulate and respond to verbal (text/speech) and non-verbal (body language/facial expressions) communication [343].

General health apps with no behavior change component were excluded, along with symptom checkers, or surveys with no conversational component. Apps that only mediated teleconsultation services between the user and a healthcare expert without a CA to converse with were not considered. Apps designed purely for educational purposes were excluded as well as apps targeting self-improvement exclusively.

9.2.2 Assessment criteria

The app selection and identification were performed using a systematic approach to minimize bias. This involved extracting data guided by a pre-defined data extraction sheet and two reviewers working in parallel to screen apps and extract data. These assessment criteria for data extraction were decided based on observations made in the scoping review, conceptual framework, and other systematic apps assessments [42, 97]. A detailed outline of the data extraction sheet is presented in Appendix 9.2.

(a) General app characteristics

General app characteristics included: app and developer name, platform (iOS/Android), user ratings, cost, language, accessibility, target users, and target health condition. Other general attributes explored whether users were able to create a profile for themselves, if the account was password protected, if gamification was incorporated, if there were any usability or technical issues, if the app had adverts and if these are disruptive, if the app was created by healthcare providers, if data could be exported and any other additional information which was relevant.

(b) CA specific characteristics

CA-specific information was also extracted. These included an assessment of the intelligence of the CA (rule-based/AI), input and output modalities, default messages, personality and appearance of the avatar (if any), health tracking functions, safety netting, educational information, human involvement, and whether the information provided was evidence-based.

(c) Mobile applications-Health On the Net (mHON) code principles

The apps were also assessed according to the updated mHON principles which, when complied with, enhance the reliability and transparency of the app [100]. The principles were concerned with authority, complementarity, confidentiality, the validity of the information provided, the justifiability of the content, user's practice, financial disclosure, and the app's advertising policy [100].

(d) Behavior change metrics

The behavior change techniques employed by the included apps were assessed using Michie's BCT taxonomy BCTTv1 [101]. The BCTs observed in the apps were first categorized according to the 16 main categories stated in the BCT taxonomy namely: goals and planning, feedback and monitoring, social support, shaping knowledge, natural consequences, comparison of behavior, associations, repetition, and substitution, comparison of outcomes, reward and threat, regulation, antecedents, identity, scheduled consequences, self-belief and covert learning [101]. Subsequently, individual BCTs under the main categories were assigned for greater specificity. The selection of BCTs was done by both reviewers individually and they later combined their results and reached a consensus on the best BCTs to allocate to each app.

9.2.3 Data analysis

The data was presented mainly narratively with some descriptive statistics. App searching, screening, and data extraction were all performed in parallel by two researchers. The data extraction sheet was first piloted with two apps between two reviewers. Once a consensus was reached on the best way to present the extracted information and the sheet was optimized, an agreement was reached on how to log extracted data and this format was applied to all the remaining apps for consistency and reliability. Any discrepancies in data extracted and app assessments were discussed between the two researchers until a consensus was reached.

9.3 Results

9.3.1 App selection

The flowchart in Figure 9.1 displays the app selection process. The initial search of iOS and Android app stores using the 26 search terms (Appendix 9.1) yielded 2898 and 2573 results, respectively. These apps were screened for inclusion and 5326 were found to meet the exclusion criteria as they were not related to healthy living. This resulted in 62 (iOS) and 173 (android) relevant apps to be further assessed. App descriptions were studied in more detail for relevance to healthy living, prospective apps were downloaded and tested for relevance, and duplicates were identified and removed. Some reasons for exclusion were that 58.7% (134/228) of the apps targeted a specific health condition like asthma or cancer patients and not healthy living for the general population. In other cases, an invitation was required to use the app and join the program (4.8%, 11/228). For some apps, once downloaded it was found that there was no chatbot component (1.3%, 3/228) or the app was not in English (0.9%, 2/228). Finally, although the 42 matters database identified certain apps, when searched for in the iOS and Android stores, the search yielded no results indicating the apps were no longer in existence (3.5%, 8/228). After excluding these apps, there was a final pool of seven CA apps applicable to healthy lifestyle behavior change.

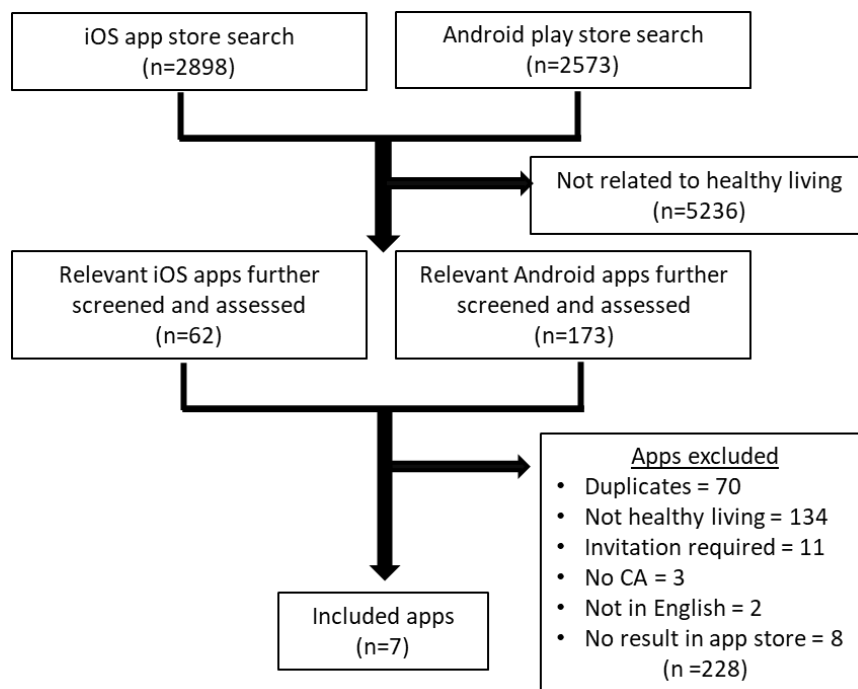


Figure 9.1 Flowchart for systematic app selection in Android and iOS stores

9.3.2 General app characteristics

Five of the apps were assessed on an android smartphone and two were downloadable from an iPhone 12 (Table 9.1). Five apps were free to download and use. The remaining two were also free to download but included in-app purchases which users could purchase for additional features. All seven apps had not received enough reviews and ratings to be displayed. All the relevant CAs used text (n=6) and speech (n=1) to communicate with the user and none were ECAs. The age suitability was rated as 4+ for all apps and the target populations comprised individuals interested in developing a healthier lifestyle for diabetes prevention (n=1), through a healthier diet (n=2), more physical activity (n=3), and improving sleep hygiene (n=1). The included apps were either rule-based (n=3), employed AI (n=2), or used a combination of both (n=2).

Table 9.1 Characteristics of included CA apps from 42 matters database search

Characteristics	Number of apps (% of total)
General app characteristics	
Store:	
Android	5 (71%)
iOS	2 (29%)
Cost:	
Free	5 (71%)
In-app purchases	2 (29%)
Aspect(s) of healthy living covered:	
Diet	2 (29%)
Exercise	3 (43%)
Sleep improvements	1 (14%)
Diabetes prevention	1 (14%)
CA specific information	
Intelligence of CA:	
Rule based	3 (43%)
AI	2 (29%)
Both	2 (29%)
Input /Output	
Text/text	6 (86%)
Speech/speech	1 (14%)

In three apps, the CA asked questions about the user's health status to provide personalized content. The questions culminated in a diabetes risk score or user's BMI and blood sugar levels. Having the user's health history meant that content and educational information could be tailored to their needs and risk factors and in one app, a personalized diet plan could be generated. Personalization in the form of tailored timing of messages from the CA was not an option offered by any of the apps included. In contrast, in three apps, the CA only interacted when the user initiated the conversation, not the other way around. Health tracking was a feature in five apps. External apps or devices like "Strava"

[344] or smartwatches could be linked to the app for physical activity tracking. For sleep, tracking was done in the form of a sleep diary [345]. Tracking of diabetes risk could also be done using the Indian diabetes risk score in one app [346].

References, for the information provided in the app, were only cited in one app, *“Sleep space-doctor approved”* [345]. Two apps mentioned the credentials of developers. *“My Dietitian, BMI, Diet Tips, Health, Food Search”* was developed as part of a Ph.D. research project, supervised by an associate professor [347]. Similarly, *“RunMotion coach”* was based on advice from experts in sport psychology and international runners [348]. No sources of evidence or information were cited by the remaining five apps. Human involvement was available through a sleep coach in the *“Sleep space-doctor approved”* app if requested by the user and this was subject to additional charges [345]. In *“PAWEN - Running app”*, the CA was not always able to respond to the user’s questions and would ask for some time to respond [349]. It could be inferred that the question was reverted to a human administrator who would offer an answer to the user’s query instead of the CA, but this was not explicitly mentioned in the app.

9.3.3 Individual app analysis of personality and BCTs

A total of 20 different BCTs were identified throughout the seven healthy lifestyle change CA apps included (Table 9.2). The individual BCTs and their frequencies are presented in Figure 9.2.

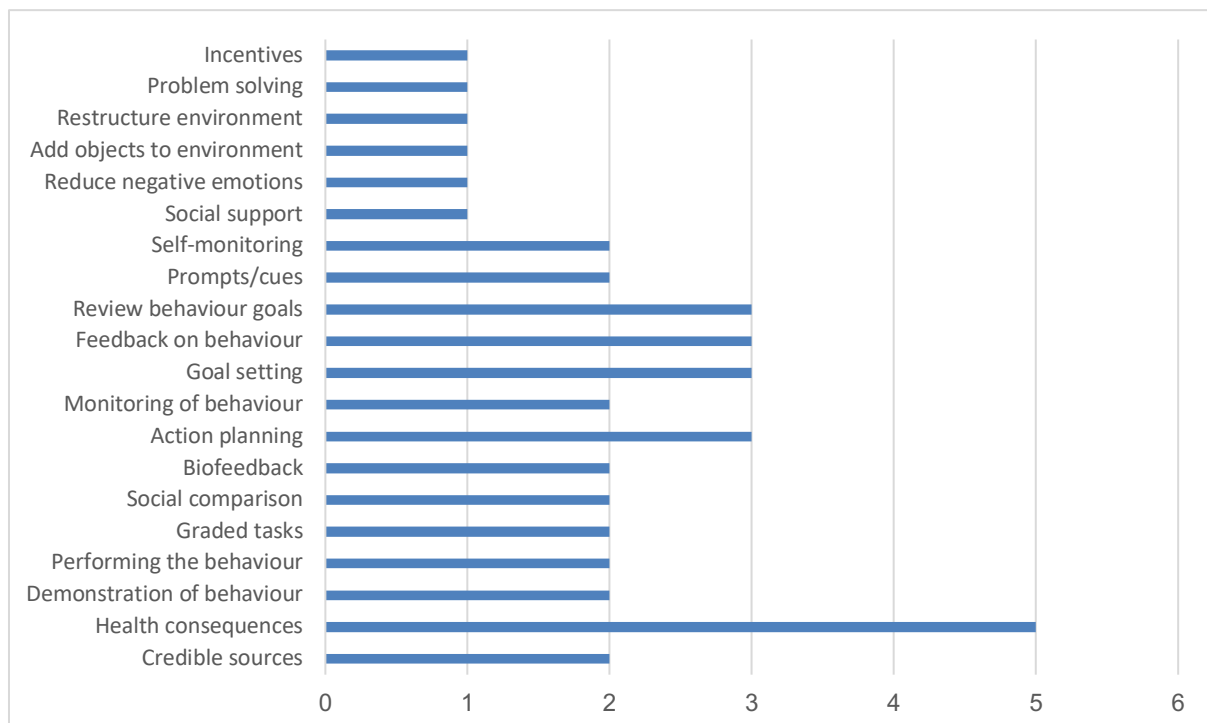


Figure 9.2 The 20 BCTs used (based on Michie’s BCT taxonomy [101]) and associated frequencies in the included apps

The “*BUD.D*” app aimed to promote diabetes prevention with a CA that was knowledgeable and factual [346]. The CA was very well organized and comprehensive with clear limits. In other words, there was no option for the user to type free-text as they could only select options provided by the CA hence, the conversation was well controlled. The factual nature of the CA meant it was quite rigid, unlike a more free and easy chat with a human. The information was provided in the form of articles that were scientific and objective. The app employed several BCTs: (1) information was from credible sources including published research, (2) information on health consequences was shared in the form of scientific articles, (3) healthy behaviors were demonstrated with model recipes, (4) graded tasks were suggested in the form of phases to transition towards healthier eating and weight loss and (5) social comparisons were made by indicating the recommended levels of physical activity for adults of the same age.

“*Briea-Wellness Coach by Hygiea*” was an app that supports physical activity and wellness for healthier living [350]. Like “*BUD.D*”, the CA had a rigid and factual nature but was also HCP-like as it supported the user in maintaining a healthy BMI. Nine BCTs were targeted in this app. Biofeedback was employed by relaying the user’s blood pressure readings to remind them to make healthier choices. Action planning was in the form of preparing for the next meal in advance. Healthy behaviors were demonstrated, such as exercises for physical activity sessions and monitored, without explicit feedback. Users were encouraged to set goals for the healthier lifestyle they envisioned, they also received feedback for their behaviors (eg. Exercise) and associated outcomes (eg. Weight loss) resulting in recommendations to review and revise their goals accordingly. The app also had a reminder function which acted as prompts/cues for healthy behaviors based on their health calendar.

“*My Dietitian, BMI, Diet Tips, Health, Food Search*” was focused on healthier eating [347]. The CA was factual and culture-specific as the food recommendations were targeted towards the Indian population. The CA provided information on how to perform the behavior with health tips and diet plans. Users’ behavior and outcomes were monitored sometimes with/without feedback. Diet plans could be set up in advance (action planning) and users could enquire about health consequences based on the food they had eaten.

Users were enlightened with a CA sharing nutritional information for a healthy diet in the “*NutriHealth - Nutrition Counselling Chat-bot*” app [351]. Information on positive health consequences (such as diet management) and social comparisons (such as recommended intake based on age groups) were the main BCTs employed. The information provided was comprehensive as the CA was knowledgeable and factual.

The “*PAWEN - Running app*” was for users interested specifically in running as a form of exercise towards healthier living [349]. The CA here was knowledgeable and coach-like, so it supported the user in goal setting, action planning and even provided social support. Users were encouraged to be independent using self-monitoring and educated on many items including how to add necessary objects to their environment to promote their running goals, how to perform a behavior (such as understanding the right pace for a run of a certain distance), how to reduce negative emotions for more success in running and information on positive health consequences (such as reduced anxiety and increased body awareness).

“*RunMotion coach*” was also an app targeting physical activity in the form of running for healthier living [348]. It targeted a range of users from those wanting to run for health wellness to those wishing to train for a marathon. The CA was friendly and coach-like, setting the user graded tasks with an adaptive training plan, helping with goal setting based on the user’s intentions, and reviewing their behavioral goals weekly. The app provided biofeedback in the form of training paces based on past runs and target timings. Users could also self-monitor their runs by syncing the app to a smartwatch or “*Strava*” app [344]. The app offered users opportunities to join individual or groups challenges and these were incentivized with badges that could be earned.

Users of the “*Sleep space-doctor approved*” app were interested in improving their sleep hygiene [345]. The CA was HCP-like and knowledgeable, providing solutions to problems, information on health consequences, and how to restructure the physical environment for better sleep hygiene from credible sources including scientists and doctors. Users also received feedback on their behaviors in the form of sleep scoring and sleep coaching and were prompted in the form of reminders to fill out the sleep diary or received cues that played the right sounds to deepen sleep cycles [345].

Table 9.2 The BCTs and personalities used by each app and CA to achieve the healthy living aim

App name	Healthy living aim	Personality	BCTs used: Evidence of BCT (quotes)
BUD.D [346]	Diabetes prevention	Knowledgeable, factual	Credible sources: <i>"This super intelligent ChatBot will answer all your questions on diabetes prevention supported by published research insights"</i> Information on health consequences: [CA] <i>"I have found some articles and videos on this topic. Happy reading"</i> Demonstration of the behavior: <i>"protein shakes to replace meals. Example recipes for morning, afternoon, and evening shakes."</i> Graded tasks: <i>"This diet plan focuses on changing your overall eating habits to encourage you towards weight loss and a risk-free lifestyle. 1st phase - starting phase, 2nd, 3rd..."</i> Social comparison: <i>"recommended daily physical activity for adults."</i>
Briea-Wellness Coach by Hygiea [350]	Physical activity and wellness	Factual (quite rigid), HCP-like	Biofeedback: Inform the person of their (blood pressure) reading to improve adoption of health behaviors Action planning: The user can ask - <i>"What is my dinner for today?"</i> Demonstration of the behavior: <i>"Provide me a video of leg exercise."</i> Monitoring behavior without feedback: <i>"As you move on with Briea, you would wish to keep on sharing several updates about health indicators."</i> Goal setting: <i>"Creating Goals is the best way to enforce lifestyle improvement. Briea helps for creating goals,"</i> Feedback on outcomes: <i>"Briea helps for creating goals, facilitates for feedback on those goals "</i> Feedback on behavior: <i>"provides analysis on your performance."</i> Review behavior goals: <i>"While doing health analysis Briea also makes a recommendation about Health Indicators for which you need to set up goals."</i> Prompts/cues: <i>"You can even ask Briea to maintain your Health Calendar. Reminders can be set up and later on, you can inquire about your reminders for today, this week, next week, this month, this year, or even next year."</i>
My Dietitian, BMI, Diet Tips, Health, Food Search [347]	A healthy diet (specifically Indian race)	Factual, culture-specific	Feedback on outcomes of behavior: <i>"overweight detected"</i> Action planning: personalized diet plans can be setup Instruction on how to perform the behavior: <i>"healthy tips and diet plans"</i> Monitoring without feedback: BMI is on display for the user to see based on the weight and height they input Information about health consequences: <i>"number of calories you will be ingesting in selected food items"</i>
NutriHealth - Nutrition Counselling Chat-bot [351]	Nutrition information for a healthy diet	Knowledgeable, factual	Information about health consequences: <i>"You'll get nutrition tips, diet management, nutrition facts, etc."</i> Social comparison: <i>"according to the DRIs, the recommended intake for total fiber for adults up to 50 years of age is 25g/per for women and 38 grams for men. For those over 50, the recommended intake is 21 grams for women and 30 grams for men."</i>
PAWEN - Running App [349]	Physical activity (specifically	Coach-like, knowledgeable	Self-monitoring:

	training running) in		Mental health - <i>"Journal everything you want: add your thoughts and feelings while journaling your state before and after each run. You can also use this time to reflect on how you're feeling about the workout in general."</i> Physical activity - <i>"Sync PAWEN with your Strava account to upload your running data into the app."</i> Social support: <i>"PAWEN is the best motivational buddy to help you keep going and not to give up... ever". "sourcing some powerful fuel to boost motivation" "Have a question about your run workouts or feel a lack of motivation this morning - message me in a chat."</i> Goal setting: <i>"PAWEN will provide you with a training plan for running, walking, and breathing workouts."</i> Information on health consequences: <i>"They will help you move forward to your goals as well as developing body awareness and decreasing anxiety levels."</i> Review behavior goals: <i>"You can provide PAWEN real feedback about your physical and emotional state before and after every workout. Back to them each time you want to reflect on your results."</i> Reduce negative emotions: <i>"Pick alternative workouts if something is not optimal that particular day. Decide what you want to do next and can still keep moving towards your goal without regrets or remorse."</i> Information on how to perform a behavior: <i>"figuring out the right pace for a certain run workout."</i> Adding objects to the environment: <i>"Find the right equipment for cold weather " to run in any weather condition the user might be in."</i> Action planning: <i>"Appoint training days during the upcoming week: pick the days you are available to train the next week, and PAWEN will adjust your training plan accordingly. It's your path, your time, and your rules!"</i>
SleepSpace – Doctor approved [345]	Better sleep quality and/or quantity	HCP-like, knowledgeable	Prompts/cues: <i>"bedtime/sleep diary reminder." "Simply place your phone on your nightstand or the foot of your bed to measure your unique sleep patterns. SleepSpace then plays the right sound at the right time to deepen your sleep cycles."</i> Information about health consequences: <i>"fall asleep faster, get deeper sleep, wake up refreshed."</i> Credible sources: <i>"SleepSpace is one of the most accurate ways to measure your sleep that has been supported by more than \$3.5 million in grant awards from the National Institute of Health. For more information, see our CEO Dr. Daniel Gartenberg's TED Talk, "The brain benefits of deep sleep - and how to get more of it."</i> Restructuring your physical environment: <i>"Sleep enhancement sounds - Experience our adaptive soundscapes designed to deepen your sleep by playing the right sounds at the right time in your sleep cycles." "Measure fluctuations of noises in your bedroom that are common causes of awakenings (i.e. snoring and traffic). Block out snoring with our noise-masking technology that targets the frequency of a snore."</i> Feedback on behavior: <i>"Sleep scoring - Track awakenings at night, sound disruptions like snoring, your sleep quality, and your unique circadian rhythm." "Sleep coaching – Want to dive deep into sleep improvement? Our team of coaches examines your sleep score to prescribe personalized actions for better sleep."</i> Problem-solving: <i>"Troubleshoot the most common ways to improve your sleep quality with our AI chatbot built by leading scientists and doctors in the sleep space."</i>

			<i>social support (emotional): "Cognitive behavioural therapy - Empower yourself with therapeutic exercise proven to help you fall asleep and stay asleep."</i>
RunMotion Coach – Running [348]	Physical activity (specifically adaptive training for wellness or marathons)	Friendly, coach-like	Graded tasks: <i>"It's important to have an adaptive training plan with various sessions to have more fun and achieve your goals." "Fit every distance: 5k, 10k, half marathon, marathon, trail running, and ultra-trail Or wellbeing goals: start running, run regularly or lose weight"</i> Goal setting: <i>"Your digital coach RunMotion Coach creates a customized training plan". "Your training plan is focused on your main goal."</i> Review behavior goals: <i>"Your digital coach RunMotion Coach creates a customized training plan whatever your schedule - which can change every week." "You can add intermediate goals"</i> <i>social comparison: "And it works! 88% of our users reach their goals!" "Join individual and group challenges "</i> Biofeedback: <i>"Training paces are based on your past races and your target time" " Know your Maximal Aerobic Speed and Endurance Index"</i> <i>credible sources: "Training paces are based on your past races and your target time, computed with a model validated by a research team at MIT."</i> Self-monitoring of behavior/outcomes: <i>"Import your activities from Strava and Adidas Running app or your GPS watch: Garmin, Suunto, Polar and Coros to get all your stats (distance, speed, calories burnt, training load...)"</i> Incentive: <i>"Join individual and group challenges and earn badges"</i>

9.3.4 mHON code guidelines

The percentages of apps complying with the mHON code guidelines are displayed in Figure 9.3. Only one out of the seven apps ("*Sleep space – doctor approved*") complied with all eight of the mHON code principles [345]. None of the included apps had any advertisements, so there was no disruption to the conversations with CAs or navigation of the app. The validity of all apps was confirmed with all apps being updated either in 2021 (71%, 5/7) or 2020 (29%, 2/7). The target audience could be deciphered based on all app descriptions and the developers were contactable in all but one app (86%, 6/7). The funding source was disclosed in most apps (71%, 5/7) whilst two apps were managed by individual developers (29%, 2/7). Whilst the information in all apps was presented objectively, references were only cited in one app (14%, 1/7). The authority of developers was mentioned in four apps (57%, 4/7) and these were experts in the area namely doctors, sports psychologists, international runners, and academics. Although not always explicitly visible to the user, six of the apps had a disclaimer indicating that the app does not replace the advice of a doctor and should be taken as information only (86%, 6/7). There was a confidentiality clause in four of the apps (57%, 4/7).

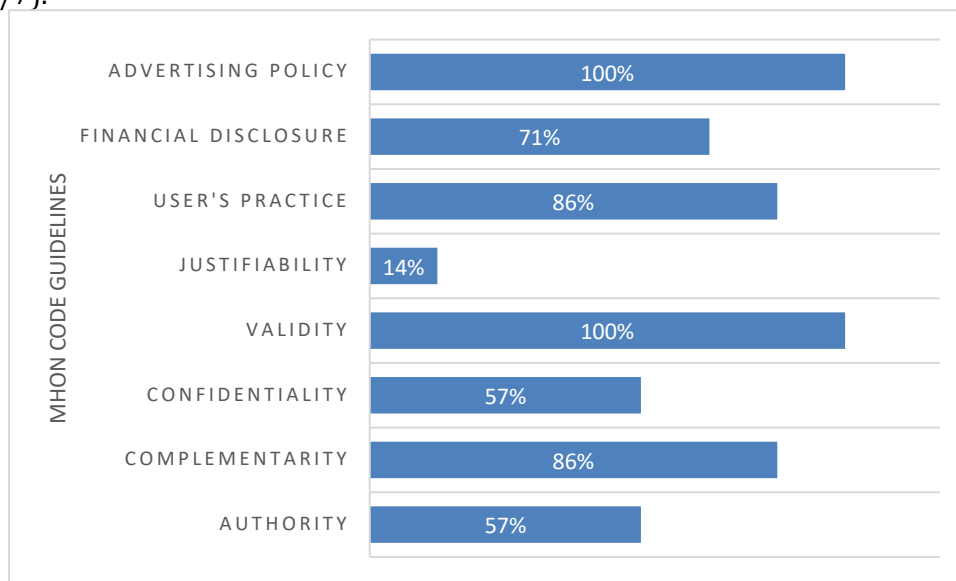


Figure 9.3 Percentage compliance of included apps with mHON code guidelines [100]

9.4 Discussion

9.4.1 Key findings

The search of the Google Play and Apple app stores using the 42 matters database and 26 search terms yielded seven CA apps relevant to healthy lifestyle behavior change. These apps targeted behavior change related to healthy eating, physical activity, or sleep improvements.

The apps varied in terms of quality and user experience. The apps with in-app purchases were of the best quality, were supported by credible sources of information, were well organized, had little (if any) technical glitches, and ensured a smooth user interaction. In some apps, the CA was part of a multifaceted app that, for example, had a diary for users to record their progress, users could consult a coach or monitor other health factors, and had options to join challenges and earn rewards. In other cases, the CA was the only part of the app with no other features to tap into. The apps of the lowest quality did not save the message history so users could not reflect and remind themselves of what was learned in the previous interaction or were not personalized to the user's needs and health history. The apps were assessed based on Michie's BCT taxonomy and 20 of these BCTs were used: credible sources, information on health consequences, demonstration of behaviors, information on how to perform the behavior, graded tasks, social comparison, biofeedback, action planning, monitoring of behaviors (without feedback), goal setting, feedback on behavior, reviewing behavior goals, prompts/cues, self-monitoring, social support, reducing negative emotions, adding objects to the physical environment, restructuring the physical environment, problem-solving, incentives. Information on health consequences was the most frequently used BCT. These health consequences could be positive or negative. Positive health consequences were more strongly emphasized by the apps. Perhaps, to get users interested in using the app, their aspirations for healthy living were promoted in these apps such as being able to sleep better and faster or eating healthier and weighing less as a result of using the app and CA.

CA personality was assessed using a personality framework derived from Haan et al. [127] (Table 4.1). Up to two personality traits were allocated to each app to maintain specificity. The personalities included were knowledgeable, factual, HCP-like, culture-specific, coach-like, and friendly. A lot of the apps were "factual" (57%, 4/7). In this project, the term was allocated to those didactic apps which were rigid, answering promptly to users' queries and needs but leaving little room to just chat and build a rapport, coming across as almost transactional. This can be helpful when the user just needs answers to their questions or needs a service provided. If to be coached or educated regularly to help improve their behavior, the CA may come across as if it's lacking some empathy and relational skills

causing the user to disconnect over time. But this may be suitable for objective users or those who are very self-motivated. The coach-like personality was used for those physical activity applications where the users want to challenge themselves to become a fitter version of themselves. This seemed like an appropriate personality trait to implement for this purpose as the coach-like nature will motivate and support the user. This coupled with a friendly and knowledgeable personality will be ideal as the user gets a mentor, coach, and educator all in one CA and app.

9.4.2 Comparisons with existing literature

Garnett al. employed the COM-B framework to reduce alcohol consumption using a smartphone app [352]. They used seven BCTs to instill behavior change. Information on effectiveness and usability came from a previous trial, user feedback, and an updated review on BCTs for alcohol interventions [352]. In this project, the number of BCTs identified ranged from two to nine with 29% (2/7) employing five and nine BCTs and 14% (1/7) using two, seven, and eight BCTs, respectively. As mentioned in Garnett et al. refinement of these apps could be achieved with user feedback however, there were insufficient reviews and ratings for the included apps in this project, making it challenging to note whether these apps were effective and usable [352]. The app creators did not mention any testing of the apps in a formal study setting but one app did reference publications that contributed to the app design and content (*Sleep space-doctor approved*) [353, 354]. Further research for those remaining apps might help assess their strengths and weaknesses and increase public confidence in them. Moreover, quantitative user engagement data and qualitative data in the form of user reviews and feedback would also be useful in determining whether these apps and CAs are effective, usable, and accepted by users for the intended purpose of healthy lifestyle change.

Band et al. posit that employing digital health interventions for self-management of health-related behaviors is a useful way of managing the burden on the healthcare system in a cost-effective manner [355]. They combine the MRC guidance that complex interventions should be based on the latest evidence, with a person-based approach in developing a digital health intervention for patients' self-management of hypertension [355]. As mentioned in the conceptual framework (chapter 6) a multifaceted team was also employed to inform intervention development. These included doctors, experts in behavioral science and hypertension as well as patient representatives [355]. In this apps assessment, members of the development team were sometimes disclosed: doctors, sports psychologists, academics and were mostly experts in their fields. Band et al. included patient representatives and employed a person-based approach indicating that the intervention was designed with the end-user in mind to increase acceptability and relevance [355]. Apps in this project

seemed to be lacking this non-expert perspective which could even reveal the needs of the target population. This is something to be considered in the healthy lifestyle behaviour change CA app development process for future apps.

Chiang et al discussed how appropriate BCTs can be selected based on the purpose of the intervention and known barriers to behavior change [356]. This was explained specifically in the context of using the behavior change wheel to improve medication adherence for antiretroviral therapy using the *WelTel* app [356]. Although relevant BCTs were identified for the included apps in this systematic assessment, a thorough understanding of the existing barriers and facilitators for healthy lifestyle change for the target populations were not explicitly disclosed hence, whether or not the developers performed this behavioral diagnosis is unknown. Chiang et al. mentioned that mapping out known barriers to behavior change onto the COM-B model could help maximize the digital health intervention and have a greater impact on behavioral outcomes. For example, incentivization and modeling were BCTs missing from the *WelTel* app, and it could be improved by considering these additional facets. This method could be adopted for future healthy living CA apps or to amend and improve existing ones.

Comparisons with Precilla

This apps assessment helped inform how *Precilla*, developed for healthy lifestyle behavior as part of this Ph.D. project, compares to existing healthy living CA apps available to the public globally via the two most frequented app stores (Android and iOS).

(a) Healthy living focus area

Unlike CA *Precilla*, which was also targeted at healthy lifestyle behavior change, the apps included in this project focus on one aspect of healthy living such as healthy eating, physical activity, or sleep improvements. *Precilla* aimed to provide help and support on four areas of healthy living (diet, exercise, sleep, and stress) whilst also informing users on diabetes and prediabetes.

(b) CA level of intelligence

Some apps in this project used AI exclusively or in combination with rule-based technologies whilst *Precilla* was exclusively rule-based. Although there were some technical difficulties with the AI features and sometimes a need for trial and error to understand which inputs would generate the right output from the CA, some rule-based apps provided a more streamlined user experience than *Precilla*. The *Sleep space-doctor-approved* app, for example, was rule-based and removed the free text bar, so users are not tempted to type any text [345]. This is something that can be implemented to improve *Precilla*. Additionally, participants commented that *Precilla* sometimes did not provide

enough answer options, the *Run Motion coach* app sometimes had up to five options for users to choose from [348]. This was a very well-organized app, exclusively rule-based and very well implemented. Employing up to five answer options for users to select from is something Precilla may benefit from by fulfilling the user's need for autonomy and variety in their answer choices.

(c) Choices of BCTs

Plenty of the BCTs identified in the included apps were also identified in healthy lifestyle behavior change CA, Precilla (chapter 7). Social comparison was present as she indicated the recommendation for weekly frequency of exercise for healthy adults. Images for exercise routines were included as a demonstration of healthy behaviors. The apps in this review also demonstrated exercises sometimes in the form of videos. This is something that can be incorporated into Precilla and it was also mentioned by participants in interviews (chapter 8) that a variety of media (such as text, images, and videos) would be appreciated. Precilla sent messages four times a week and these notifications acted as prompts/cues to actively get the user to engage. Other prompts/cues used by apps in this review were reminders for health activities based on their calendars or specific sounds to deepen the user's sleep cycle at night. Besides engagement, prompts/cues can be used by Precilla for these purposes too. In a similar way to the *PAWEN running app* [349], Precilla helped to reduce negative emotions associated with healthy living such as stress and anxiety from an inability to perform healthy behaviors. Instructions on how to perform healthy behaviors were included in the form of food recommendations, for example. Credible sources were employed to inform the content for Precilla including published scientific evidence, the expertise of researchers, and non-expert input from prospective users ie. the general population in Singapore. These sources were similar to those mentioned in the included apps – doctors, experts, sports psychologists, etc.

9.4.3 Strengths and limitations

Strengths and limitations of included apps

The technology in some apps was not yet optimized and this compromised the user experience. *Briea* used speech recognition technology, but the speech recognition was not always very good as the user sometimes had to repeat several times before the CA understood them [350]. Efforts to improve such speech recognition technology would be beneficial as voice bots can provide a more human-like interaction, might enable multitasking on the smartphone, and provide a more-inclusive solution (eg. for those lacking some motor abilities, or illiterate persons). In other circumstances, the AI specifically was inefficient at sustaining a conversation with the user. *My Dietician* used rule-based and AI technology however, the AI was limited and not very effective [347]. Users were told what

they could type to get a response from the CA, even in such cases, the user's input generated an error message. In such cases, a purely rule-based CA might be more appropriate. *Nutrihealth* was similar in that the AI capabilities of the CA were limited but this was not explicitly disclosed by the CA so the user had to use trial and error to find the right input which would generate a response from the CA [351]. Perhaps, a user manual for appropriate inputs or a button-based CA might have been easier to navigate and could have boosted user experience. Prior to the optimization of AI or speech recognition technology, an alternative could be to direct the user to a live chat with a human. For example, in cases when the CA did not have an answer readily available for the question posed by the user, the *PAWEN – running app* would seemingly consult a human administrator [349]. Unfortunately, this meant the response would only come hours later, almost negating the very point of a CA which is meant to be readily available for timely responses.

Strengths and limitations of this apps assessment

The BCT taxonomy was used to assess the included apps. This is a world-renowned taxonomy and has been used for several mHealth assessments with success and may have been the best option for the app's assessment in this project. The 42 matters database was used to search for apps in both app stores. This ensured a more sensitive search compared to manual searching of each app store individually. The search strategy aimed to be as comprehensive as possible, but some search terms yielded no results whilst others picked up hundreds of apps that were of no relevance indicating that the specificity of the search strategy can be improved upon. A lack of phone numbers for certain geographic regions and a need for special invitations to access certain apps meant that some apps of potential relevance could not be assessed as part of this project. Mental wellness has a major impact on healthy living, but mental health apps were excluded due to the large number of CA apps covering this area and the diversity in their approaches and applications which may not always relate to healthy living. Mental health CA apps cover a broad range of applications from scientific methods (eg. CBT) to self-improvement and wellness and need to be studied more extensively in a separate review.

9.4.4 Conclusions

To conclude, this systematic apps assessment identified a conservative number of apps from the Android and iOS app stores targeting healthy lifestyle behavior change for healthy populations (n=7). The apps considered the mHON code guidelines, did not employ advertisements, had CAs with a factual personality trait, and used up to nine BCTs for their healthy living-related aims. The apps targeted different apps of healthy living including diabetes prevention, exercise and wellness, sleep improvements, and healthy eating, respectively.

Chapter 10 – Discussion and conclusions

The first part of the chapter reiterates the aims of the thesis (Section 10.1), presents an objective presentation of the summary of findings (Section 10.2) from the PhD project. This is followed by comparisons with existing work (Section 10.3), the impact the findings will have in general (Section 10.4) and on the policies in Singapore (Section 10.5). The second half of the chapter critically appraises CAs in general along with the researcher's reflections on the projects. This considers how the wide adoption of healthcare CAs might lead to certain challenges (Section 10.6), the limitations of the project (Section 10.7) and reflections by the researcher (Sections 10.8 and 10.9). The chapter concludes with recommendations for the future (Section 10.10) and concluding remarks (10.11).

10.1 Aims of the thesis

The main aim of this Ph.D. project was to develop a healthy lifestyle change conversational agent for the general population in Singapore. This revolved around contributing to the evidence base around the use of conversational agents for healthcare, with a specific focus on rule-based text CAs for healthy lifestyle changes. To provide a comprehensive summary and synthesis on this topic, the following six objectives were formulated and executed:

1. Map and analyze the current uses of conversational agents in healthcare and identify knowledge gaps by performing a **scoping review**
2. Propose a systematic list of steps on how to develop a rule-based text healthcare conversational agent and present it in a **conceptual framework**, using evidence from the literature and a validation study
3. Understand the public's perceptions and needs with regards to conversational agents for healthy living and diabetes prevention in Singapore, using **semi-structured interviews** and subsequent **qualitative analysis**
4. Develop and test the feasibility and acceptability of a healthy lifestyle behavior change conversational agent for the general population in Singapore in an **online feasibility study**
5. Understand the public's views on a conversational agent for healthy lifestyle change and obtain their feedback on areas for improvement using **semi-structured interviews** in the follow-up phase of a pilot study
6. Identify and describe the main characteristics and evaluate the reliability, transparency, and BCTs used in healthy lifestyle change conversational agents through a **systematic assessment of apps**

10.2 Summary of findings and contributions:

Each main chapter of the thesis contributed towards addressing each objective successively.

Chapter 4 aimed to map and analyze the current uses of conversational agents in healthcare and identify knowledge gaps by performing a scoping review. The scoping review employed Arksey and O'Malley methodology and 63 search terms were used to search grey and bibliographic databases. The key findings are summarised in point form here:

- (i) 47 studies were included and most of the literature had been published in the last two years before the review was conducted (2016-2018), reiterating the nascent nature of this field of research, along with the large number of case studies and lack of large-scale RCTs.
- (ii) The majority of the CAs identified in the papers used text as the main mode of communication and were machine learning based.
- (iii) The CAs were categorized based on their applications and treatment and monitoring, health services support and patient education were the most frequently reported applications.
- (iv) Only studies including an evaluation of CAs were included and the results noted that effectiveness, accuracy, and acceptability were generally positive.
- (v) Future reviews could investigate extrapolating the CAs to other usage populations, beyond just patients to caregivers (formal/informal) and even HCPs.
- (vi) Speech-based CAs appeared to be underrepresented and should be studied in future studies as they could be advantageous to populations lacking the dexterity and eyesight for typing and reading text from a small screen on a smartphone.

The interest to develop more CAs was indicated by the sudden spike in healthcare CA-related publications in the most recent years. It was also noted that not all healthcare researchers or research teams had access to computer scientists and other experts well versed in CA development and publication. As such, **chapter 5** proposes a systematic list of steps on how to develop a rule-based text healthcare conversational agent in the form of a conceptual framework, using evidence from the literature and a validation study.

The key findings for this study are presented in point form here:

- (i) The conceptual framework was based on the analysis of 23 mHealth frameworks (general mHealth, smartphone apps, text messaging), 6 CA frameworks, 17 primary studies reporting on the design and development of rule-based healthcare CAs.
- (ii) The framework comprised 5 steps and two overarching themes: (1) Defining the goal, (2) Creating an identity, (3) Developing the content, (4) Building the conversation, (5) Piloting & Evaluation. Co-design and privacy & security were identified as overarching concepts, relevant at each stage of development.
- (iii) A pilot study was also conducted to validate the framework (chapter 7).
- (iv) This framework provided guidance, which was specific to CA development, something currently missing in more generalized mHealth frameworks for smartphone apps, for example. CA specific details included a consideration for CA personality, tone of speech, personalization, building a conversation flow, etc.
- (v) As this framework was developed specifically for rule-based, text CAs, further work should look into developing a similar framework for AI-based CAs as the interest in this area is also growing.

The study in **chapter 6**, aimed to understand the public's perceptions and needs with regards to conversational agents for healthy living and diabetes prevention in Singapore. This was achieved using semi-structured interviews and subsequent qualitative analysis. The key findings from this study are presented below:

- (i) Participants talked about: (1) the uses of conversational agents (definitions, beneficial and unfavorable traits), (2) the ubiquity of smartphone apps including their usage in daily life, advantages and disadvantages, (3) understanding diabetes, prediabetes, what they mean, sources of information and where they feel knowledge is lacking and (4) finally the barriers and facilitators to healthy living in Singapore.
- (ii) Recommendations on content and delivery for future conversational agent interventions for healthy lifestyle behavior change were the main takeaway points from this study.
- (iii) The public would appreciate more information shared on the implications, symptoms, and prevention methods for diabetes and prediabetes.
- (iv) More novel and trustworthy content and actionable advice were also suggested around healthy eating, physical activity, stress management, and sleep.

- (v) Beneficial CA delivery methods to employ for healthy living applications were personalization, suitable frequency, and duration of messages, easy, engaging, and relevant content.
- (vi) Additional items like incentives (points, rewards, discounts) and group support could also be considered if applicable for the intended purpose.

To apply some of these suggestions from the co-design/needs assessment (chapter 6) and to validate the conceptual framework (chapter 5), CA *Precilla* was assessed for its feasibility and acceptability in a pilot study and discussed in **chapter 7**. The key findings from this study are presented below:

- (i) In this online trial, using a Facebook Messenger CA for healthy lifestyle behavior change in Singapore, it was possible to recruit 37.5% (75/200) of the target sample size in 1 month.
- (ii) Retention was high at 93% (56/60); as was CA engagement, with all the conversations being completed by at least 50% (28/56) of the participants consistently. Acceptability, usability, and satisfaction were also high.
- (iii) Any definitive signs of improvement in QoL, knowledge, diet, physical activity, sleep, and stress would have to be studied in more detail in a future study of effectiveness.
- (iv) The CA engagement data revealed that there was a slight decrease in the number of interactions that the participants completed fully with each successive week. It is possible that the content and delivery methods were considered repetitive because a review of take-home points from the previous interactions were provided from week 2 onwards.
- (v) Introducing some novelty in the timing of messages or messaging format could help in the future to keep interest rates higher.

To follow up on users' experience of using CA *Precilla* during the 4-week pilot period, follow-up interviews were conducted, and **chapter 8** aimed to outline the findings and qualitative analysis of these interviews. The key findings are presented in point form here:

- (i) 20 participants were interviewed, and four themes emerged from the analysis: (1) enjoyable and acceptable experience of *Precilla*, (2) sub-optimal experience(s) of *Precilla*, (3) comparison of *Precilla* to other sources, (4) favorable aspects of conversational agent interventions, (5) recommendations on how to improve *Precilla* in the future.

- (ii) The main recommendations for the future involved selecting appropriate message timings, involving participants in answer option generation, selecting an acceptable delivery platform based on target population usage, and finally, engagement strategies to keep users entertained and interested in future conversations.

The study in **Chapter 9** aimed to describe the main characteristics and evaluate the reliability, transparency, and BCTs used in healthy lifestyle change CA apps through a systematic assessment of apps. The android and iOS app stores were searched using 26 search terms and the 42 matters apps database. The key findings are presented in point form below:

- (i) Seven apps were found to be applicable for the assessment.
- (ii) The quality of the apps varied from those offering in-app purchases having the greatest quality and offered the best user experience in the sense that there were no technical glitches, and the content was well-informed and trustworthy.
- (iii) Although the CA was a central part of the apps, in some cases the apps were multifaceted and included reminder functions, options to journal progress, consult a coach, link up with a smartwatch, or join challenges and earn rewards.
- (iv) The apps of the lowest quality did not save the message history so users could not reflect and remind themselves of what was learned in the previous interaction or were not personalized to the user's needs and health history.
- (v) The apps were assessed based on Michie's BCT taxonomy and 20 of these BCTs were used: credible sources, information on health consequences, demonstration of behaviors, information on how to perform the behavior, graded tasks, social comparison, biofeedback, action planning, monitoring of behaviors (without feedback), goal setting, feedback on behavior, reviewing behavior goals, prompts/cues, self-monitoring, social support, reducing negative emotions, adding objects to the physical environment, restructuring the physical environment, problem-solving, incentives.
- (vi) The CA personalities included were: knowledgeable, factual, HCP-like, culture-specific, coach-like, and friendly. A lot of the apps were "factual" (57%, 4/7) and came across as rigid, lacking the ease and flow of regular human-human conversations.

10.3 Comparisons with existing literature

Over the last couple of years, the coronavirus pandemic has had an impact on multiple facets of life and certainly also made a dent in the research landscape around conversational agents. Of the CA applications identified in the scoping review performed (chapter 4), usage applications of CAs

spanned treatment and monitoring, health services support, education, behavior change, and diagnosis. An additional theme would now be identified, specifically for COVID-19 as CAs have been developed for diagnostic evaluation and recommendations on patient management following a positive test result [357]. Additional applications related to COVID-19 were about self-triage, exposure monitoring, tracking symptoms, combating fake news; as identified in a scoping review [358].

The broadening of CA usage applications was also implicated by Dingler *et al.* who posited that CA usage for mental health applications has dominated in the past but there is now a shift towards more context-aware agents which are more proactive and able to engage users better to deliver health information and services [5].

The intention for researchers to conduct timely reviews on the rapidly growing field of healthcare CA research is evidenced by Guerreiro *et al.* who presented a protocol for a very broad and comprehensive evidence map on CAs for health and wellbeing for all age groups [359]. They aim to look into evidence not just from primary studies but also from meta-research to be as thorough as possible in presenting all the research out there [359]. Similarly, Martinengo *et al.* propose a protocol for a scoping review exploring the BCTs employed in healthcare CAs [360].

Other methods for qualitative research, such as user feedback, have been conducted to identify user perceptions around healthcare CAs [361]. Here, the main issues raised by users were that some CAs had limited vocabulary, poor voice recognition accuracy, or didn't have a robust default message or strategy to manage errors in inputs from users. In contrast, positive points raised were regarding the personality of CAs and the empathy they displayed [361]. The points mentioned were very much aligned with the feedback collated from the qualitative studies presented in this thesis around the healthy living CA Precilla. This indicates some level of consistency in users' expectations and perceptions of CAs not just in Singapore, but more internationally.

The main example of quantitative empirical research presented in this thesis was in the form of a pilot study (chapter 7). Other methods of evaluation can also be explored in the future such as Micro-RCTs, where randomization happens every time that intervention component is delivered. This has been used for a CA designed to promote physical activity [266]. Alternatively, ecological momentary assessments (EMAs) can be considered. These study an individual's behavior (using surveys, for example) at a specific moment in their natural environment at regular intervals [267]. These have been used in mHealth interventions to reduce alcohol intake in adolescents and have the advantage of overcoming recall bias which can happen in conventional surveys [267].

10.4 Impact of this work

Conducting this research on CAs and the subsequent publications have been pivotal in securing collaborations with similar research teams. One such group we have had the privilege to work closely and exchange ideas with are Future Health Technologies (FHT), affiliated with ETH Zurich. These collaborations include invitations to present at their scientific seminars and critical feedback on manuscripts for publication.

The publications generated from a lot of the work presented in this thesis (chapters 4, 6 & 7) have been cited collectively up to 59 times for a variety of publications. They have been cited for a wide variety of CA applications such as CAs related to the COVID-19 pandemic, CA literature reviews, and papers discussing CAs for specific health conditions such as cancer or diabetes.

10.5 Potential implications on policies in Singapore

“Affordable Health Care” was a report passed by the government in 1993 as a way of containing costs, enabling subsidies all whilst maintaining a high quality of healthcare in Singapore [49]. The White Paper encompassed five main objectives namely: (1) to promote good health to nurture a healthy nation, (2) to instill responsibility in each individual to take care of their health and not be over-reliant on subsidies and insurance schemes, (3) to ensure that basic healthcare is affordable to all citizens, (4) to encourage competition for better care and efficiency, (5) to step in when the healthcare sector needs support of when the market is unable to keep healthcare costs in check [49].

The aims of this Ph.D. project and the findings are very much in line with this “Affordable health care” plan. A healthy lifestyle CA has the intention of promoting good health through healthy lifestyle choices, with the added function of preventing diabetes, a condition of chronic concern in Singapore. Furthermore, Singapore is not a welfare state, and each individual is responsible for their healthcare funding. In line with empowering people to take responsibility for their health is the notion of prevention over treatment. Existing examples which advocate for this mindset are the “war on diabetes”, national step challenge, ban on advertisement for sugary drinks, and stronger barriers to control smoking in the country [362]. The healthy living CA, *Precilla*, proposed in this thesis has the same motivations as these campaigns. The CA is easily accessible given the ubiquity of Wi-Fi, it can provide personalized tips and advice, and may even consider the use of appropriate BCTs to implement the necessary behavior change to nurture a culture of prevention over treatment all from the convenience and comfort of your smartphone.

The alignment with healthcare aims and policies in Singapore, the urge to evolve towards a smart nation, the high smartphone penetration rate, and the willingness for the public to communicate digitally are all factors which support and encourage the implementation of healthcare CAs like *Precilla* in Singapore.

10.6 Potential challenges to wide adoption – and how these issues might be mitigated

As promising as CAs are to healthcare, they do not come without their setbacks. The literature has highlighted a multitude of reasons for scepticism about CA applications in healthcare. For example, as promising as novel mHealth interventions seem in theory, their potential effectiveness is not always guaranteed [363]. Furthermore, barriers to implementation are highly likely and other issues around digital (in)equity and safety concerns also exist [44, 364, 365]. Consideration for these collateral factors helped dilute any possible thoughts for the researcher around “silver bulletism” which would assume the CA in this project will be the “be-all, end-all” solution for the aims presented in this thesis [284]. As promising and helpful as CAs have the potential to be, their limitations, along with the challenges associated with their widespread usage are recognised.

A lack of potential effectiveness for very promising initiatives in mHealth have been demonstrated in a few studies, for example no-cost mobile phone reminders were tested in a pilot study to improve the attendance for HIV test results in rural Swaziland [363]. The intervention was free, mobile phone penetration rate was high (>80%) and yet, there were no differences in attendance before and after the intervention [363]. Low levels of effectiveness for novel mHealth initiatives such as smartphone apps or CAs maybe due to their implementation in isolation. In an article “Bots are good, but they can’t do it alone,” a technology writer presents the ideal use-case scenario for these disruptive agents, using customer service as an example [366]. *“We see these chatbots becoming co-pilots. When an agent is involved, the chatbot is right there being the note taker or transcriber, offering suggestions [to the agent] in real time about what to do next. [366]”* This tandem relationship has been applauded as the human employs its indispensable skills of talking to the customer and understanding their needs while the CA picks up on all the minute details so it can collaborate with the live agent and a useful resolution can be reached more efficiently [366]. Dingler et al. present a similar sentiment where they affirm the aim for CAs is not to replace HCPs but rather to help them navigate the healthcare system and empower them to manage their own health [5]. They propose CAs which integrate the patient’s medical history with data points from the wearable technologies, monitors their symptoms and proactively shares important data with

their GP and automatically books appointments when deemed necessary [5]. As such, introducing CAs as part of a blended programme, integrated with in-person components or in collaboration with educational programmes may boost effectiveness.

Barriers to implementation are another cause for concern with healthcare CAs. Such barriers could manifest as: contextual factors, low technological competency in the target group or even a lack of streamlined integration with health services which might lead to fragmentation of data [367]. On a large scale, the high cost of software or technology malfunction are all possible which can cause inconvenience and delays in the delivery of healthcare or provision of materials, feedback and educational resources to the user [367]. Other barriers that may arise could be decreased adoption of the CA due to perceived limited functionality or perhaps if the platform of choice (eg. FB/iOS/Android) is not device-agnostic [368]. Additionally, bureaucracy could pose as a noteworthy barrier to implementation as it has been cited as a hindrance to the adoption of connected care technology (concerned with health technology, digital media and mobile devices) in multiple countries, including Singapore, where this was rated as a top barrier by 45% of healthcare professionals [369]. However, such issues are likely to pose less of a battle in the near future, as Singapore aspires towards becoming a smart nation, it must make way for promising digital health initiatives [370]. Furthermore, a qualitative study on the acceptance and resistance of new digital technologies in medicine reported some promise for the uptake of new technologies [371]. This was insinuated by medical staff embracing technology to foster more communication with patients and their carers. Similarly, patients were enthusiastic about new technology as it enable them to have greater autonomy in directing their healthcare [371].

Even as smartphone usage thrives, globally as well as in Singapore, the uptake by certain sub-groups is lower than average. For example, despite Singapore's attempts to be more digitally inclusive, it was noted that older adults (those aged 60 and above) lacked internet usage due to low digital literacy and access to the internet [364]. An additional observation was that 1 in 15 of these older adults faced the added caveat of a health condition which hindered their usage of digital devices [364]. This survey with 4000 older adults in Singapore revealed that the digital divide amongst Singaporeans was caused by more than just access to digital tools and internet access, but that health status played a key role too [365]. The study further revealed a gender divide where older males were less likely to use digital devices for health-related reasons while older females were less likely to use technology for non-health related reasons [365]. These findings make a crucial point on digital inequity and indicate that the newest CA inventions, even for healthcare,

may face challenges being implemented in the ecosystem due to a multitude of factors. As such, as propositioned by Assistant Professor Malhotra at the Duke-NUS Health Services and Systems Research Programme, *“policymakers should consider supporting research and development into assistive technology and design aimed at helping older internet users overcome health-related difficulty in internet use [364].”* Additionally, context specific research also needs to be done to overcome non-health related barriers to adoption and to narrow the gap caused by digital divides associated with age, health, socioeconomic status etc. The first stage stated in the conceptual framework (Chapter 5) of defining the target population can help alleviate this issue around digital equity by first assessing the socioeconomic status of the target population, access to smartphones, familiarity with digital health as well as technological competency. This could inform the level of complexity and role to be played by the CA in helping to manage their healthcare.

10.7 Limitations and methods for improvement

This project focussed exclusively on rule-based healthcare conversational agents in all areas of its empirical research. As beneficial and pragmatic as this approach is, the public is attracted to the slick nature of AI and researchers are trying to match those expectations. Although the work presented here is not directly relevant to AI CAs, it remains largely relevant but will need to be expanded upon to provide a clearer view and guidance for AI healthcare CAs.

If given the opportunity to perform the studies outlined in this thesis again, some changes could be made. Automating emails to participants regarding study participation, obtaining informed consent, and reimbursement vouchers would reduce the burden on administrative activities whilst also being time-efficient. Piloting the CA Precilla in multiple rounds with colleagues could help to identify more limitations of the CA such that amendments and improvements could be implemented before testing with a study population.

To aim for a larger sample size for a more RCT-styled study design, it would be beneficial to apportion more time for recruitment and post-study posters on multiple FB pages and groups, with more snowballing, and other (free) social media platforms to widen the reach for prospective participants. To prevent any attrition, later, it might be helpful to invite participants to an in-person event (with incentives such as gift bags, vouchers, or reimbursement for their travel) to legitimize the study and put a face to the researcher which may enhance trust.

A point for improvement for a rule-based CA such as Precilla would be to remove the text box and keypad, from the chat so that users are not tempted to type free text which Precilla is not designed

to manage (a method noted from some apps in the apps assessment) (chapter 9). Additionally, providing more answer options makes the user feel listened to, like their needs are being fulfilled, and like they maintain autonomy in how the conversation flows. Implementing the CA on Telegram in the future might be a worthy change as that seems to be a neutral method of communication (the messaging app was mentioned numerous times in interviews) amongst diverse age groups and is used for both personal and professional purposes in Singapore. As feasibility and acceptability were the main areas of focus for this pilot study evaluating Precilla, effectiveness should be the priority in future studies. This may be explored by using BCTs more intentionally in the conversations and the messaging delivery style.

10.8 Impact of the research on the researcher

These projects related to my Ph.D. have made me realize that there is so much more to the role of a researcher than scoping the literature, analyzing data, and writing up publications. There are administrative tasks that need to be learned and done, new software that needs to be navigated, myriad different personalities to communicate and work with, repetitive and menial tasks that also need to be done so that the research can progress and flourish. This experience has been eye-opening and pivotal in moulding me into a more all-rounded person and a researcher with a solid grounding.

As optimistic as I was about a novel CA and the potential it had to help with healthy lifestyle changes, talking to the interview participants, dissecting the interviews and studying behaviour change theories revealed that such an intervention is not a final answer to the challenges Singapore faces with diabetes and pre-diabetes. I also acknowledge that it would not work best on its own. What about those who need individualised in-person accountability? Those who cannot work alone to make lifestyle changes? Those who are not comfortable with using technology for health due to privacy concerns or poor digital literacy? Blending it with other platforms and treatment programs would be might be a more realistic starting point. As noted earlier, CAs should not be developed with the aim of replacing HCPs but rather to support them in providing the most efficient and effective care for the target user.

If given the opportunity to do the projects again, in hindsight, there are some things I may have done differently. Taking into consideration the credibility, transferability, dependability and confirmability framework in assessing the quality of the qualitative studies [301, 302], I would be inclined to implement the “member check” component. This would allow me to compare the codes

generated by the two researchers with the opinions of the participants. This would help to corroborate our understanding of their comments and views expressed in the interviews and confirm whether the definitions we interpreted were accurate. This would also be an effective way of applying co-design at another stage of CA development as advised in Chapter 5.

With regards to the feasibility study in Chapter 7, I would certainly extend the recruitment period to at least 3 months, as suggested by the literature. I would add an additional inclusion criteria of individuals keen on changing their lifestyle (ie. motivated) to observe if the levels of preliminary effectiveness are improved. Motivation is a prime component of the COM-B model of behaviour and is supported by the PRIME theory of motivation which explains how reflective, emotional and habitual processes come together to perform a behaviour [331]. Considering how important motivation is in encouraging behaviour change, the existing inclusion criteria doesn't account for this. An example of a question which could be included in the eligibility criteria could be, "Are you willing to change your habits for a healthier lifestyle within the next month?"

If given the opportunity to do the feasibility study again, I would not implement the CA in isolation but instead supplement it with videos from researchers and doctors talking about diabetes, pre-diabetes, prevalence, risk factors and methods for prevention. The CA could also provide links to in-person events that promote physical activity, or free community programmes for stress management and better sleep quality. This kind of blended approach may plausibly be implemented with the resources available and could provide the users with a more holistic experience not just limited to their smartphones but also enabling them to interact with other humans for support, potentially urging their healthy lifestyle change along.

10.9 Quality and trustworthiness of the Qualitative Research

The qualitative studies discussed in Chapters 6 and 8 contributed significantly to this PhD project however, the quality and trustworthiness of these projects were not addressed within the chapters. Given the subjective nature of qualitative research, it was necessary to perform this assessment. The following quality criteria were employed: credibility, transferability, dependability and confirmability [301, 302]. In addition, reflexivity was also included to ensure the highest level of transparency in data presentation. Table 10.1 shows the quality criteria as defined by Lincoln and Guba [301]; Sim and Sharp [302] followed by an explanation of how each item was addressed in this study.

Table 10.1 Quality criteria to ensure trustworthiness in qualitative research. Based on definitions by Lincoln and Guba [301]; Sim and Sharp [302].

Credibility	Prolonged engagement	✓
	Persistent observation	✓
	Triangulation	✓
	Member check	x
Transferability	Thick description	✓
Dependability and confirmability	Audit trail	✓
Reflexivity	Diary	✓

Prolonged engagement. Multiple distinct questions were asked on each topic of interest (please see interview guide Appendices 6.2,8.2). Participants were encouraged to support their statements with examples, and the interviewer asked follow-up questions based on participants' unique responses as they were semi-structured interviews. The researchers studied the data from their raw interview material thoroughly until codes and themes emerged to provide valuable conclusions.

Persistent observation. The two coders constantly read and re-read the transcripts to analyse them, code them and revise the coding as new concepts emerged. The initial stages of coding helped to familiarise them with the data and examine the core concepts and categories. In subsequent rounds of coding, the themes were re-labelled to better encompass concepts that could be condensed together. The researchers studied the data and went through several rounds of coding before the data could be used to narrate a holistic, meaningful, and representative

Triangulation. Methodological triangulation involved collating data from the interview transcripts with field notes taken by the interviewer during the interviews. Investigator triangulation was achieved as two researchers analysing the transcripts in parallel. The first five transcripts were done independently. They then reconvened for a discussion on their interpretations, after which a codebook was generated to ensure uniformity in the assignment of codes and a shared understanding of their meanings. The researchers continued to meet regularly through the data analysis process and even consulted the PhD student's co-supervisor for feedback on the suitability of the codes and themes generated. Data triangulation was enabled through the combination of raw material from the interviews with codes, themes, and concepts noted from the literature.

Member check. This technique was not employed in these qualitative studies but would be considered in future work. The idea of sending analysed transcripts to participants to confirm the accuracy of interpretations of their findings is valuable and mirrors the idea that co-design is relevant at all stages of CA development as highlighted in the conceptual framework (Chapter 5).

Transferability. Context and interview experiences are elaborated in the Reflexivity account (Section 10.9.1)

Dependability and confirmability. The research process, team meetings, reflections, resources, interview guides, coded transcripts and all other research materials, emerging findings and information on data management were all systematically organised in a secure folder on a data storage system provided by the University. These were all collated to ensure transparency in the research and dependability and confirmability can be assessed. The study was audited by the NTU IRB committee who signed off on the study. All the necessary data was presented and satisfactory for the audit.

Reflexivity. Please see the Reflexivity account in Section 10.9.1.

10.9.1 Reflexivity

Reflexivity is an important part of the qualitative research process. Whilst quantitative research thrives on the absence of bias, qualitative researchers prepare a reflexive account to indicate how the subjectivity helped to shape the results, data collection and subsequent findings [372]. Table 10.2 outlines the reflexivity by the researcher following the guidance set out by Olmos-Vega et al [372]. Here they describe four facets of reflexivity: personal, interpersonal, methodological and contextual. In personal reflexivity, the researcher is required to state and reflect on their expectations and assumptions in relation to the setting, participants and data. Interpersonal reflexivity is more concerned with how the relationships formed during data collection (between researcher and participant) and data analysis (between researcher and colleagues) might have influenced the people and the results. In methodological reflexivity, the researcher is required to critically reflect on their methodological decisions. Finally, contextual reflexivity asks the researcher to reflect on how aspects of the context would affect the research and those involved [372].

Table 10.2 Reflexivity diary for qualitative studies (Chapters 6 and 8). Adapted from Olmos-Vega et al [372].

Personal	I (the PhD candidate) was the primary researcher in both qualitative studies flanking the pilot feasibility trial (Chapter 7) as these studies formed a major of my PhD project. This was my first-time conducting interviews with participants, both in-person and over the telephone. In preparation for this qualitative study, I had completed online courses on “Coursera” and studied the literature on conducting semi-structured interviews and performing analysis on qualitative data. Having written up the interview guide and rationale for the study, I also knew about the “War on diabetes” in Singapore, that Singapore has the second highest prevalence of diabetes among developed countries and had some awareness that the population faced issues around healthy eating, exercise, sleep and stress management. Having looked through the Facebook pages and groups to recruit participants, I did note that a number of individuals were advocating the “Keto-diet” as an excellent method for healthier eating. I noticed the diversity in the educational levels of participants with some being white collar professionals, being very eloquent in English whilst others were students or older adults with a much lower monthly household income and were not particularly fluent in English suggesting, it was perhaps their second language. This language barrier may have restricted some individuals from expressing their full, honest opinions and needs. However, the interviews proceeded uneventfully with all participants being able to answer the questions and provided relevant responses. I was pleased to see that participants were willing to share their thoughts quite generously and came away from the interviews always feeling like I’d collected some rich data. Some interviews were more similar than others. I felt like in conversations with people who had an academic or scientific background, the data was richer than with those who did not. Nevertheless, I appreciated having a range of views both from the layman and the academic. I had some misconceptions prior to starting my interviews (Chapters 6 and 8) which, if abolished could have helped me to collect richer data from longer interviews. I was pleasantly surprised to see so many volunteers for the interviews. Having this preconceived idea that Singaporeans are always in a rush, I really did not think we would get so many volunteers willing to take time out of their working days to do me a favour by discussing our CA Precilla. On the same point of volunteering time, I was wary of taking up too much time and aimed to make the interviews as streamlined as possible. Considering how forthcoming the participants were, I would have liked to ask more questions possibly those specifically related to future implementation of a CA. Some additional questions would be: <i>“If a doctor prescribed a CA and asked you to download it onto your phone, would you be willing to do so?”</i> <i>“Are you looking forward to Singapore progressing as a smart nation?”</i> This has been an important learning point and pivotal transition for me as a data collection instrument. In the future, I think I will approach such interview

	situations viewing the interviewee as a fellow researcher, here to provide valuable insight and give them the time and respect that they deserve considering the contribution they are making to the subject. I would potentially even add a few more lines to my introduction reiterating the value they are about to provide to empower them before the interview begins.
Interpersonal	I contacted interview participants using the “WhatsApp” application on the departmental smartphone. The profile picture featured our study team members in semi-professional attire. This was done to prove the legitimacy of our study and research team. At the beginning of each conversation, I began by introducing myself as a PhD student and reiterated the aims of the study. This was done to orientate the participant and provide a firm grounding for the project and demonstrate methodological rigour. However, to avoid creating much of a hierarchy between the researcher and participant and in an attempt to make the participants feel at ease, I asked multiple questions through the introduction so they wouldn’t feel like I was the main speaker directing that conversation and they were a secondary counterpart. Some of these questions were to ask if a 20 min duration was alright; if they could hear me clearly; and to inform them that they could pause the interview at any time if they need to ask questions or take a short break. Some participants were also post-graduate researchers, familiar with the journey of a PhD student. This familiarity may have encouraged them to be more forthcoming in their responses, wanting to be as helpful as possible as they understood the importance of data collection in their own academic careers. Being at a similar educational standpoint, the issue around hierarchies may have dissipated, making those participants feel more like they were sharing their thoughts with a colleague as opposed to a researcher supervising the study they were a part of. Furthermore, being a Singaporean myself I think was a strength as I could understand the slang and “Singlish” people use in Singapore and was not perturbed and did not have to ask for clarification like a foreigner (or someone unfamiliar with the slang in Singapore) may have needed. This was an advantage in ensuring the interviews proceeded smoothly without much interruption. Participants also felt at ease speaking to someone with a similar accent. I noted that they were open in sharing their opinions. Even the quieter ones opened up when I asked follow-up questions. I did not sense any hierarchy induced barriers as participants did not explicitly indicate that they were nervous, did not feel their views were unimportant etc. Considering the subjective nature of the data analysis, this was done in parallel between two researchers, myself and my PhD supervisor Assistant Professor of Evidence Based Medicine, Dr Lorainne Tudor Car. The derivation of themes and thematic analysis is described at length in Sections 6.2.3 and 6.3.2. Being supervised by an expert in evidence-based medicine, this was always the priority for both researchers, throughout the projects. I was intentional with grounding our findings with existing evidence and contributing to the evidence base in a stepwise manner. This was prioritised over other things like technological complexity, which was of more interest for some participants in our interviews, for example.
Methodological	This PhD project was approached with a pragmatic paradigm in mind which posits that methods should be selected based on “what works best” for the study aims,

	setting and research group to help answer the research question [81]. Furthermore, the qualitative facets of the project depended on a constructivist epistemology which aims to identify the themes and views set out by the participants and also consult these individuals when aiming to construct answers to our research questions - as was done in the needs assessment in Chapter 6. The semi-structured interview design was selected to guide the conversations towards identifying the needs of the population (Chapter 6) and their views on the CA Precilla (Chapter 8). Other data collection methods certainly exist, such as, fully structured interviews. This would have enabled even more uniformity amongst all interviews, enhancing reliability. However, the rigid nature of such interviews would not have allowed me to elaborate on interesting points raised by participants. For example, some preferred to receive messages from Precilla early in the morning whilst other preferred messages to be sent in the evenings after work. The fluid nature of the interview allowed me to elaborate on this point which revealed that Wi-Fi availability was an important consideration in for their willingness to use the CA. Hence, being at home in the morning or evening would guarantee Wi-Fi availability and no risk of data usage or overuse charges. Alternatively, group-based discussions were a plausible study design eg. Focus group discussions or thinking out loud testing. Based on the literature and research done for the conceptual framework, these group-based methods appeared to be more applicable to the piloting stages prior to initial prototype development. This is something to consider when upgrading Precilla. However, when exploring feasibility of uptake by the population, acceptability and satisfaction, individual interviews allow users to be as honest as possible without susceptibility to any Hawthorne effects.
Contextual	While conducting the interviews and whilst transcribing them, I noted these observations. Some interviews were conducted over the telephone using a “WhatsApp” call which ensured that the call was free and is also a commonly used method of communication in Singapore (as opposed to a normal voice call). In some cases, the WiFi connection was not optimal, and the connection was cut a few times, disrupting the flow of the conversation. The telephone interviews came with the advantages of being remote and convenient (eliminating the need for the participant to travel), it also had its shortcomings such as never getting to meet the research team as well as connection issues. Also, with telephone interviews, and considering this was a purely remote study where study participants never saw the researchers, it is positive in the sense that their anonymity was protected however, they missed out on the eye contact and non-verbal nuances of a live conversation. Perhaps this could have allowed them to be more honest in sharing their opinions as there may be less of a fear of judgement or confrontation. In-person interviews were all conducted in private meeting rooms at the Lee Kong Chain School of Medicine. In this way, participants were reassured that the conversations were private and their opinions and confidentiality would be protected. I picked up on some hesitation from participants when answering the question regarding household income. It potentially came across as a little intrusive and I contemplated whether we really needed this question. However, in line with

	maintaining consistency in all interviews and in the hopes of collating basic demographic information about our participants, this question was retained. In hindsight, I think it would be appropriate to remove the question as it did not subsequently play a major role in our data analysis. It indicated we had a diverse group of household incomes but did not directly impact on our conclusions. If we were going to follow up with questions regarding what proportion of your monthly salary gets allocated to grocery expenses? Or memberships to gyms/other fitness programmes? Purchase of stress management tools or sleep enabling apps? Then this question would be more justified and worth asking despite the discomfort it imbues in the interviewee.

Reflexivity was somewhat relevant to all the studies conducted in the PhD project from the feasibility study, which was almost exclusively objective, to the qualitative studies at the other end of the pole. Figure 10.1 outlines some of these reflections.

Feasibility study	Scoping Review	Apps assessment	Conceptual framework	Interviews
•Data was fully quantitative and collected via surveys •Not open to interpretation	•2 reviewers assigned extracted data into three distinct categories: acceptability, effectiveness and diagnostic accuracy. •Discussions between reviewers helped to eliminate discrepancies and reduce bias	•Assignment of BCTs was done by two reviewers. There was a little ambiguity given that multiple BCTs overlapped and may have been relevant. •Subjectivity was reduced through discussions to explicitly define BCTs and how they would be allocated for this study specifically	•Given that I had no computing or tech background, the codes and themes generated were very relatable to the lay man and non-tech expert researcher making it relevant to our target population.	•Reflexivity outlined in table 10.2

Figure 10.1 Reflexive considerations for studies in this PhD project ordered from most objective (left) to most subjective (right).

10.10 Recommendations for the future

This thesis has focussed on building a strong foundation for the research around healthcare CAs. As it draws to a close, some important points have become apparent and point towards work to be undertaken soon to increase the quality, robustness, and relevance of research published on healthcare CAs. As identified in the apps assessment (chapter 9) and the pilot study (chapter 7)

BCTs are a growing area of research and have the potential to be effective when implemented in CA interventions. This is evidenced by a protocol for a scoping review of BCTs used in CA applications [360] and a behavior change model for the development of AI-based CAs [220]. Future work should look into the ideal combinations of BCTs for the intended purposes to assert effectiveness at a higher level.

Privacy and security issues have been mentioned in several papers relating to CAs, especially AI-based CAs [44, 185]. Issues related to confidentiality need to be explored in future studies, and this point is reiterated by several apps identified in chapter 9 as not having a clear privacy policy made easily available to the user. Personalization for rule-based agents does pose a challenge given the lack of machine learning capabilities but can be worked around. The ability to seamlessly deliver a personalized rule-based CA for healthcare should also be explored in future work.

10.11 Concluding remarks

To conclude, this Ph.D. project has employed a mixed-methods approach informed by a pragmatic paradigm to investigate the use of CAs in healthcare. Scoping methodology helped to provide a broad overview of the current usage applications of CAs in the literature and those smartphone apps readily available for public usage without formal peer-review. This thesis has also combined evidence from the literature with empirical research to develop a healthy lifestyle change CA which was tested for its feasibility and acceptability in Singapore. The feedback was largely positive with some valuable pointers for improvement. Most notably, this Ph.D. project has contributed to filling knowledge gaps concerning guidelines on how to design and develop a rule-based healthcare CA for non-experts. All the evidence, results, and conclusions provided have been generated through rigorous methodology and were ethically and morally generated. Collaborations with other researchers, study team members, and non-experts were pivotal in providing a holistic view and in ensuring the success of this project. This Ph.D. project was embedded in the tenets of evidence-based medicine, tapped into existing evidence for guidance, and contributed generously to the evidence base as well.

References

1. Administration USFaD. What is Digital Health? 2020; Available from: <https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health>.
2. Sophie P, Javier G-P, Andrew C, Zsuzsanna V. From treatment to prevention: The evolution of digital healthcare. <https://media.nature.com/original/magazine-assets/d42473-019-00274-6/d42473-019-00274-6.pdf>.
3. Consulting A. Today's consumers reveal the future of healthcare2019. Available from: https://www.accenture.com/_acnmedia/PDF-99/Accenture-2019-Digital-Health-Consumer-Survey-AU.pdf.
4. Holman T. DEFINITION - mHealth (mobile health). Tech Target2018 [cited 2022]; Available from: <https://www.techtarget.com/searchhealthit/definition/mHealth>.
5. Dingler T, Kwasnicka D, Wei J, Gong E, Oldenburg B. The Use and Promise of Conversational Agents in Digital Health. Yearbook of Medical Informatics. 2021;30(01):191-9.
6. Dictionaries OL. Chatbot. Oxford Living Dictionaries2018 [cited 2018]; Available from: <https://en.oxforddictionaries.com/definition/chatbot>. Archived at <http://www.webcitation.org/746pRgbsi>.
7. Veretskaya O. What is a chatbot and how to use it for your business. In: Anadea, editor. Anadea2017.
8. Press A. Smart speaker sales more than tripled in 2017. Billboard2017 [cited 2018]; Available from: <https://www.billboard.com/articles/business/8085524/smart-speaker-sales-tripled-25-million-year-2017>. Archived at <http://www.webcitation.org/746pXZ3d7>.
9. Perez S. 39 million Americans now own a smart speaker, report claims. Tech Crunch2018 [cited 2018]; Available from: <https://techcrunch.com/2018/01/12/39-million-americans-now->

[own-a-smart-speaker-report-claims/](#). Archived at <http://www.webcitation.org/746puKdDT>.

10. Markets Ra. Smart speakers 2018: World Market Forecast to 2023 - Key Vendors covered include Alphabet, Amazon, Harman International, Alibaba and Sonos. Research and Markets. 2018.

11. Saeed H. Developing a chatbot? Learn the difference between AI, Machine Learning, and NLP. Chatbots life2016.

12. Card D. The “black box” metaphor in machine learning2017. Available from: <https://dallascard.medium.com/the-black-box-metaphor-in-machine-learning-4e57a3a1d2b0>.

13. Peiris D, Miranda JJ, Mohr DC. Going beyond killer apps: building a better mHealth evidence base. BMJ Global Health. 2018;3(1):e000676. doi: 10.1136/bmjgh-2017-000676.

14. Scott-Sheldon LAJ, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text Messaging-Based Interventions for Smoking Cessation: A Systematic Review and Meta-Analysis. JMIR mHealth and uHealth. 2016;4(2):e49-e. PMID: 27207211. doi: 10.2196/mhealth.5436.

15. Rubinstein A, Miranda JJ, Beratarrechea A, Diez-Canseco F, Kanter R, Gutierrez L, et al. Effectiveness of an mHealth intervention to improve the cardiometabolic profile of people with prehypertension in low-resource urban settings in Latin America: a randomised controlled trial. The Lancet Diabetes & Endocrinology. 2016;4(1):52-63.

16. Patel A. PPC Chatbots Explained. Acquisio: Acquisio; 2019 [cited 2021]; Available from: <https://www.acquisio.com/blog/innovation/ppc-chatbots-explained/>.

17. Microsoft. ReDial: Recommendation dialogs for bridging the gap between chit-chat and goal-oriented chatbots. Microsoft research blog2018 [cited 2021]; Available from: https://www.microsoft.com/en-us/research/blog/redial-recommendation-dialogs-for-bridging-the-gap-between-chit-chat-and-goal-oriented-chatbots/?OCID=msr_blog_redial_neurips_tw.

18. Schuetzler RM, Giboney JS, Grimes GM, Nunamaker Jr JF. The influence of conversational agent embodiment and conversational relevance on socially desirable responding. *Decision Support Systems*. 2018;114:94-102.
19. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun ACM*. 1966;9(1):36-45. doi: 10.1145/365153.365168.
20. Deryugina OV. Chatterbots. *Scientific and Technical Information Processing*. 2010 April 01;37(2):143-7. doi: 10.3103/s0147688210020097.
21. Futurism. The history of chatbots. *Futurism*2016 [cited 2018]; Available from: <https://futurism.com/images/the-history-of-chatbots-infographic/>. Archived at <http://www.webcitation.org/746qKMpDn>.
22. Colby KM. Artificial paranoia: A computer simulation of paranoid processes: Elsevier; 2013. ISBN: 1483153266.
23. McTear M, Callejas Z, Griol D. *The Conversational Interface*. Switzerland: Springer International Publishing; 2016.
24. Letting fingers do the talking. Computer makes patient satisfaction surveys a snap. *Rose Medical Center, Denver, CO. Profiles in healthcare marketing*. 1992 Jul-Aug(48):40-4. PMID: 10120010.
25. Delichatsios HK, Friedman RH, Glanz K, Tennstedt S, Smigelski C, Pinto BM, et al. Randomized trial of a "talking computer" to improve adults' eating habits. *American journal of health promotion : AJHP*. 2001 Mar-Apr;15(4):215-24. PMID: 11349340. doi: 10.4278/0890-1171-15.4.215.
26. Friedman E. Your friendly neighborhood diagnosis-aiding talking computer. *Hospitals*. 1981 Oct 16;55(20):105-6, 8, 13-4. PMID: 7275074.
27. Friedman RB, Newsom RS, Entine SM, Cheung S, Schultz JV. A simulated patient-physician encounter using a talking computer. *Jama*. 1977 Oct 31;238(18):1927-9. PMID: 578551.

28. Migneault JP, Farzanfar R, Wright JA, Friedman RH. How to write health dialog for a talking computer. Journal of biomedical informatics. 2006 Oct;39(5):468-81. PMID: 16564749. doi: 10.1016/j.jbi.2006.02.009.
29. Haider A. Microsoft's Cortana to focus on enterprise, leaving smart home to Google, Amazon2020. Available from: <https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/microsoft-s-cortana-to-focus-on-enterprise-leaving-smart-home-to-google-amazon-57417892>.
30. Mayo J. 2016: The year when chatbots were hot. 2016.
31. Bruner J. Why 2016 is shaping up to be the Year of the Bot. O'Reilly2016 [cited 2018]; Available from: <https://www.oreilly.com/ideas/why-2016-is-shaping-up-to-be-the-year-of-the-bot>. Archived at <http://www.webcitation.org/746q0alqL>.
32. Aspect. The year of chatbots for business. Aspect blogs2016.
33. Jack A. Rwanda venture tests digital health potential in developing world2021. Available from: <https://www.ft.com/content/4fe33c92-cbd5-459a-8df6-20d0d1f57ec8>.
34. Fitzpatrick KK, Darcy A, Vierhile M. 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 Mental Health. 2017;4(2):e19. PMID: PMC5478797. doi: 10.2196/mental.7785.
35. Kleinman A. This App Thinks Conversations Are The Key To Better Fitness And Sleep2015. Available from: https://www.huffpost.com/entry/lark-app_n_7152178?guccounter=1.
36. Department SR. Mobile messenger and communication apps - Statistics and Facts. Statista2021; Available from: <https://www.statista.com/topics/1523/mobile-messenger-apps/#dossierKeyfigures>.

37. department Sr. Most popular global mobile messenger apps as of October 2021, based on number of monthly active users. Statista2021; Available from: <https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/>.
38. Neville R, Greene A, McLeod J, Tracey A, Surie J. Mobile phone text messaging can help young people manage asthma. *BMJ*. 2002;325(7364):600. PMID: 12228151. doi: 10.1136/bmj.39552.645775.AD.
39. Yeager VA, Menachemi N. Text messaging in health care: a systematic review of impact studies. *Advances in health care management*. 2011;11:235-61. PMID: 22908672.
40. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile Text Messaging for Health: A Systematic Review of Reviews. *Annual Review of Public Health*. 2015;36(1):393-415. PMID: 25785892. doi: 10.1146/annurev-publhealth-031914-122855.
41. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*. 2018;25(9):1248-58.
42. Car LT, Dhinakaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y-L, et al. Conversational agents in health care: scoping review and conceptual analysis. *Journal of medical Internet research*. 2020;22(8):e17158.
43. Lomas N. UK's MHRA says it has 'concerns' about Babylon Health — and flags legal gap around triage chatbots2021. Available from: <https://techcrunch.com/2021/03/05/uks-mhra-says-it-has-concerns-about-babylon-health-and-flags-legal-gap-around-triage-chatbots/#:~:text=In%20the%20summary%20of%20his,top%20Dra ted%20practicing%20clinicians.%E2%80%9D>.
44. Reuter E. Babylon Health, physician tussle over triage chatbot's safety2020. Available from: <https://medcitynews.com/2020/02/babylon-physician-tussle-over-triage-chatbots-safety/>.

45. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. *Journal of medical Internet research*. 2019;21(4):e12887.
46. Salleh NAM. Singapore a rare, precious example of harmonious multiracial, multi-religious society: PM Lee2017. Available from: <https://www.straitstimes.com/singapore/singapore-a-rare-and-precious-example-of-a-harmonious-multi-racial-multi-religious-society>.
47. Gardens SB. Singapore's First UNESCO World Heritage Site. Singapore Botanic Gardens2020; Available from: <https://www.nparks.gov.sg/sbg/about/singapore-s-first-unesco-world-heritage-site>.
48. Poh J. How Does the Singapore Healthcare System Work Anyway?2021. Available from: <https://blog.moneysmart.sg/healthcare/singapore-healthcare-public-private-hospitals/>.
49. Haseltine WA. Affordable excellence: the Singapore healthcare story: how to create and manage sustainable healthcare systems: Brookings Institution Press; 2013. ISBN: 0815724160.
50. Yong LMO, Koe LWP. War on Diabetes in Singapore: a policy analysis. *Health Research Policy and Systems*. 2021;19(1):1-10.
51. Singapore Sn. National Steps Challenge™ & Healthy 365 App. 2021; Available from: <https://www.smartnation.gov.sg/initiatives/health/national-steps-challenge>.
52. Hub H. Healthy Hawker Food. 2020; Available from: <https://www.healthhub.sg/live-healthy/24/healthyhawkerfood>.
53. Khim KB. Singapore is 'most fatigued' country in the world – and we're tired of it2021. Available from: <https://cnalifestyle.channelnewsasia.com/wellness/singapore-most-fatigued-country-world-we-are-tired-278311>.

54. Federation ID. IDF Atlas - 8th edition 2017. Available from: <https://idf.org/our-network/regions-members/western-pacific/members/113-singapore.html>.
55. Lai L. One in three Singaporeans not active enough. The Straits Times. 2018.
56. Dale AC, Nilsen TI, Vatten L, Midthjell K, Wiseth R. Diabetes mellitus and risk of fatal ischaemic heart disease by gender: 18 years follow-up of 74 914 individuals in the HUNT 1 Study. *Eur Heart J*. 2007 Dec;28(23):2924-9. PMID: WOS:000251506200020. doi: 10.1093/eurheartj/ehm447.
57. Gu K, Cowie CC, Harris MI. Mortality in adults with and without diabetes in a national cohort of the US population, 1971-1993. *Diabetes Care*. 1998 Jul;21(7):1138-45. PMID: WOS:000074416800020. doi: DOI 10.2337/diacare.21.7.1138.
58. Wong LY, Toh MPHS, Tham LWC. Projection of prediabetes and diabetes population size in Singapore using a dynamic Markov model. *J Diabetes*. 2017 Jan;9(1):65-75. PMID: WOS:000390565200009. doi: 10.1111/1753-0407.12384.
59. Tabak AG, Herder C, Rathmann W, Brunner EJ, Kivimaki M. Prediabetes: a high-risk state for diabetes development. *Lancet*. 2012 Jun 16;379(9833):2279-90. PMID: 22683128. doi: 10.1016/S0140-6736(12)60283-9.
60. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *The New England journal of medicine*. 2001;344(18):1343-50. PMID: 11333990 doi: 10.1056/NEJM200105033441801.
61. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *The New England journal of medicine*. 2002;346(6):393-403. PMID: 11832527. doi: 10.1056/NEJMoa012512.
62. Ramachandran A, Snehalatha C, Mary S, Mukesh B, Bhaskar AD, Vijay V. The Indian Diabetes Prevention Programme

shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). *Diabetologia*. 2006;49(2):289-97. PMID: 16391903. doi: 10.1007/s00125-005-0097-z.

63. Al-Maskari F. Lifestyle diseases: An economic burden on the health services. *UN Chronicle The Magazine of the United Nations*. 2010;5:1-2.

64. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS medicine*. 2006;3(11):e442.

65. Ford ES, Bergmann MM, Kröger J, Schienkiewitz A, Weikert C, Boeing H. Healthy living is the best revenge: findings from the European Prospective Investigation Into Cancer and Nutrition-Potsdam study. *Archives of internal medicine*. 2009;169(15):1355-62.

66. Singapore G. History of internet in Singapore – from niche toy to must-have essential 2021. Available from:

<https://www.tech.gov.sg/media/technews/history-of-the-internet>.

67. Baker JA. Chatbots, robotics to help ease manpower crunch in healthcare sector. *Channel News Asia*. 2018 17 Jul 2018.

68. times Ts. Chatbots show the way in using tech to boost healthcare. *The straits times* 2018 [cited 2021]; Available from:

https://www.ihis.com.sg/Latest_News/News_Article/Pages/Chatbots-show-the-way-in-using-tech-to-boost-healthcare-.aspx.

69. SingHealth. SingHealth and A*STAR co-develop smart chatbot to enhance care for COVID-19 patients at community care facilities 2020. Available from:

<https://www.singhealth.com.sg/news/patient-care/doctor-covid-chatbot>.

70. Bhattacharya A. Four ways Singapore is redesigning patients' experience. Available from:

<https://govinsider.asia/innovation/singapore-patient-experience-ihis-chatbot-virtual-reality/>.

71. Han GY. New 'emotionally intelligent' chatbot to help Singaporeans stressed by pandemic 2021. Available from:

<https://www.straitstimes.com/singapore/community/new-emotionally-intelligent-chatbot-to-help-singaporeans-stressed-by-pandemic>.

72. DBS. digibot Services. DBS.com2019; Available from: <https://www.dbs.com.sg/personal/support/general-digital-assistance.html>.

73. Lim J. 'Bus Uncle' is the most Singaporean bus arrival timing app ever2016. Available from: <https://mothership.sg/2016/10/bus-uncle-is-the-most-singaporean-bus-arrival-timing-app-ever/>.

74. Spectrm. Messaging App Statistics: The Most Popular Communication Method Globally in 2020 & Beyond2020. Available from: <https://spectrm.io/insights/blog/messaging-app-statistics-most-popular-communication-method-2020/>.

75. Kobori Y, Osaka A, Soh S, Okada H. NOVEL APPLICATION FOR SEXUAL TRANSMITTED INFECTION SCREENING WITH AN AI CHATBOT. The Journal of Urology. 2018 2018/04/01/;199(4, Supplement):e189-e90. doi: <https://doi.org/10.1016/j.juro.2018.02.516>.

76. Masic I, Miokovic M, Muhamedagic B. Evidence based medicine–new approaches and challenges. Acta Informatica Medica. 2008;16(4):219.

77. Oliver D. David Oliver: Lessons from the Babylon Health saga. BMJ. 2019;365.

78. Klassen AC, Creswell J, Clark VLP, Smith KC, Meissner HI. Best practices in mixed methods for quality of life research. Quality of life Research. 2012;21(3):377-80.

79. Caffery LJ, Martin-Khan M, Wade V. Mixed methods for telehealth research. Journal of telemedicine and telecare. 2017;23(9):764-9.

80. Tran T. Pragmatism: How to connect positivism and constructivism in doing research. Retrieved from webQDA: <https://www.webqda.net/pragmatism-how-to-...>; 2017.

81. Creswell JW. Mapping the field of mixed methods research. SAGE publications Sage CA: Los Angeles, CA; 2009. p. 95-108.

82. Greene JC, Caracelli VJ. Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms: Jossey-Bass; 1997. ISBN: 0787998222.
83. Morgan DL. Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of mixed methods research*. 2007;1(1):48-76.
84. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS medicine*. 2010;7(9):e1000326.
85. Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. *Evaluation & the health professions*. 2002;25(1):12-37.
86. Jordan Z, Munn Z, Aromataris E, Lockwood C. Now that we're here, where are we? The JBI approach to evidence-based healthcare 20 years on. *JBI Evidence Implementation*. 2015;13(3):117-20.
87. Aromataris E, Pearson A. The systematic review: an overview. *AJN The American Journal of Nursing*. 2014;114(3):53-8.
88. Julian PT Higgins, Green S. Cochrane Handbook for Systematic Reviews of Interventions. 2011. In: *Cochrane Handbook for Systematic Reviews of Interventions* [Internet]. The Cochrane Collaboration: The Cochrane Collaboration. Available from: <https://handbook-5-1.cochrane.org/>.
89. Pearson A, Jordan Z, Munn Z. Translational science and evidence-based healthcare: a clarification and reconceptualization of how knowledge is generated and used in healthcare. *Nursing research and practice*. 2012;2012.
90. Anderson S, Allen P, Peckham S, Goodwin N. Asking the right questions: scoping studies in the commissioning of research on the organisation and delivery of health services. *Health research policy and systems*. 2008;6(1):1-12.

91. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *International journal of social research methodology*. 2005;8(1):19-32.
92. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC medical research methodology*. 2018;18(1):1-7.
93. Yin H, Wardenaar KJ, Wang Y, Wang N, Chen W, Zhang Y, et al. Mobile Mental Health Apps in China: Systematic App Store Search. *Journal of Medical Internet Research*. 2020;22(7):e14915.
94. Flors-Sidro JJ, Househ M, Abd-Alrazaq A, Vidal-Alaball J, Fernandez-Luque L, Sanchez-Bocanegra CL. Analysis of Diabetes Apps to Assess Privacy-Related Permissions: Systematic Search of Apps. *JMIR diabetes*. 2021;6(1):e16146.
95. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. *JMIR mHealth and uHealth*. 2015;3(2):e4283.
96. Huckvale K, Adomaviciute S, Prieto JT, Leow MK-S, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. *BMC medicine*. 2015;13(1):1-10.
97. Martinengo L, Van Galen L, Lum E, Kowalski M, Subramaniam M, Car J. Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. *BMC medicine*. 2019;17(1):1-12.
98. Van Galen L, Xu X, Koh M, Thng S, Car J. Eczema apps conformance with clinical guidelines: a systematic assessment of functions, tools and content. *British Journal of Dermatology*. 2020;182(2):444-53.
99. Lum E, Jimenez G, Huang Z, Thai L, Semwal M, Boehm BO, et al. Decision support and alerts of apps for self-management of blood glucose for type 2 diabetes. *Jama*. 2019;321(15):1530-2.

100. Team H. The mHONcode guidelines. Team HON2020 [cited 2021 26/10/2021]; Available from: <https://www.hon.ch/en/guidelines-mhoncode.html>.
101. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Annals of behavioral medicine*. 2013;46(1):81-95.
102. Jabareen Y. Building a Conceptual Framework: Philosophy, Definitions, and Procedure. 2009.
103. Miles MB, Huberman AM. Qualitative data analysis: An expanded sourcebook: sage; 1994. ISBN: 0803955405.
104. Leamy M, Bird V, Le Boutillier C, Williams J, Slade M. Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *The British Journal of Psychiatry*. 2011;199(6):445-52.
105. Ndlovu K, Mars M, Scott RE. Development of a conceptual framework for linking mHealth applications to eRecord systems in Botswana. *BMC Health Serv Res*. 2021;21(1):1-10.
106. Bauer MS, Kirchner J. Implementation science: What is it and why should I care? *Psychiatry research*. 2020;283:112376.
107. Colditz GA, Emmons KM. The promise and challenges of dissemination and implementation research. *Dissemination and implementation research in health: Translating science to practice*. 2012;2:1-17.
108. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. *American journal of preventive medicine*. 2009;36(5):452-7. PMID: 19362699. doi: 10.1016/j.amepre.2009.02.002.
109. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *bmj*. 2021;374.
110. Good A, Omisade O. Linking Activity Theory with User Centred Design: A Human Computer Interaction Framework for

the Design and Evaluation of. Applied Interdisciplinary Theory in Health Informatics: A Knowledge Base for Practitioners. 2019;263:49.

111. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Diequez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). Journal of biomedical informatics. 2016;60:243-51.

112. DeJonckheere M, Vaughn LM. Semistructured interviewing in primary care research: a balance of relationship and rigour. Family Medicine and Community Health. 2019;7(2).

113. Warren K. Qualitative Data Analysis Methods 101: The “Big 6” Methods + Examples. GradCoach2020; Available from: <https://gradcoach.com/qualitative-data-analysis-methods/>.

114. Caulfield J. How to do thematic analysis. Scribbr2021; Available from: <https://www.scribbr.com/methodology/thematic-analysis/>.

115. Tropea P, Schlieter H, Sterpi I, Judica E, Gand K, Caprino M, et al. Rehabilitation, the great absentee of virtual coaching in medical care: scoping review. Journal of medical Internet research. 2019;21(10):e12805.

116. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. Journal of medical Internet research. 2019;21(11):e15360.

117. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. Journal of medical Internet research. 2017;19(5):e151.

118. Pereira J, Díaz Ó. Using health chatbots for behavior change: a mapping study. Journal of medical systems. 2019;43(5):1-13.

119. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. Journal of medical Internet research. 2017;19(8):e7023.

120. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *The Canadian Journal of Psychiatry*. 2019;64(7):456-64.
121. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. *Expert Systems with Applications*. 2019;129:56-67.
122. Xing Z, Yu F, Du J, Walker JS, Paulson CB, Mani NS, et al. Conversational interfaces for health: bibliometric analysis of grants, publications, and patents. *Journal of medical Internet research*. 2019;21(11):e14672.
123. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine*. 2018;169(7):467-73.
124. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. *JBI Evidence Implementation*. 2015;13(3):141-6.
125. Chatbots.org. Embodied Conversational Agent. Available from: https://www.chatbots.org/embodied_conversational_agent/.
126. Bickmore TW. Relational Agents Northeastern University; Available from: <https://www.ccs.neu.edu/home/bickmore/agents/>.
127. de Haan H, Snijder J, van Nimwegen C, Beun RJ. Chatbot Personality and Customer Satisfaction. 2018.
128. Kowatsch T, Nißen M, Shih C-H, Rüegger D, Volland D, Filler A, et al. Text-based Healthcare Chatbots Supporting Patient and Health Professional Teams: Preliminary Results of a Randomized Controlled Trial on Childhood Obesity 2017.
129. Ballati F, Corno F, De Russis L. "Hey Siri, do you understand me?": Virtual Assistants and Dysarthria 2018.
130. Vita S, Marocco R, Pozzetto I, Morlino G, Vigilante E, Palmacci V, et al. The "doctor Apollo" chatbot: A digital health tool to improve engagement of people living with HIV. *Journal of the International AIDS Society*. 2018 October;21 (Supplement 8):67-8.

131. Casas J, Mugellini E, Khaled OA, editors. Food diary coaching chatbot. Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers; 2018.
132. Denecke K, Hochreutener SL, Pöpel A, May R. Self-Anamnesis with a Conversational User Interface: Concept and Usability Study. *Methods of Information in Medicine*. 2018;57(5-6):243-52. doi: 10.1055/s-0038-1675822.
133. Heldt K, Buchter DJ, Brogle B, Shih CI, Ruegger D, Filler A, et al. Telemedicine Therapy for Overweight Adolescents: First Results of a Novel Smartphone App Intervention using a Behavioural Health Platform. *Obesity Facts*. 2018 May;11 (Supplement 1):214-5.
134. Kowatsch T, Volland D, Shih I, Rüegger D, Künzler F, Barata F, et al., editors. Design and Evaluation of a Mobile Chat App for the Open Source Behavioral Health Intervention Platform MobileCoach. *International Conference on Design Science Research in Information Systems*; 2017: Springer.
135. L'Allemand D, Shih C, Heldt K, Buchter D, Brogle B, Ruegger D. Design and interim evaluation of a smartphone app for overweight adolescents using a behavioural health intervention platform. *Obesity Reviews*. 2018 (Suppl 1):102.
136. NCT03384550. Can a Smartphone App That Includes a Chatbot-based Coaching and Incentives Increase Physical Activity in Healthy Adults? 2017.
137. Stasinaki A, Brogle B, Buchter D, Shih CHI, Heldt K, White C, et al. A novel digital health intervention improves physical performance in obese youth. *Swiss Medical Weekly*. 2018 May;147 (Supplement 228):10S.
138. Chaix B, Bibault J-E, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When chatbots meet patients: a one-year retrospective study of conversations between patients with breast cancer and a chatbot. (Preprint)2018.

139. Nct. Artificial Intelligence vs Physicians for Breast Cancer Patients' Information. <https://clinicaltrials.gov/show/nct03556813>. 2018. PMID: CN-01660339.
140. Lobo J, Ferreira L, Ferreira AJ. CARMIE: A Conversational Medication Assistant for Heart Failure. *International Journal of E-Health and Medical Communications (IJEHMC)*. 2017;8(4):21-37.
141. Crutzen R, Peters GJY, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: An exploratory study. *Journal of Adolescent Health*. 2011 May;48(5):514-9. PMID: 51210569. doi: <http://dx.doi.org/10.1016/j.jadohealth.2010.09.002>.
142. Gaffney H, Mansell W, Edwards R, Wright J. Manage Your Life Online (MYLO): A Pilot Trial of a Conversational Computer-Based Intervention for Problem Solving in a Student Sample. *Behavioural and Cognitive Psychotherapy*. 2013;42(6):731-46. doi: 10.1017/S135246581300060X.
143. Middleton K, Butt M, Hammerla N, Hamblin S, Mehta K, Parsa A. Sorting out symptoms: design and evaluation of the 'babylon check' automated triage system. *arXiv preprint arXiv:160602041*. 2016.
144. Nadarzynski T, Bayley J, Llewellyn C, Graham C, Kidsley S. The acceptability of artificial intelligence (AI)-led chatbot, an automated advice system for sexual health. *HIV Medicine*. 2018 April;19 (Supplement 2):S109.
145. Razzaki S, Baker A, Perov Y, Middleton K, Baxter J, Mullarkey D, et al. A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. *arXiv preprint arXiv:180610698*. 2018.
146. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*. 2018;6(11):e12106.
147. Griol D, Callejas Z. Mobile Conversational Agents for Context-Aware Care Applications. *Cognitive Computation*. 2016 April 01;8(2):336-56. doi: 10.1007/s12559-015-9352-x.

148. Griol D, Molina J, Callejas Z. Towards Emotionally Sensitive Conversational Interfaces for E-therapy 2015. 498-507 p. ISBN: 978-3-319-18913-0.
149. Ly KH, Ly A-M, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions*. 2017 2017/12/01/;10:39-46. doi: <https://doi.org/10.1016/j.invent.2017.10.002>.
150. Comendador BEV, Francisco BMB, Medenilla JS, Mae S. Pharmabot: a pediatric generic medicine consultant chatbot'. *Journal of Automation and Control Engineering* Vol. 2015;3(2).
151. Huang J, Li Q, Xue Y, Cheng T, Xu S, Jia J, et al. Teenchat: A chatterbot system for sensing and releasing adolescents stress. *Lect Notes Comput Sci Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2015;9085:133-45.
152. Kamita T, Ito T, Matsumoto A, Munakata T, Inoue T. A chatbot system for mental healthcare based on SAT counseling method. *Mobile Information Systems*. 2019;2019.
153. Mujeeb S, Javed MH, Arshad T. Aquabot: A Diagnostic Chatbot for Achluophobia and Autism. *International Journal Of Advanced Computer Science And Applications*. 2017;8(9):209-16.
154. Shaikh M, Shah A, Chawla R, Makkar BM, Kesavadev J, Joshi S, et al. Real-world effectiveness of a digital therapeutic in improving glycaemic control in south Asians living with type 2 diabetes. *Diabetes*. 2018 July;67 (Supplement 1):A225.
155. Danda P, Srivastava BML, Shrivastava M, editors. Vaidya: A Spoken Dialog System for Health Domain. *Proceedings of the 13th International Conference on Natural Language Processing*; 2016.
156. Wang H, Zhang Q, Ip M, Lau JTF. Social Media-based Conversational Agents for Health Management and Interventions. *Computer*. 2018;51(8):26-33.

157. Elmasri D, Maeder A, editors. A conversational agent for an online mental health intervention. International Conference on Brain Informatics; 2016: Springer.
158. Ghosh S, Bhatia S, Bhatia A. Quro: Facilitating User Symptom Check Using a Personalised Chatbot-Oriented Dialogue System. Studies in health technology and informatics. 2018;252:51-6. PMID: 30040682.
159. Joerin A, Rauws M, Ackerman ML. Psychological Artificial Intelligence Service, Tess: Delivering On-demand Support to Patients and Their Caregivers: Technical Report. Cureus. 2019 Jan 28;11(1):e3972. PMID: 30956924. doi: 10.7759/cureus.3972.
160. Wilson N, MacDonald EJ, Mansoor OD, Morgan J. In bed with Siri and Google Assistant: a comparison of sexual health advice. BMJ. 2017;359. doi: 10.1136/bmj.j5635.
161. Ni L, Lu C, Liu N, Liu J, editors. Mandy: Towards a smart primary care chatbot application. International Symposium on Knowledge and Systems Sciences; 2017: Springer.
162. van Heerden A, Ntinga X, Vilakazi K, editors. The potential of conversational agents to provide a rapid HIV counseling and testing services. 2017 International Conference on the Frontiers and Advances in Data Science (FADS); 2017: IEEE.
163. Cheng A, Raghavaraju V, Kanugo J, Handrianto YP, Shang Y, editors. Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes. 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC); 2018: IEEE.
164. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Mental Health.5(4):e64. PMID: 30545815.
165. Galescu L, Allen J, Ferguson G, Quinn J, Swift M, editors. Speech recognition in a dialog system for patient health monitoring. Proceedings - 2009 IEEE International Conference on Bioinformatics and Biomedicine Workshops, BIBMW 2009; 2009.

166. Liu B, Sundar SS. Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot. *Cyberpsychology, behavior and social networking*. 2018;21(10):625-36. PMID: 30334655.
167. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. *JAMA internal medicine*. 2016 May 1;176(5):619-25. PMID: 26974260. doi: 10.1001/jamainternmed.2016.0400.
168. Rhee H, Allen J, Mammen J, Swift M. Mobile phone-based asthma self-management aid for adolescents (mASMAA): a feasibility study. Patient preference and adherence. 2014;8:63-72. PMID: 24470755. doi: 10.2147/ppa.S53504.
169. Stein N, Brooks K. A Fully Automated Conversational Artificial Intelligence for Weight Loss: Longitudinal Observational Study Among Overweight and Obese Adults. *JMIR diabetes*. 2017 Nov 1;2(2):e28. PMID: 30291087. doi: 10.2196/diabetes.8590.
170. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. *Journal of medical Internet research*. 2018;20(9):e11510.
171. Ghandeharioun A, McDuff D, Czerwinski M, Rowan K. EMMA: An Emotionally Intelligent Personal Assistant for Improving Wellbeing. *arXiv preprint arXiv:1812.11423*. 2018.
172. Kocielnik R, Xiao L, Avrahami D, Hsieh G. Reflection companion: a conversational system for engaging users in reflection on physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. 2018;2(2):1-26.
173. Kowatsch T, Nißen M, Rügger D, Stieger M, Flückiger C, Allemand M, et al. The impact of interpersonal closeness cues in

text-based healthcare chatbots on attachment bond and the desire to continue interacting: an experimental design. 2018.

174. Fitzpatrick KK, Darcy A, Vierhile M. 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*. 2017 Jun 6;4(2):e19. PMID: 28588005. doi: 10.2196/mental.7785.

175. Ghosh S, Bhatia S, Bhatia A. Quro: Facilitating User Symptom Check Using a Personalised Chatbot-Oriented Dialogue System. *Studies in Health Technology & Informatics*. 2018;252:51-6. PMID: 30040682.

176. Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth and uHealth*. 2018 Nov 23;6(11):e12106. PMID: 30470676. doi: 10.2196/12106.

177. Comendador BEV, Francisco BMB, Medenilla JS, Mae S. Pharmabot: a pediatric generic medicine consultant chatbot. *Journal of Automation and Control Engineering* Vol. 2015;3(2).

178. Lobo J, Ferreira L, Ferreira AJ. CARMIE: A conversational medication assistant for heart failure. *International Journal of E-Health and Medical Communications*. 2017;8(4):21-37. doi: 10.4018/IJEHMC.2017100102.

179. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health. *JAMA Internal Medicine*. 2016 05 01;176(5):619-25. PMID: 26974260. doi: <https://dx.doi.org/10.1001/jamainternmed.2016.0400>.

180. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. *Journal of medical Internet research*. 2018 Sep 4;20(9):e11510. PMID: 30181110. doi: 10.2196/11510.

181. Blog Editor M. ReDial: Recommendation dialogs for bridging the gap between chit-chat and goal-oriented chatbots. Microsoft; 2018.
182. Onwumere J, Amaral F, Valmaggia LR. Digital Technology for Caregivers of People With Psychosis: Systematic Review. *JMIR Mental Health*.5(3):e55. PMID: 30185402.
183. Ruggiano N, Brown EL, Li J, Scaccianoce M. Rural Dementia Caregivers and Technology: What Is the Evidence? *Research in Gerontological Nursing*. 2018;11(4):216-24. PMID: 30036405. doi: 10.3928/19404921-20180628-04.
184. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. *PloS one*. 2018;13(7):e0201166-e. PMID: 30048546. doi: 10.1371/journal.pone.0201166.
185. Coiera E. Paper Review: the Babylon Chatbot. *WordPress.com*2018.
186. Thompson D, Baranowski T. Chatbots as extenders of pediatric obesity intervention: an invited commentary on “Feasibility of Pediatric Obesity & Pre-Diabetes Treatment Support through Tess, the AI Behavioral Coaching Chatbot”. *Translational behavioral medicine*. 2019;9(3):448-50.
187. D'Alfonso S, Santesteban-Echarri O, Rice S, Wadley G, Lederman R, Miles C, et al. Artificial Intelligence-Assisted Online Social Therapy for Youth Mental Health. *Frontiers in psychology*. 2017;8:796. PMID: 28626431. doi: 10.3389/fpsyg.2017.00796.
188. Mascitti I, Feituri M, Funghi F, Correnti S, Galassi LA, editors. COACH BOT: Modular e-course with virtual coach tool support. *ICAART 2010 - 2nd International Conference on Agents and Artificial Intelligence, Proceedings*; 2010.
189. Amrita DB. Health care social media: expectations of users in a developing country. *Medicine* 20. 2013;2(2). PMID: 25075239.
190. Kerr D, Axelrod C, Hoppe C, Klonoff DC. Diabetes and technology in 2030: a utopian or dystopian future? *Diabetic*

- medicine : a journal of the British Diabetic Association. 2018 Apr;35(4):498-503. PMID: 29356078. doi: 10.1111/dme.13586.
191. Garcia DMP, Lopez SS, Donis H, editors. Voice activated virtual assistants personality perceptions and desires: comparing personality evaluation frameworks. Proceedings of the 32nd International BCS Human Computer Interaction Conference 32; 2018.
 192. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. Journal of the American Medical Informatics Association. 2018;25(8):1089-98.
 193. Fedele DA, McConville A, Moon J, Thomas JG. Topical review: design considerations when creating pediatric mobile health interventions: applying the IDEAS framework. Journal of pediatric psychology. 2019;44(3):343-8.
 194. Denecke K, Warren J. How to Evaluate Health Applications with Conversational User Interface? Studies in health technology and informatics. 2020;270:976-80.
 195. Jennings HM, Morrison J, Akter K, Kuddus A, Ahmed N, Kumer Shaha S, et al. Developing a theory-driven contextually relevant mHealth intervention. Global health action. 2019;12(1):1550736.
 196. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mHealth interventions: examples from New Zealand. Journal of health communication. 2012;17(sup1):11-21.
 197. Medicines T. What do all these 'health'-terms actually mean? Talking medicines: Listen to what your patients are saying 2017 [cited 2021]; Available from: <https://talkingmedicines.com/2017/03/digital-health-terms-ehealth-mhealth-telehealth-telemedicine/>.
 198. Dhinakaran DA, Sathish T, Soong A, Theng Y-L, Best J, Car LT. Conversational Agent for Healthy Lifestyle Behavior Change: Web-Based Feasibility Study. JMIR Formative Research. 2021;5(12):e27956.

199. Depp C, Torous J, Thompson W. Technology-based early warning systems for bipolar disorder: a conceptual framework. *JMIR mental health*. 2016;3(3):e42.
200. Direito A, Walsh D, Hinbarji M, Albatal R, Tooley M, Whittaker R, et al. Using the intervention mapping and behavioral intervention technology frameworks: development of an mHealth intervention for physical activity and sedentary behavior change. *Health Education & Behavior*. 2018;45(3):331-48.
201. Horvath KJ, Ecklund AM, Hunt SL, Nelson TF, Toomey TL. Developing Internet-based health interventions: a guide for public health researchers and practitioners. *Journal of medical Internet research*. 2015;17(1):e28.
202. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. *Journal of medical Internet research*. 2014;16(6):e146.
203. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. *Journal of medical Internet research*. 2016;18(12):e317.
204. Wittenberg E, Xu J, Goldsmith J, Mendoza Y. Caregiver communication about cancer: development of a mhealth resource to support family caregiver communication burden. *Psycho - oncology*. 2019;28(2):365-71.
205. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. *Journal of medical Internet research*. 2015;17(1):e30.
206. Duff O, Walsh D, Malone S, McDermott L, Furlong B, O'Connor N, et al. MedFit app, a behavior-changing, theoretically informed mobile app for patient self-management of cardiovascular disease: user-centered development. *JMIR formative research*. 2018;2(1):e8.

207. Pham Q, Graham G, Lalloo C, Morita PP, Seto E, Stinson JN, et al. An analytics platform to evaluate effective engagement with pediatric mobile health apps: design, development, and formative evaluation. *JMIR mHealth and uHealth*. 2018;6(12):e11447.
208. Ravn Jakobsen P, Hermann AP, Søndergaard J, Wiil UK, Clemensen J. Development of an mHealth application for women newly diagnosed with osteoporosis without preceding fractures: a participatory design approach. *International journal of environmental research and public health*. 2018;15(2):330.
209. Schellong J, Lorenz P, Weidner K. Proposing a standardized, step-by-step model for creating post-traumatic stress disorder (PTSD) related mobile mental health apps in a framework based on technical and medical norms. *European journal of psychotraumatology*. 2019;10(1):1611090.
210. Tay I, Garland S, Gorelik A, Wark JD. Development and testing of a mobile phone app for self-monitoring of calcium intake in young women. *JMIR mHealth and uHealth*. 2017;5(3):e27.
211. van Agteren JE, Lawn S, Bonevski B, Smith BJ. Kick. it: the development of an evidence-based smoking cessation smartphone app. *Translational behavioral medicine*. 2018;8(2):243-67.
212. Vilardaga R, Rizo J, Zeng E, Kientz JA, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. *JMIR Serious Games*. 2018;6(1):e8881.
213. Wilhide III CC, Peeples MM, Kouyaté RCA. Evidence-based mHealth chronic disease mobile app intervention design: development of a framework. *JMIR research protocols*. 2016;5(1):e4838.
214. Woods L, Cummings E, Duff J, Walker K. Design thinking for mHealth application co-design to support heart failure self-management. *Studies in health technology and informatics*. 2017;241:97-102.

215. Fjeldsoe BS, Miller YD, O'Brien JL, Marshall AL. Iterative development of MobileMums: a physical activity intervention for women with young children. *International Journal of Behavioral Nutrition and Physical Activity*. 2012;9(1):1-11.
216. Ross XS, Gunn KM, Patterson P, Olver I. Development of a smartphone program to support adherence to oral chemotherapy in people with cancer. *Patient preference and adherence*. 2019;13:2207.
217. Denecke K, Warren J. How to Evaluate Health Applications with Conversational User Interface? *Studies in health technology and informatics*. 2020 Jun 16;270:976-80. PMID: 32570527. doi: 10.3233/shti200307.
218. Hussain S, Sianaki OA, Ababneh N, editors. A survey on conversational agents/chatbots classification and design techniques. *Workshops of the International Conference on Advanced Information Networking and Applications*; 2019: Springer.
219. ter Stal S, Kramer LL, Tabak M, op den Akker H, Hermens H. Design features of embodied conversational agents in eHealth: a literature review. *International Journal of Human-Computer Studies*. 2020;138:102409.
220. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet. *Journal of medical Internet research*. 2020;22(9):e22845.
221. Zierau N, Elshan E, Visini C, Janson A. A Review of the Empirical Literature on Conversational Agents and Future Research Directions. 2020.
222. Diederich S, Brendel AB, Kolbe LM. Towards a taxonomy of platforms for conversational agent design. 2019.
223. Chaix B, Bibault JE, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When Chatbots Meet Patients: One-Year Prospective Study of Conversations Between Patients With

Breast Cancer and a Chatbot. JMIR cancer. 2019 May 2;5(1):e12856. PMID: 31045505. doi: 10.2196/12856.

224. Echeazarra L, Pereira J, Saracho R. TensioBot: a Chatbot Assistant for Self-Managed in-House Blood Pressure Checking. Journal of medical systems. 2021 Mar 15;45(4):54. PMID: 33723721. doi: 10.1007/s10916-021-01730-x.

225. Gabrielli S, Rizzi S, Carbone S, Donisi V. A Chatbot-Based Coaching Intervention for Adolescents to Promote Life Skills: Pilot Study. JMIR human factors. 2020 Feb 14;7(1):e16762. PMID: 32130128. doi: 10.2196/16762.

226. Greer S, Ramo D, Chang YJ, Fu M, Moskowitz J, Haritatos J. Use of the Chatbot "Vivibot" to Deliver Positive Psychology Skills and Promote Well-Being Among Young People After Cancer Treatment: Randomized Controlled Feasibility Trial. JMIR Mhealth Uhealth. 2019 Oct 31;7(10):e15018. PMID: 31674920. doi: 10.2196/15018.

227. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A Smartphone-Based Health Care Chatbot to Promote Self-Management of Chronic Pain (SELMA): Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth. 2020 Apr 3;8(4):e15806. PMID: 32242820. doi: 10.2196/15806.

228. Kowatsch T, Schachner T, Harperink S, Barata F, Dittler U, Xiao G, et al. Conversational Agents as Mediating Social Actors in Chronic Disease Management Involving Health Care Professionals, Patients, and Family Members: Multisite Single-Arm Feasibility Study. J Med Internet Res. 2021 Feb 17;23(2):e25060. PMID: 33484114. doi: 10.2196/25060.

229. Kowatsch T, Volland D, Shih I, Rüegger D, Künzler F, Barata F, et al., editors. Design and evaluation of a mobile chat app for the open source behavioral health intervention platform MobileCoach. International Conference on Design Science Research in Information System and Technology; 2017: Springer.

230. Carfora V, Bertolotti M, Catellani P. Informational and emotional daily messages to reduce red and processed meat

consumption. *Appetite*. 2019 Oct 1;141:104331. PMID: 31276710. doi: 10.1016/j.appet.2019.104331.

231. Stasinaki A, Büchter D, Shih C-H, Heldt K, Güsewell S, Brogle B, et al. Effects of a novel mobile health intervention compared to a multi-component behaviour changing program on body mass index, physical capacities and stress parameters in adolescents with obesity: a randomized controlled trial. *BMC pediatrics*. 2021;21(1):1-16.

232. Denecke K, Hochreutener SL, Pöpel A, May R. Self-anamnesis with a conversational user interface: concept and usability study. *Methods of information in medicine*. 2018;57(05/06):243-52.

233. Lee HY, Koopmeiners JS, Rhee TG, Raveis VH, Ahluwalia JS. Mobile phone text messaging intervention for cervical cancer screening: changes in knowledge and behavior pre-post intervention. *Journal of Medical Internet Research*. 2014 Aug 27;16(8):e196. PMID: 25164545. doi: <https://dx.doi.org/10.2196/jmir.3576>.

234. Bock BC, Rosen RK, Barnett NP, Thind H, Walaska K, Foster R, et al. Translating Behavioral Interventions Onto mHealth Platforms: Developing Text Message Interventions for Smoking and Alcohol. *JMIR MHealth and UHealth*. 2015 Feb 24;3(1):e22. PMID: 25714907. doi: <https://dx.doi.org/10.2196/mhealth.3779>.

235. Militello LK, Kelly SA, Melnyk BM. Systematic review of text - messaging interventions to promote healthy behaviors in pediatric and adolescent populations: implications for clinical practice and research. *Worldviews on Evidence - Based Nursing*. 2012;9(2):66-77.

236. Reynoso R. A Complete Guide on How to Build a Chatbot (Easy to Hard). G22019 [cited 2021]; Available from: <https://www.g2.com/articles/how-to-build-a-chatbot>.

237. Ching C. How To Make An App in 2021. Code with Chris2021 [cited 2021]; Available from:

<https://codewithchris.com/how-to-make-iphone-apps-with-no-programming-experience/>.

238. Solomon M. If Chatbots Win, Customers Lose, Says Zappos Customer Service Expert. Forbes2017.

239. Morrissey K, Kirakowski J, editors. 'Realness' in Chatbots: Establishing Quantifiable Criteria. International Conference on Human-Computer Interaction; 2013: Springer.

240. Apps Ng. 5 Steps to Create Chatbots that Stick. New Gen Apps2018.

241. Morton K, Sutton S, Hardeman W, Troughton J, Yates T, Griffin S, et al. A Text-Messaging and Pedometer Program to Promote Physical Activity in People at High Risk of Type 2 Diabetes: The Development of the PROPELS Follow-On Support Program. JMIR MHealth and UHealth. 2015 Dec 15;3(4):e105.

PMID: 26678750. doi: <https://dx.doi.org/10.2196/mhealth.5026>.

242. McWhorter J. Txtng is killing language. JK!!! TED talks2013.

243. Fadhil A, Schiavo G, Wang Y, Yilma BA, editors. The Effect of Emojis when interacting with Conversational Interface Assisted Health Coaching System. Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare; 2018: ACM.

244. Menacho LA, Blas MM, Alva IE, Roberto Orellana E. Short Text Messages to Motivate HIV Testing Among Men Who have Sex with Men: A Qualitative Study in Lima, Peru. Open AIDS J. 2013;7:1-6. PMID: 23802032. doi:

<https://dx.doi.org/10.2174/1874613601307010001>.

245. Coniam D. The linguistic accuracy of chatbots: usability from an ESL perspective. Text & Talk. 2014;34(5):545-67.

246. Mac OA, Thayre A, Tan S, Dodd RH. Web-Based Health Information Following the Renewal of the Cervical Screening Program in Australia: Evaluation of Readability, Understandability, and Credibility. J Med Internet Res. 2020;22(6):e16701. PMID: 32442134. doi: 10.2196/16701.

247. Kincaid JP, Fishburne Jr RP, Rogers RL, Chissom BS. Derivation of new readability formulas (automated readability

index, fog count and flesch reading ease formula) for navy enlisted personnel. Naval Technical Training Command Millington TN Research Branch, 1975.

248. Bobrow K, Farmer A, Cishe N, Nwagi N, Namane M, Brennan TP, et al. Using the Medical Research Council framework for development and evaluation of complex interventions in a low resource setting to develop a theory-based treatment support intervention delivered via SMS text message to improve blood pressure control. BMC Health Serv Res. 2018 01 23;18(1):33. PMID: 29361934. doi:

<https://dx.doi.org/10.1186/s12913-017-2808-9>.

249. Chatterjee S, Price A. Healthy living with persuasive technologies: framework, issues, and challenges. Journal of the American Medical Informatics Association. 2009;16(2):171-8.

250. Aldoory L, Roberts EB, Bushar J, Assini-Meytin LC. Exploring the Use of Theory in a National Text Message Campaign: Addressing Problem Recognition and Constraint Recognition for Publics of Pregnant Women. Health Commun. 2018 Jan;33(1):41-8. PMID: 27858460. doi:

<https://dx.doi.org/10.1080/10410236.2016.1242034>.

251. Klok T, Kaptein AA, Brand PL. Non - adherence in children with asthma reviewed: The need for improvement of asthma care and medical education. Pediatric Allergy and Immunology. 2015;26(3):197-205.

252. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. MIS quarterly. 2003:425-78.

253. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Transactions on Computer-Human Interaction (TOCHI). 2005;12(2):293-327.

254. Fadhil A. Can a Chatbot Determine My Diet?: Addressing Challenges of Chatbot Application for Meal Recommendation. arXiv preprint arXiv:180209100. 2018.

255. Radziwill NM, Benton MC. Evaluating Quality of Chatbots and Intelligent Conversational Agents. CoRR. 2017;abs/1704.04579
256. Meadows R, Hine C, Suddaby E. Conversational agents and the making of mental health recovery. Digital health. 2020;6:2055207620966170.
257. Lum E, Jimenez G, Huang Z, Thai L, Car J. Appropriateness of action prompts for hypoglycaemia and hyperglycaemia in type 2 diabetes self - management apps. Diabetes/metabolism research and reviews. 2020;36(2):e3235.
258. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. Journal of Medical Internet Research. 2010 Feb 17;12(1):e4. PMID: 20164043. doi: <https://dx.doi.org/10.2196/jmir.1376>.
259. Fadhil A. A Conversational Interface to Improve Medication Adherence: Towards AI Support in Patient's Treatment. arXiv preprint arXiv:180309844. 2018.
260. Baron J. Here are the eight steps you'll go through building a bot. Recastai2017.
261. XMind. XMind - Ideas grow on trees. XMind2019 [cited 2019]; Available from: <https://www.xmind.net/>.
262. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annual review of public health. 2015;36:393-415.
263. Coomes CM, Lewis MA, Uhrig JD, Furberg RD, Harris JL, Bann CM. Beyond reminders: a conceptual framework for using short message service to promote prevention and improve healthcare quality and clinical outcomes for people living with HIV. AIDS Care. 2012;24(3):348-57. PMID: 21933036. doi: <https://dx.doi.org/10.1080/09540121.2011.608421>.
264. Mbuagbaw L, Mursleen S, Lytvyn L, Smieja M, Dolovich L, Thabane L. Mobile phone text messaging interventions for HIV and other chronic diseases: an overview of systematic reviews

and framework for evidence transfer. BMC Health Serv Res. 2015 Jan 22;15:33. PMID: 25609559. doi:

<https://dx.doi.org/10.1186/s12913-014-0654-6>.

265. Finitis DJ, Pellowski JA, Johnson BT. Text message intervention designs to promote adherence to antiretroviral therapy (ART): a meta-analysis of randomized controlled trials. PloS one. 2014;9(2):e88166.

266. Kramer J-N, Künzler F, Mishra V, Presset B, Kotz D, Smith S, et al. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. JMIR research protocols. 2019;8(1):e11540.

267. Wright C, Dietze PM, Agius PA, Kuntsche E, Livingston M, Black OC, et al. Mobile phone-based ecological momentary intervention to reduce young adults' alcohol use in the event: a three-armed randomized controlled trial. JMIR mHealth and uHealth. 2018;6(7):e149.

268. Mason AE, Saslow L, Moran PJ, Kim S, Wali PK, Abousleiman H, et al. Examining the effects of mindful eating training on adherence to a Carbohydrate-Restricted diet in patients with type 2 diabetes (the DELISH study): protocol for a randomized controlled trial. JMIR research protocols. 2019;8(2):e11002.

269. Mason AE, Jhaveri K, Cohn M, Brewer JA. Testing a mobile mindful eating intervention targeting craving-related eating: feasibility and proof of concept. Journal of behavioral medicine. 2018;41(2):160-73.

270. Tam. Bot Analytics—The Case For Third Party Tools. Chatbots magazine 2017.

271. Dombrowski C. The best Chatbot Third Parties Analytics for your Business. Medium 2018.

272. Phillips C. Chatbot Analytics 101: The Essential Metrics You Need to Track. Chatbots Magazine 2018.

273. Yan AF, Stevens P, Wang Y, Weinhardt L, Holt CL, O'Connor C, et al. mHealth text messaging for physical activity

promotion in college students: a formative participatory approach. Am J Health Behav. 2015 May;39(3):395-408. PMID: 25741684. doi: <https://dx.doi.org/10.5993/AJHB.39.3.12>.

274. Georgsson M, Staggars N. An evaluation of patients' experienced usability of a diabetes mHealth system using a multi-method approach. Journal of biomedical informatics. 2016;59:115-29.

275. O'Connor S. Actionable Tips to Protect Your Chatbot Data and User Privacy. Aithority2020; Available from: <https://aithority.com/guest-authors/actionable-tips-to-protect-your-chatbot-data-and-user-privacy/>.

276. Sundareswaran V, Sarkar A. Chatbots RESET A Framework for Governing Responsible Use of Conversational AI in Healthcare2020. Available from: http://www3.weforum.org/docs/WEF_Governance_of_Chatbots_in_Healthcare_2020.pdf.

277. Morrissey D. Chatbots, Healthcare, and Legal Advice: Best Practices. 2020; Available from: <https://botpublication.com/chatbots-healthcare-and-legal-advice-best-practices-b191d26bf510>.

278. Rumbold JMM, Pierscionek B. The effect of the general data protection regulation on medical research. Journal of medical Internet research. 2017;19(2):e47.

279. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR mHealth and uHealth. 2019;7(4):e11223.

280. Wong B. Data Privacy Law in Singapore: The Personal Data Protection Act 2012. 2017.

281. Daruwalla ZJ, Loh JL, Dong C. Spaced education and the importance of raising awareness of the personal data protection act: a medical student population-based study. JMIR medical education. 2016;2(2):e5586.

282. Dhinakaran DA, Sathish T, Kowatsch T, Griva K, Best JD, Car LT. Public Perceptions of Diabetes, Healthy Living, and

Conversational Agents in Singapore: Needs Assessment. JMIR Formative Research. 2021;5(11):e30435.

283. Bucher B. WhatsApp, WeChat and Facebook Messenger Apps – Global usage of Messaging Apps, Penetration and Statistics. Messenger people2020 [cited 2021]; Available from: <https://www.messengerpeople.com/global-messenger-usage-statistics/#Singapore>.

284. Kunze L. We're thinking about the Turing Test all wrong. Quartz2018 [cited 2018]; Available from: <https://qz.com/1487101/the-turing-test-shows-how-chatbots-ultimate-goal-isnt-intelligence-its-language/amp/>.

285. Zhu E, Lilienthal A, Shluzas LA, Masiello I, Zary N. Design of Mobile Augmented Reality in Health Care Education: A Theory-Driven Framework. JMIR Medical Education. 2015;1(2):e10. PMID: 27731839. doi: 10.2196/mededu.4443.

286. Mothership. S'poreans life expectancy increased, but more living with diabetes & hypertension2019. Available from: <https://mothership.sg/2019/03/singaporean-life-expectancy-diabetes-hypertension/>.

287. Group DPPR. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. New England journal of medicine. 2002;346(6):393-403.

288. Erickson G, Rothberg H. Knowledge and knowledge-related assets: Design for optimal application and impact. Successes and Failures of Knowledge Management: Elsevier; 2016. p. 113-27.

289. Schachner T, Keller R, Von Wangenheim F. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. Journal of medical Internet research. 2020;22(9):e20701.

290. Ang BW, Tan MY, Goh CM, Rahardja S, Lim BY, Chiew W, et al. Impact of knowledge and attitudes on lifestyle practices in preventing type 2 diabetes mellitus. Ann Acad Med Singap. 2019;48(8):247-63.

291. Kulandaivelu Y, Laloo C, Ward R, Zempsky WT, Kirby-Allen M, Breakey VR, et al. Exploring the needs of adolescents with sickle cell disease to inform a digital self-management and transitional care program: qualitative study. *JMIR pediatrics and parenting*. 2018;1(2):e11058.
292. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International journal for quality in health care*. 2007;19(6):349-57.
293. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. *Journal of medical Internet research*. 2015;17(10):e240.
294. Fukuoka Y, Gay CL, Joiner KL, Vittinghoff E. A novel diabetes prevention intervention using a mobile app: a randomized controlled trial with overweight adults at risk. *American journal of preventive medicine*. 2015;49(2):223-37.
295. Wong CK, Fung CS, Siu S, Lo YY, Wong K, Fong DY, et al. A short message service (SMS) intervention to prevent diabetes in Chinese professional drivers with pre-diabetes: a pilot single-blinded randomized controlled trial. *Diabetes Research and Clinical Practice*. 2013;102(3):158-66.
296. Burnard P. A method of analysing interview transcripts in qualitative research. *Nurse education today*. 1991;11(6):461-6.
297. Fridlund B, Hildingh C. *Qualitative research methods in the service of health: Studentlitteratur*; 2000. ISBN: 9144012489.
298. Chartered S. Are You Earning Enough? Singapore's Average Household Income Revealed. 2019; Available from: https://www.sc.com/sg/stories/financial-tips/are-you-earning-enough-singapores-average-household-income-revealed/#_ftn2.
299. Given L. Thematic coding and analysis. *The SAGE Encyclopedia of Qualitative Research Methods* Sage <https://dx.doi.org/10.4135/9781412963909>. 2008 (451).

300. Tesch R. Qualitative research—Analysis types and software protocols. Hampshire, UK: The Falmer Press. 1990.
301. Lincoln YS, Guba EG, Pilotta J. Naturalistic Inquiry California. Sage Publications; 1985.
302. Sim J, Sharp K. A critical appraisal of the role of triangulation in nursing research. *International journal of nursing studies*. 1998;35(1-2):23-31.
303. Lim RBT, Wee WK, Ananthanarayanan JA, Soh YH, Goh LML, Tham DKT, et al. Correlates, facilitators and barriers of healthy eating among primary care patients with prediabetes in Singapore—A mixed methods approach. *Nutrients*. 2019;11(5):1014.
304. Lim RBT, Wee WK, For WC, Ananthanarayanan JA, Soh YH, Goh LML, et al. Health education and communication needs among primary care patients with prediabetes in Singapore: a mixed methods approach. *Primary care diabetes*. 2020;14(3):254-64.
305. Ta V, Griffith C, Boatfield C, Wang X, Civitello M, Bader H, et al. User experiences of social support from companion chatbots in everyday contexts: Thematic analysis. *Journal of medical Internet research*. 2020;22(3).
306. Yates T, Davies MJ, Khunti K. Obesity and chronic disease in younger people: an unfolding crisis. *British Journal of General Practice*; 2012.
307. Phan TP, Alkema L, Tai ES, Tan KH, Yang Q, Lim W-Y, et al. Forecasting the burden of type 2 diabetes in Singapore using a demographic epidemiological model of Singapore. *BMJ Open Diabetes Research and Care*. 2014;2(1):e000012.
308. Liu S, Ko Q, Heng KQ, Ngiam KY, Feng M. Healthcare Transformation in Singapore with Artificial Intelligence. *Frontiers in Digital Health*. 2020;2:35.
309. Chatfuel. Chatfuel. Available from: <https://chatfuel.com/>.
310. CDC Ndpp. Lifestyle change program details. Available from: <https://www.cdc.gov/diabetes/prevention/lcp-details.html>.

311. Cappa. Improving lives through digital therapeutics. Available from: <https://cappahealth.com/>.
312. Lindström J, Louheranta A, Mannelin M, Rastas M, Salminen V, Eriksson J, et al. The Finnish Diabetes Prevention Study (DPS): Lifestyle intervention and 3-year results on diet and physical activity. *Diabetes Care*. 2003;26(12):3230-6.
313. Freeman D, Sheaves B, Goodwin GM, Yu L-M, Nickless A, Harrison PJ, et al. The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis. *The Lancet Psychiatry*. 2017;4(10):749-58.
314. Martín DB, De La Torre I, Garcia-Zapirain B, Lopez-Coronado M, Rodrigues J. Managing and controlling stress using mHealth: systematic search in app stores. *JMIR mHealth and uHealth*. 2018;6(5):e8866.
315. hub H. Aim for 150 minutes of physical activity every week. Available from: <https://www.healthhub.sg/live-healthy/386/150minutes>.
316. Heinrich KM, Patel PM, O'Neal JL, Heinrich BS. High-intensity compared to moderate-intensity training for exercise initiation, enjoyment, adherence, and intentions: an intervention study. *BMC public health*. 2014;14(1):1-6.
317. Stevanovic D. Quality of Life Enjoyment and Satisfaction Questionnaire—short form for quality of life assessments in clinical practice: A psychometric study. *Journal of Psychiatric and Mental Health Nursing*. 2011;18(8):744-50.
318. Centre MDR. The next set of questions ask about what you think of your risk or chance for getting diabetes. . [cited 2021]; Available from: <http://www.med.umich.edu/borc/profs/documents/svi/RPS-DDwritten.pdf>.
319. Centre NCBR. Food frequency questionnaires. Available from: <https://dapa-toolkit.mrc.ac.uk/diet/subjective-methods/food-frequency-questionnaire>.

320. Questionnaire IPA. Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ) - Short Form, . 2004.
321. Buysse DJ, Reynolds III CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry research*. 1989;28(2):193-213.
322. Cohen S, Kamarck T, Mermelstein R. Perceived stress scale. *Measuring stress: A guide for health and social scientists*. 1994;10(2):1-2.
323. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *bmj*. 2016;355.
324. Brahnam S, Weaver M, editors. *Re/framing virtual conversational partners: A feminist critique and tentative move towards a new design paradigm*. International Conference of Design, User Experience, and Usability; 2015: Springer.
325. Teo AR, Liebow SB, Chan B, Dobscha SK, Graham AL. Reaching those at risk for psychiatric disorders and suicidal ideation: Facebook advertisements to recruit military veterans. *JMIR mental health*. 2018;5(3):e10078.
326. Kayrouz R, Dear BF, Karin E, Titov N. Facebook as an effective recruitment strategy for mental health research of hard to reach populations. *Internet Interventions*. 2016;4:1-10.
327. Kwan RY, Lee D, Lee PH, Tse M, Cheung DS, Thiamwong L, et al. Effects of an mHealth brisk walking intervention on increasing physical activity in older people with cognitive frailty: Pilot randomized controlled trial. *JMIR mHealth and uHealth*. 2020;8(7):e16596.
328. Carter-Harris L, Ellis RB, Warrick A, Rawl S. Beyond traditional newspaper advertisement: leveraging Facebook-targeted advertisement to recruit long-term smokers for research. *Journal of medical Internet research*. 2016;18(6):e5502.
329. Sharman A, Zhussupov B, Sharman D, Kim I. Evaluating mobile apps and biosensing devices to monitor physical activity

and respiratory function in smokers with and without respiratory symptoms or chronic obstructive pulmonary disease: Protocol for a proof-of-concept, open-label, feasibility study. *JMIR research protocols*. 2020;9(3):e16461.

330. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. *Journal of medical Internet research*. 2017;19(8):e290.

331. West R, Michie S. A brief introduction to the COM-B Model of behaviour and the PRIME Theory of motivation. *Qeios*. 2020 2020/04/09. doi: 10.32388/WW04E6.2.

332. Michie S, van Stralen MM, West R. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*. 2011 2011/04/23;6(1):42. doi: 10.1186/1748-5908-6-42.

333. Piñeiro B, López-Durán A, Del Río EF, Martínez Ú, Brandon TH, Becoña E. Motivation to quit as a predictor of smoking cessation and abstinence maintenance among treated Spanish smokers. *Addictive behaviors*. 2016;53:40-5.

334. Dale H, Brassington L, King K. The impact of healthy lifestyle interventions on mental health and wellbeing: a systematic review. *Mental Health Review Journal*. 2014.

335. Health IoM. Latest nationwide study shows 1 in 7 people in Singapore has experienced a mental disorder in their lifetime. *News releases - Institute of Mental Health* 2018.

336. Ramachandran A, Snehalatha C, Ram J, Selvam S, Simon M, Nanditha A, et al. Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: a prospective, parallel-group, randomised controlled trial. *The lancet Diabetes & endocrinology*. 2013;1(3):191-8.

337. Baptista S, Wadley G, Bird D, Oldenburg B, Speight J, Group MDCR. User experiences with a type 2 diabetes coaching app: qualitative study. *JMIR diabetes*. 2020;5(3):e16692.

338. Stonbraker S, Cho H, Hermosi G, Pichon A, Schnall R. Usability testing of a mHealth app to support self-management of

HIV-associated non-AIDS related symptoms. *Studies in health technology and informatics*. 2018;250:106.

339. Gong E, Baptista S, Russell A, Scuffham P, Riddell M, Speight J, et al. My Diabetes Coach, a Mobile App–Based Interactive Conversational Agent to Support Type 2 Diabetes Self-Management: Randomized Effectiveness-Implementation Trial. *Journal of medical Internet research*. 2020;22(11):e20322.

340. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: a meta-analysis. *American journal of preventive medicine*. 2017;52(3):391-402.

341. Michie S, Van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation science*. 2011;6(1):1-12.

342. Michie S, Wood CE, Johnston M, Abraham C, Francis J, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). *Health technology assessment*. 2015;19(99).

343. Kramer LL, Ter Stal S, Mulder BC, de Vet E, van Velsen L. Developing embodied conversational agents for coaching people in a healthy lifestyle: Scoping review. *Journal of medical Internet research*. 2020;22(2):e14058.

344. Strava I. Strava: Run, Ride, Swim Apple app store2021.

345. Proactive Life I. SleepSpace - Doctor Approved. Google Play store2021.

346. centre DMsds. BUD.D. Google Play store2020.

347. Ltd ETP. My Dietitian, BMI, Diet Tips, Health, Food Search. Google Play store2021.

348. Motion R. RunMotion Coach - Running. Apple App store2018.

349. LTD NFL. PAWEN - Running App. Apple App store2021.

350. Anik Technologies I. Briea-Wellness Coach by Hygiea. Google Play store 2021.
351. Kalakheti R. NutriHealth - Nutrition Counselling Chat-bot. Google Play store 2020.
352. Garnett C, Oldham M, Angus C, Beard E, Burton R, Field M, et al. Evaluating the effectiveness of the smartphone app, Drink Less, compared with the NHS alcohol advice webpage, for the reduction of alcohol consumption among hazardous and harmful adult drinkers in the UK at 6 - month follow - up: protocol for a randomised controlled trial. *Addiction*. 2021;116(2):412-25.
353. Roberts DM, Schade MM, Mathew GM, Gartenberg D, Buxton OM. Detecting sleep using heart rate and motion data from multisensor consumer-grade wearables, relative to wrist actigraphy and polysomnography. *Sleep*. 2020;43(7):zsaa045.
354. Schade MM, Mathew GM, Roberts DM, Gartenberg D, Buxton OM. Enhancing Slow Oscillations and Increasing N3 Sleep Proportion with Supervised, Non-Phase-Locked Pink Noise and Other Non-Standard Auditory Stimulation During NREM Sleep. *Nature and Science of Sleep*. 2020;12:411.
355. Band R, Bradbury K, Morton K, May C, Michie S, Mair FS, et al. Intervention planning for a digital intervention for self-management of hypertension: a theory-, evidence- and person-based approach. *Implementation Science*. 2017;12(1):1-13.
356. Chiang N, Guo M, Amico KR, Atkins L, Lester RT. Interactive two-way mHealth interventions for improving medication adherence: an evaluation using the behaviour change wheel framework. *JMIR mHealth and uHealth*. 2018;6(4):e9187.
357. Battineni G, Chintalapudi N, Amenta F, editors. AI chatbot design during an epidemic like the novel coronavirus. *Healthcare*; 2020: Multidisciplinary Digital Publishing Institute.
358. Almalki M, Azeez F. Health Chatbots for Fighting COVID-19: a scoping review. *Acta Informatica Medica*. 2020;28(4):241.
359. Guerreiro MP, Angelini L, Henriques HR, El Kamali M, Baixinho C, Balsa J, et al. Conversational Agents for Health and

Well-being Across the Life Course: Protocol for an Evidence Map. JMIR research protocols. 2021;10(9):e26680.

360. Martinengo L, Lo NY, Goh WI, Car LT. Choice of Behavioral Change Techniques in Health Care Conversational Agents: Protocol for a Scoping Review. JMIR research protocols. 2021;10(7):e30166.

361. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: Systematic review. Journal of Medical Internet Research. 2020;22(10):e20346.

362. LIM J. Understanding the changes in Singapore's healthcare system in the last decade2020. Available from:

<https://www.todayonline.com/commentary/understanding-changes-singapores-healthcare-system-last-decade>.

363. Kliner M, Knight A, Mamvura C, Wright J, Walley J. Using no-cost mobile phone reminders to improve attendance for HIV test results: a pilot study in rural Swaziland. Infectious Diseases of poverty. 2013;2(1):1-7.

364. Sreekumar L. Poor health contributing to digital divide among older Singaporeans2020. Available from:

<https://www.duke-nus.edu.sg/allnews/poor-health-contributing-to-digital-divide-among-older-singaporeans>.

365. Ang S, Lim E, Malhotra R. Health-Related Difficulty in Internet Use Among Older Adults: Correlates and Mediation of Its Association With Quality of Life Through Social Support Networks. The Gerontologist. 2020;61(5):693-702. doi: 10.1093/geront/gnaa096.

366. Britt P. Bots Are Good, but They Can't Do It Alone2021 25/06/2022. Available from:

<https://www.destinationcrm.com/Articles/Editorial/Magazine-Features/Bots-Are-Good-but-They-Cant-Do-It-Alone-150267.aspx>.

367. van Olmen J, Erwin E, García-Ulloa AC, Meessen B, Miranda JJ, Bobrow K, et al. Implementation barriers for mHealth

for non-communicable diseases management in low and middle income countries: a scoping review and field-based views from implementers. Wellcome Open Res. 2020;5:7. PMID: 32399497. doi: 10.12688/wellcomeopenres.15581.2.

368. Aggarwal R, Visram S, Martin G, Sounderajah V, Gautama S, Jarrold K, et al. Defining the Enablers and Barriers to the Implementation of Large-scale, Health Care–Related Mobile Technology: Qualitative Case Study in a Tertiary Hospital Setting. JMIR Mhealth Uhealth. 2022;10(2):e31497. PMID: 35133287. doi: 10.2196/31497.

369. Ng E. Singapore ranked fifth in its readiness to achieve a fully integrated health system reveals first Future Health Index2016. Available from: https://www.philips.com.sg/about/news/archive/standard/news/press/2016/20160613_philips_singapore_ranked_fifth_in_its_readiness_to_achieve_a_fully_integrated_health_system_reveals_first_future_health_index.html.

370. Office SNaDG. Smart Initiatives to Improve Healthcare. SmartNation.gov.sg2022 [cited 2022]; Available from: <https://www.smartnation.gov.sg/initiatives/health>.

371. Safi S, Thiessen T, Schmailzl KJ. Acceptance and resistance of new digital technologies in medicine: qualitative study. JMIR research protocols. 2018;7(12):e11072.

372. Olmos-Vega FM, Stalmeijer RE, Varpio L, Kahlke R. A practical guide to reflexivity in qualitative research: AMEE Guide No. 149. Medical Teacher. 2022:1-11.

373. Asia CN. War on diabetes: Singapore to host global conference on tackling epidemic2018. Available from: <https://www.channelnewsasia.com/news/singapore/diabetes-singapore-to-host-global-ministerial-conference-10966520>.

374. Diabetes J. What is prediabetes. Joslin Diabetes Center, Inc.2019; Available from: <https://www.joslin.org/patient-care/diabetes-education/diabetes-learning-center/what-prediabetes>.

375. clinic M. Prediabetes. 2020; Available from: <https://www.mayoclinic.org/diseases-conditions/prediabetes/diagnosis-treatment/drc-20355284>.
376. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *New England Journal of Medicine*. 2001;344(18):1343-50.
377. George AM, Jacob AG, Fogelfeld L. Lean diabetes mellitus: an emerging entity in the era of obesity. *World journal of diabetes*. 2015;6(4):613.
378. Pan A, Sun Q, Manson JE, Willett WC, Hu FB. Walnut consumption is associated with lower risk of type 2 diabetes in women. *The Journal of nutrition*. 2013;143(4):512-8.
379. Kalergis M, Leung Yinko SSL, Nedelcu R. Dairy products and prevention of type 2 diabetes: implications for research and practice. *Frontiers in endocrinology*. 2013;4:90.
380. Nguyen V, Cooper L, Lowndes J, Melanson K, Angelopoulos TJ, Rippe JM, et al. Popcorn is more satiating than potato chips in normal-weight adults. *Nutrition Journal*. 2012;11(1):71.
381. Carter P, Gray LJ, Troughton J, Khunti K, Davies MJ. Fruit and vegetable intake and incidence of type 2 diabetes mellitus: systematic review and meta-analysis. *Bmj*. 2010;341.
382. Bazzano LA, Li TY, Joshipura KJ, Hu FB. Intake of fruit, vegetables, and fruit juices and risk of diabetes in women. *Diabetes care*. 2008;31(7):1311-7.
383. Salas-Salvadó J, Martínez-Gonzalez M, Bullo M, Ros E. The role of diet in the prevention of type 2 diabetes. *Nutrition, Metabolism and Cardiovascular Diseases*. 2011;21:B32-B48.
384. Katsanis J, Newman-Smith KC. Sleepless No More: Techniques and Interventions for Sleep Disorders.
385. Feng F, Zhang Y, Hou J, Cai J, Jiang Q, Li X, et al. Can music improve sleep quality in adults with primary insomnia? A systematic review and network meta-analysis. *International journal of nursing studies*. 2018;77:189-96.

386. Care IfQaEiH. Insomnia: Relaxation techniques and sleeping habits. 2017. InformedHealth.org [Internet]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK279320/>.
387. Watch HWsH. Too early to get up, too late to get back to sleep. Harvard Health Publishing2020; Available from: <https://www.health.harvard.edu/staying-healthy/too-early-to-get-up-too-late-to-get-back-to-sleep>.
388. Miller CB, Espie CA, Epstein DR, Friedman L, Morin CM, Pigeon WR, et al. The evidence base of sleep restriction therapy for treating insomnia disorder. Sleep medicine reviews. 2014;18(5):415-24.
389. Heinrich KM, Patel PM, O'Neal JL, Heinrich BS. High-intensity compared to moderate-intensity training for exercise initiation, enjoyment, adherence, and intentions: an intervention study. BMC public health. 2014;14(1):789.
390. Colberg SR, Sigal RJ, Yardley JE, Riddell MC, Dunstan DW, Dempsey PC, et al. Physical activity/exercise and diabetes: a position statement of the American Diabetes Association. Diabetes care. 2016;39(11):2065-79.
391. Association AD. Exercise for diabetes and get a leg up. 2020 [cited 2019]; Available from: <https://www.diabetes.org/fitness>.
392. Health A. Benefits of exercise and physical activity. Allina Health [cited 2019]; Available from: <https://www.allinahealth.org/health-conditions-and-treatments/health-library/patient-education/prediabetes/exercise-and-physical-activity>.
393. Migala J. Reversing Diabetes Through Weight Loss: How Much Should You Lose? Everyday Health2017 [cited 2019]; Available from: <https://www.everydayhealth.com/type-2-diabetes/treatment/can-type-2-diabetes-reversed-through-weight-loss/>.
394. Hub H. Different types of physical activity. Health hub SG2020; Available from: <https://www.healthhub.sg/live-healthy/387/TypesOfActivities>.

395. Christmann CA, Hoffmann A, Bleser G. Stress management apps with regard to emotion-focused coping and behavior change techniques: a content analysis. JMIR mHealth and uHealth. 2017;5(2):e22.
396. Martín DB, De La Torre I, Garcia-Zapirain B, Lopez-Coronado M, Rodrigues J. Managing and controlling stress using mHealth: systematic search in app stores. JMIR mHealth and uHealth. 2018;6(5):e111.
397. Michigan Uo. Stress Management: Breathing Exercises for Relaxation. Michigan Medicine2020 [cited 2020]; Available from: <https://www.uofmhealth.org/health-library/uz2255>.
398. Mindful. Getting started with Mindfulness. Mindful2020 [cited 2020]; Available from: <https://www.mindful.org/meditation/mindfulness-getting-started/>.
399. Varvogli L, Darviri C. Stress management techniques: evidence-based procedures that reduce stress and promote health. Health Science Journal. 2011.

Appendices

Appendix 4.1 – Search terms for scoping review

Date of Search: 8th April 2019

Search terms :

1. Artificially intelligent chatbot
2. Artificially intelligent chat agent
3. Automated virtual agent
4. Automated chat agent
5. Artificial intelligence chatbot
6. Artificial conversational entity
7. AI agent
8. Chatterbox
9. Chatbot
10. Chatterbot
11. Chat assistant
12. Conversational agent
13. Conversive agent
14. Conversational system
15. Conversational assistant
16. Conversational User Interface
17. Conversational Personal Assistant
18. Conversational interface
19. Conversational avatar
20. Conversational computer
21. Conversational humanoid
22. Conversational Character
23. Conversational bot
24. Conversational AI
25. Cyber individual
26. Dialog system
27. Interactive online character
28. Interactive talking program
29. Interactive virtual agent
30. Interactive agent
31. Interactive conversational assistant
32. Intelligent virtual agent
33. Intelligent virtual assistant
34. Intelligent conversational assistant
35. Intelligent conversational avatar
36. Intelligent agent
37. Online chat agent
38. Smartbot
39. Smart virtual assistant

40. Sociable agent
41. Talk bot
42. Talking agent
43. Talking avatar
44. Text-based healthcare chatbot
45. Text based dialogue system
46. Text-based synchronous chat
47. Virtual chat agent
48. Virtual personal assistant
49. Virtual online assistnat
50. Virtual host
51. Virtual hostess
52. Virtual human
53. Virtual human agent
54. Virtual human avatar
55. Virtual human persona
56. Virtual chat agent
57. Virtual chat expert
58. Virtual coach
59. Virtual consultant
60. Virtual conversational agent
61. Virtual advisor
62. Virtual agent
63. Virtual assistant

Appendix 4.2 - Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist [123]

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	Reported in Section # of the thesis
TITLE			
Title	1	Identify the report as a scoping review.	Chapter 4
ABSTRACT			
Structured summary	2	Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives.	Presented in published Scoping Review [42]
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach.	Chapter 4 – Section 4.1
Objectives	4	Provide an explicit statement of the questions and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives.	Chapter 4 – Section 4.1.1
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number.	NA
Eligibility criteria	6	Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale.	Chapter 4 – Section 4.2.2
Information sources*	7	Describe all information sources in the search (e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed.	Chapter 4 – Section 4.2.1
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	Appendix 4.1
Selection of sources of evidence†	9	State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review.	Chapter 4 – Section 4.2.3
Data charting process‡	10	Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators.	Chapter 4 – Section 4.2.3
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made.	Chapter 4 – Section 4.2.3

Critical appraisal of individual sources of evidence§	12	If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate).	NA
Synthesis of results	13	Describe the methods of handling and summarizing the data that were charted.	Chapter 4 – Section 4.2.3
RESULTS			
Selection of sources of evidence	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow	Chapter 4 – Section 4.3.1 and Figure 4.1
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	Chapter 4 – Section 4.3.2
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12).	NA
Results of individual sources of evidence	17	For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives.	Chapter 4 – Sections 4.3.2 – 4.3.4, Table 4.2 and Appendix 4.4.
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives.	Chapter 4 – Sections 4.3.3 & 4.3.4 and Section 4.4.1
DISCUSSION			
Summary of evidence	19	Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups.	Chapter 4 – Section 4.4.1
Limitations	20	Discuss the limitations of the scoping review process.	Chapter 4 – Section 4.4.3
Conclusions	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps.	Chapter 4 – Section 4.4.4
FUNDING	22	Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review.	Presented in published Scoping Review [42]

Appendix 4.3 – List of excluded studies and reasons for exclusion

Study ID	Title	Reason for exclusion
Abashev et al. (2017) [101]	Programming Tools for Messenger-Based Chatbot System Organization: Implication for Outpatient and Translational Medicines	No evaluation data*
Ahmad et al. (2018) [102]	Conversational bot for pharmacy: A natural language approach	Limited evaluation data**
Alexander 1973 [103]	Computer assisted optometry a tutorial with examples	No evaluation data
Amato et al. (2017) [104]	Chatbots meet eHealth: automatizing healthcare	No evaluation data
Atay et al. (2016) [105]	Can a smartphone-based chatbot engage older community group members? the impact of specialised content	Limited evaluation data**
Beun et al. (2016) [106]	Improving adherence in automated e-coaching	No evaluation data
Brixy (2017) [107]	SHIHbot: A Facebook chatbot for sexual health information on HIV/AIDS	No evaluation data
Callejas et al. (2014) [108]	A virtual coach for active ageing based on sentient computing and m-health	No evaluation data
Cameron et al. (2017) [109]	Towards a chatbot for digital counselling	No evaluation data
Cameron et al. (2018) [110]	Best Practices for Designing Chatbots in Mental Healthcare—A Case Study on iHelpr	No evaluation data
Chung et al. (2018) [111]	Chatbot-based healthcare service with a knowledge base for cloud computing	No evaluation data
Cooper et al. (2018) [112]	Designing a Chat-Bot for Non-Verbal Children on the Autism Spectrum	No evaluation data
D'Alfonso et al. (2017) [97]	Artificial Intelligence-Assisted Online Social Therapy for Youth Mental Health	No evaluation data
Denecke et al. (2018) [113]	Talking to Ana: A Mobile Self-Anamnesis Application with Conversational User Interface	No evaluation data
Dharwadkar et al. (2018) [114]	A Medical chatbot	No evaluation data
Divya et al. (2018) [115]	A self-diagnosis medical chatbot using artificial intelligence	No evaluation data
Do et al. (2016) [116]	Empathic virtual assistant for healthcare information with positive emotional experience	No evaluation data
Dubosson et al. (2017) [117]	Going beyond the relapse peak on social network smoking cessation programmes: ChatBot opportunities	No evaluation data
Fadhil (2017) (a) [118]	Addressing Challenges in Promoting Healthy Lifestyles	No evaluation data
Fadhil (2017) (b) [119]	An adaptive learning with gamification & conversational UIs: The rise of CiboPoliBot	No evaluation data
Fadhil et al. (2017) (c) [120]	CoachAI_A conversational UI assisted e-coaching platform	No evaluation data
Fadhil (2018) (a) [121]	Beyond Patient Monitoring: Conversational Agents Role in Telemedicine & Healthcare Support For Home-Living Elderly Individuals	No evaluation data
Fadhil (2018) (b) [122]	Can a Chatbot Determine My Diet?: Addressing Challenges of Chatbot Application for Meal Recommendation	No evaluation data
Fadhil et al. (2018) (c) [123]	A Conversational Interface to Improve Medication Adherence: Towards AI Support in Patient's Treatment	No evaluation data
Fadhil et al. (2018) (d) [124]	Assistive Conversational Agent for Health Coaching: A Validation Study	No evaluation data
Ferguson (2009) [125]	CARDIAC: An intelligent conversational assistant for chronic heart failure patient health monitoring	No evaluation data

Ferguson (2010) [126]	Towards a Personal Health Management Assistant	No evaluation data
Fernandez (2018) [127]	Implementation and feasibility study of a tailored health education bot in Telegram for mothers of children with obesity and overweight	No evaluation data
Fischer (2016) [128]	From books to bots: Using medical literature to create a chat bot	No evaluation data
Gabrielli (2018) [129]	SLOWBot (chatbot) Lifestyle Assistant	No evaluation data
Gatius (2012) [130]	Conversational System to Assist the User when Accessing Web Sources in the Medical Domain	No evaluation data
Hassoon (2018) [131]	Increasing Physical Activity Amongst Overweight and Obese Cancer Survivors Using an Alexa-Based Intelligent Agent for Patient Coaching: Protocol for the Physical Activity by Technology Help (PATH) Trial	No evaluation data
Hsu (2017) [132]	Allergybot: A chatbot technology intervention for young adults with food allergies dining out	Limited evaluation data**
Ireland (2018) [133]	Designing a Chat-Bot for Non-Verbal Children on the Autism Spectrum	No evaluation data
Ireland et al. (2015) [134]	Chat-Bots for People with Parkinson's Disease: Science Fiction or Reality?	No evaluation data
Ireland et al. (2016) [135]	Hello Harlie: Enabling Speech Monitoring Through Chat-Bot Conversations	Limited evaluation data**
Karpagam (2014) [136]	An intelligent conversation agent for health care domain	Limited evaluation data**
Kramer (2019) [137]	Investigating Intervention Components and Exploring States of Receptivity for a Smartphone App to Promote Physical Activity: Protocol of a Micro randomized Trial	No evaluation data
Kyo-Joong et al. (2018) [138]	Empathy Bot: Conversational Service for Psychiatric Counseling with Chat Assistant	No evaluation data
Lee (2017) [139]	The chatbot feels you-a counseling service using emotional response generation	No evaluation data
Lokman and Zain (2010) [140]	One-Match and All-Match Categories for Keywords Matching in Chatbot	Limited evaluation data**
Lokman et al. (2009) (a) [141]	An architectural design of Virtual Dietitian (ViDi) for diabetic patients	Limited evaluation data**
Lokman et al. (2009) (b) [142]	Designing a Chatbot for diabetic patients	No evaluation data
Madhu et al. (2017) [143]	A novel approach for medical assistance using trained chatbot	No evaluation data
Marciel et al. (2010) [144]	Cell phone intervention to improve adherence: cystic fibrosis care team, patient, and parent perspectives	No evaluation data
Marsh (2018) [145]	The use of a chatbot in radiology education	No evaluation data
Mascitti et al. (2010) [98]	COACH BOT: Modular e-course with virtual coach tool support	No evaluation data
Morris (2018) [146]	Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions	No evaluation data
Oh et al. (2017) [147]	A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation	No evaluation data
Oyebode et al. (2018) [148]	Likita: A Medical Chatbot To Improve HealthCare Delivery In Africa	No evaluation data
Pereira and Díaz (2018) [149]	Chatbot dimensions that matter: Lessons from the trenches	No evaluation data
Rarhi et al. (2017) [150]	Automated medical chatbot	No evaluation data
Ricardo et al. (2018) [151]	Proposal for the development of a mobile virtual assistant for treatment of tuberculosis	No evaluation data
Rick et al. (2019) [152]	SleepBot: encouraging sleep hygiene using an intelligent chatbot	No evaluation data

Rojas-Barahona et al. (2009) [153]	HomeNL: Homecare assistance in natural language. An intelligent conversational agent for hypertensive patients' management	No evaluation data
Rosruen et al. (2019) [154]	Chatbot Utilization for Medical Consultant System	No evaluation data
V. Manoj kumar et al. (2016) [155]	Sanative Chatbot For Health Seekers	No evaluation data
Vaira et al. (2018) [156]	Mama bot: A system based on ML and NLP for supporting women and families during pregnancy	No evaluation data
Valtolina et al. (2018) [157]	Chatbots and conversational interfaces: Three domains of use	Limited evaluation data**
Blanson Henkemans et al. (2009) [158]	An online lifestyle diary with a persuasive computer assistant providing feedback on self-management.	No evaluation data. Not a conversational agent.***
Both et al. (2010) [159]	Towards Fully Automated Psychotherapy for Adults: BAS - Behavioral Activation Scheduling via web and mobile phone	No evaluation data. Not a conversational agent.
Allen (2006) [160]	Chester: towards a personal medication advisor	Limited evaluation data**
Bickmore et al. (2013) [161]	Automated interventions for multiple health behaviors using conversational agents	No evaluation data. Not a conversational agent.*** (ECA)
Radziwill et al. (2017) [162]	Evaluating Quality of Chatbots and Intelligent Conversational Agents	No evaluation data
Miner et al. (2016) [163]	Conversational agents and mental health: Theory-informed assessment of language and affect	No evaluation data
Rizzo et al. (2011) [164]	An intelligent virtual human system for providing healthcare information and support	No evaluation data. Not a conversational agent.***
Shawar et al. (2007) [165]	Different measurements metrics to evaluate a chatbot system	Limited evaluation data**
Følstad et al. (2017) [184]	Chatbots and the new world of HCI	Limited evaluation data**
Hoermmann et al. (2017) [36]	Application of Synchronous Text Based Dialogue Systems in Mental Health Interventions: Systematic Review	No evaluation data
Mishra et al. (2018) [166]	Dr.Vdoc: A Medical Chatbot that Acts as a Virtual Doctor	Limited evaluation data**
Kazi et al. (2012) [167]	MedChatBot: An UMLS based Chatbot for Medical Students	Limited evaluation data**
Fadhil et al. (2016) [168]	Coach me: a platform for promoting healthy lifestyle	Limited evaluation data**
Bickmore et al. (2010) [169]	Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials	No evaluation data. Not a conversational agent.*** (ECA)
Lindenberg et al. (2011) [170]	E-health for individualized prevention of eating disorders	No evaluation data. Not a conversational agent.
Dowling et al. (2016) [171]	Exploring hope and expectations in the youth mental health online counselling environment	No evaluation data. Not a conversational agent.
Dowling et al. (2014) [172]	Investigating individual online synchronous chat counselling processes and treatment outcomes for young people	No evaluation data. Not a conversational agent.
Dowling et al. (2015) [173]	A naturalistic study of the effects of synchronous online chat counselling on young people's psychological distress, life satisfaction and hope	No evaluation data. Not a conversational agent.
Azevedo et al. (2018) [174]	Using conversational agents to explain medication instructions to older adults	No evaluation data. Not a conversational agent.*** (ECA)

Brown et al. (1981) [175]	A conversational information computer system for health and safety operation: The occupational surveillance interactive system (OSIS)	No evaluation data. Not a conversational agent.***
Crutzen et al. (2014) [176]	What can we learn from a failed trial: insight into non-participation in a chat-based intervention trial for adolescents with psychosocial problems	No evaluation data. Not a conversational agent.***
Denecke et al. (2019) [177]	Intelligent Conversational Agents in Healthcare: Hype or Hope?	No evaluation data
Fadhil et al. (2019) [178]	Designing for Health Chatbots	Limited evaluation data**
Yang et al. (2016) [179]	Mindbot: A Social-Based Medical Virtual Assistant	No evaluation data
Stieger et al. (2018) [180]	PEACH, a smartphone- and conversational agent-based coaching intervention for intentional personality change: study protocol of a randomized, wait-list controlled trial	No evaluation data
Palanica et al. (2019) [181]	Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey	No evaluation data
Moura et al. (2018) [182]	Intelligent chatbot for analysis and diagnosis of the psychiatric disorders	No evaluation data
Kowatsch et al. (2018) [183]	Designing Just-in-time Adaptive Interventions and Healthcare Chatbots with the Open Source Platform MobileCoach	Limited evaluation data**

***No evaluation data:**

- Papers with no health-related outcomes tested or presented. We only report on health-related outcomes. Details on technical outcomes are beyond the scope of this paper
- Some literature in the format of posters or protocols for example, do not present any evaluation data

****Limited evaluation data:**

eg. Limited qualitative data presented, with no supporting quantitative findings.

*****No evaluation data. Not a conversational agent:**

In some papers, the agent described (eg. ECA) did not match our definition of a conversational agent in this paper (ie. A non-embodied agent using text, speech or images as the main modality of communication with users).

Appendix 4.4 – Characteristics of CAs included in scoping review

Study ID	Conversational agent name	Conversational agent delivery channel	Conversational Agent Personality	Human Involvement	Short/Long term goals (Description)
Ballati <i>et al.</i> (2018)	Siri, Google Assistant, Amazon Alexa	Smartphone embedded software*	NA	NA	Short (To test the CA ability to comprehend individual spoken sentences from the dysarthric user and do so consistently)
Bickmore <i>et al.</i> (2018)	Siri, Alexa, Google Assistant	Smartphone embedded software*	NA	NA	Short (to provide advice or information for the user's specific medical question)
Casas <i>et al.</i> (2018)	Rupert le nutritioniste	Facebook messenger	Coach-like, informal, culture-specific	NA	Long (To encourage long term healthy eating habits - food coaching)
Chaix <i>et al.</i> (2018)	Vik	Website, app, messenger	Culture-specific, informal	NA	Long (To support patients with breast cancer and their relatives via personalized text messages addressing their queries)
Cheng <i>et al.</i> (2018)	Healthy Coping with Diabetes	Google Home assistant application (speaker/website)	Human-like	NA	Long (To promote the self-management of type 2 diabetes by older people)
Clinical trial NCT03384550 (2017)	Ally	Smartphone based platform (either android or iOS)	NA	NA	Long (to increase physical activity in healthy adults through coaching and incentive provision)
Clinical trial NCT (2018)	Vik	Unspecified	NA	NA	Long (to provide long term support for breast cancer patients by responding to any questions they have about their therapeutic management)
Comendador <i>et al.</i> (2015)	Pharmabot	computer-based platform	Informal, conversational agent identity	No.	Short (to provide information on generic medicine to patients)
Crutzen <i>et al.</i> (2011)	Bzz	Windows Live Messenger	Informal, knowledgeable	No.	Short & Long. Short = answer adolescent questions regarding sex, drugs and alcohol. Long = aim to increase knowledge in general or influence determinants of behaviour
Danda <i>et al.</i> (2016)	Vaidya	Smartphone application	Healthcare professional- like	NA	Short (diagnosis is provided based on user's symptoms)
Denecke <i>et al.</i> (2018)	Ana	Smartphone application	Culture-specific, human-like	NA	Short (Collect a comprehensive patient history via a thorough questionnaire prior to a doctor's appointment)
Elmasri <i>et al.</i> (2016)	Unspecified	web-based	Healthcare professional-like	NA	Long (Provide an alcohol risk assessment based on user's drinking habits & provides educational

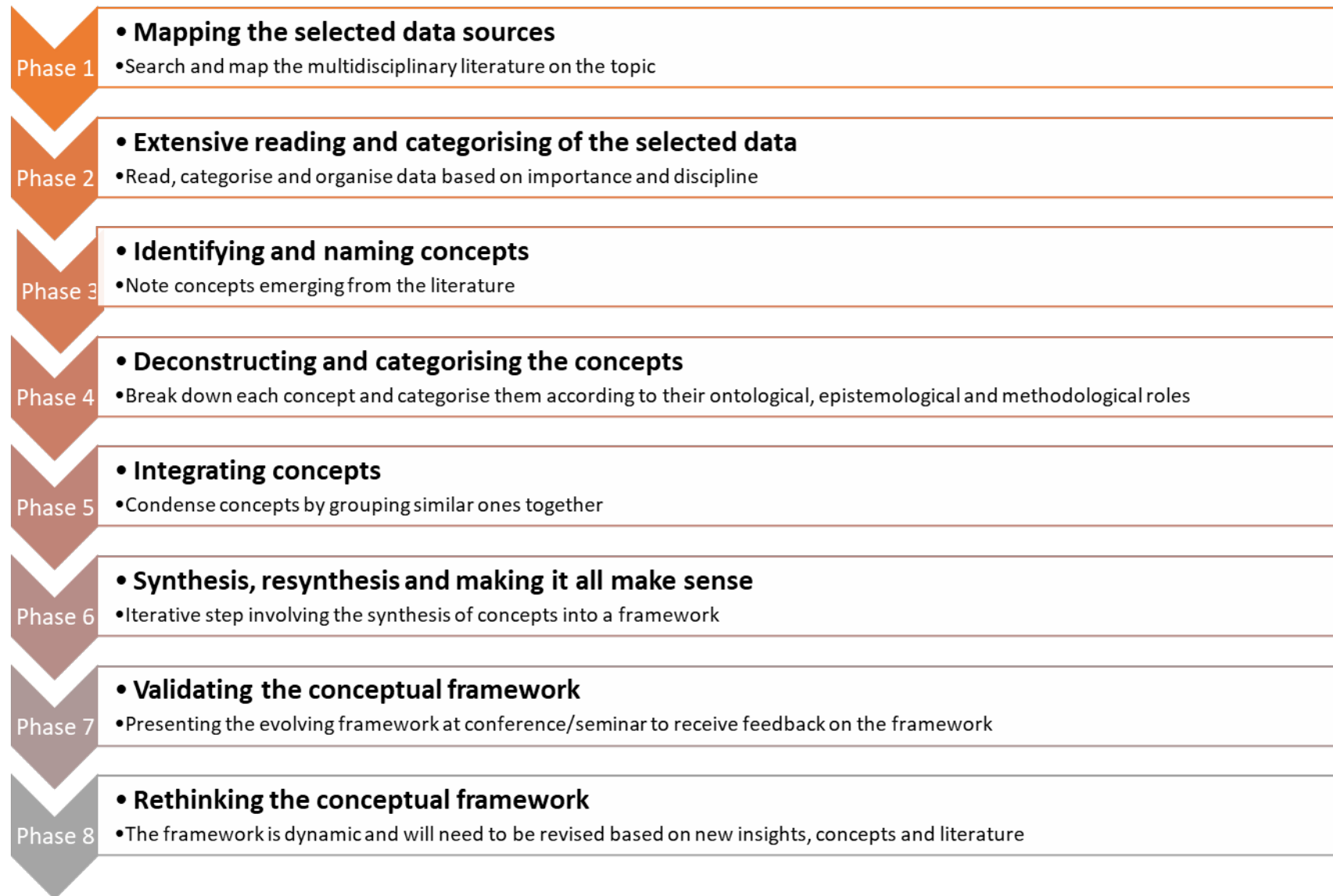
					content to imbue behavior change to reduce alcohol misuse)
Fitzpatrick <i>et al.</i> (2017)	Woebot	Facebook Messenger	Coach-like	No. (not during the chat. Encourage users to contain 911 for emergencies)	Long (Cognitive behavioral therapy)
Fulmer <i>et al.</i> (2018)	Tess	Smartphone application (Fb messenger, Slack, SMS)	Informal	No.	Short (To reduce self-identified symptoms of depression and anxiety in college students)
Gaffney <i>et al.</i> (2013)	MYLO	Web-based	Healthcare professional-like	NA	Long (To guide patients with problem management)
Galescu <i>et al.</i> (2009)	CARDIAC	Web-based	NA	No.	Short (To develop a plan-based CA to help chronic care patients look after themselves and provide comprehensive health care monitoring)
Ghandeharioun <i>et al.</i> (2018)	EMMA	Smartphone application	Human-like	No.	Short (the agent assesses the user's affect and provides a response/task for the user accordingly)
Ghosh <i>et al.</i> (2018)	Quro	Web-based	Healthcare professional-like	No.	Short (analyses user's symptoms, infers likely conditions and provides advice on what to do next)
Griol <i>et al.</i> (2015)	Unspecified	Unspecified	NA	No.	Short (An emotionally sensitive conversational system intended to support patients suffering from chronic pulmonary disease)
Griol and Callejas (2016)	Unspecified	Smartphone based (android) platform	Healthcare professional-like	NA	Long (multimodal conversational agent that combines user- preferences as with characteristics about the interaction environment, to improve and personalize the healthcare service provided)
Heldt <i>et al.</i> (2018)	Mobile coach (tele)	Smartphone application	Coach-like	NA	Long (To encourage an increase in physical activity for overweight individuals)
Huang <i>et al.</i> (2015)	Teenchat	Web-based	Healthcare professional-like, coach-like	NA	Long (Stress management)
Inkster <i>et al.</i> (2018)	WYSA	Smartphone application	Informal	NA	Long (to improve user's self-reported depression symptoms)
Joerin <i>et al.</i> (2019)	Tess	Smartphone application (Fb messenger, SMS, Amazon Alexa/Google Home)	NA	No.	Short (to provide customized psychological support for caregiving professionals, patients, and family caregiver)
Kamita <i>et al.</i> (2019)	SAT BOT	Smartphone application (LINE)	NA	No.	Short (stress reduction and motivation maintenance)

Kobori <i>et al.</i> (2018)	Unspecified	Web-based/Facebook messenger	factual, knowledgeable	NA	Short (STI diagnosis)
Kocielnik <i>et al.</i> (2018)	Reflection companion	SMS/MMS	knowledgeable	NA	Long (promote physical activity)
Kowatsch <i>et al.</i> (2017) [a]	MobileCoach	Smartphone application	Gender specific, informal, conversational agent identity	Yes. Manual chat channel for direct contact with their health professionals	Long (curb obesity)
Kowatsch <i>et al.</i> (2017) [b]	MobileCoach	Smartphone application	NA	Yes. Dedicated chat channel for patients and health professionals	Long (curb obesity)
L'Allemand <i>et al.</i> (2018)	Unspecified	Smartphone application	NA	No.	Long (to motivate overweight participates in a lifestyle intervention including relaxation and activity exercises)
Liu <i>et al.</i> (2018)	Unspecified	Smartphone application/Web-based	NA	No.	Short (To test the effect of three types of empathic expression—sympathy, cognitive empathy, and affective empathy—on individuals' perceptions of the service and the conversational agent)
Lobo <i>et al.</i> (2017)	CARMIE	Smartphone application	NA	No.	Short (to provide the patient with a response to their medication related query)
Ly <i>et al.</i> (2017)	Shim	Smartphone application	Coach-like	No	Long (to manage mental health issues/stress with positive psychology and CBT)
Middleton <i>et al.</i> (2016)	Babylon Check	Smartphone application	Healthcare professional-like	No.	Short (automated triaging)
Miner <i>et al.</i> (2016)	Siri, Google Now, S Voice, Cortana	Smartphone embedded software*	NA	No. (but some referral to human involvement if patient is susceptible to harm)	Short (to provides instant, automated answers to patients' queries)

Mujeeb <i>et al.</i> (2017)	Aquabot	Unspecified	NA	No.	Short (diagnosis of achluophobia (fear of darkness) and autism disorder)
Nadarzynski <i>et al.</i> (2018)	Unspecified	Smartphone application/Web-based	NA	NA	Short (To provide sexual health advice using an automated advice system)
Ni <i>et al.</i> (2017)	MANDY	Smartphone application	NA	No.	Short (to generate a report for the doctor (prior to consultation) regarding the patient's symptoms and likely causes)
Razzaki <i>et al.</i> (2018)	Babylon Triage and Diagnostic System	Smartphone application	Knowledgeable	No.	Short (online symptom checker for triage and diagnosis)
Rhee <i>et al.</i> (2014)	mASMAA	Smartphone application	Informal	Yes.	Long (ongoing symptom monitoring, treatment adherence, and nurturing adolescent-parent partnership)
Shaikh <i>et al.</i> (2018)	Wellthy diabetes (WD)	Smartphone application	Coach-like	Yes. (personalized coaching from an expert diabetes coach in additional to conversational agent)	Long (behavioral modification for better diabetes management and risk reduction)
Stasinaki <i>et al.</i> (2018)	Unspecified	Smartphone application	Coach-like	Yes. (One option for direct communication with a health professional)	Long (to promote weight loss and improved ability to perform physical activity for obesity therapy)
Stein <i>et al.</i> (2017)	Lark Weight Loss health Coach AI (HCA)	Smartphone application	Healthcare professional-like	No.	Long (Continued guidance to promote healthy behavior change through improved meal quality and weight loss for diabetes prevention)
Van Heerden <i>et al.</i> (2017)	Lwazi (male), Nolwazi (female)	Telegram bot	Human-like, gender specific,	Yes. (upon request or if there is risk of self-harm)	Short (Short counselling session to help the user overcome their fears and concerns to go ahead and get a HIV test)

Vita <i>et al.</i> (2018)	Doctor Apollo	Telegram bot	NA	n/a	Short & Long (Short = book appointment, medication reminders, long = ongoing support for their HIV management)
Wang <i>et al.</i> (2018)	Unspecified	WeChat platform	Knowledgeable	Yes. (a human administrator is also involved in conversations, and can instruct and guide participants if needed)	Long (Smoking cessation)
Wilson <i>et al.</i> (2017)	Siri, Google assistant	Smartphone embedded software	Factual	No.	Short (to answer sexual health related questions)

Appendix 5.1 – Jabareen’s methodology for conceptual framework development [102].



Appendix 5.2 – List of excluded studies (from mHealth framework literature search) and reasons for exclusion

Study ID	Title	Reason for exclusion
Blok et al. (2019)	Nurse-driven mHealth implementation using the technology inpatient program for smokers (TIPS): Mixed methods study	Program/framework doesn't describe the design/development process of the app, eg. part of the framework involves motivating mHealth usage by users, this is beyond the scope of our paper.
Cameron et al. (2017)	An ontology of and roadmap for mHealth research	Doesn't provide advice on mHealth development.
Chiang et al. (2018)	Interactive Two-Way mHealth Interventions for Improving Medication Adherence: An Evaluation Using The Behaviour Change Wheel Framework	Doesn't describe use of the framework to develop the intervention or how to do it.
DiStefano et al. (2016)	mHealth for Tuberculosis Treatment Adherence: A Framework to Guide Ethical Planning, Implementation, and Evaluation	Focus is on an ethical framework as opposed to a developmental framework
Lobelo et al. (2016)	The Wild Wild West: A Framework to Integrate mHealth Software Applications and Wearables to Support Physical Activity Assessment, Counseling and Interventions for Cardiovascular Disease Risk Reduction	focus on wearable activity monitors
Maar et al. (2017)	A Framework for the Study of Complex mHealth Interventions in Diverse Cultural Settings	focus is on human organisational level domains, not on design and development process
Sharp et al. (2017)	Mobile Medical Apps and mHealth Devices: A Framework to Build Medical Apps and mHealth Devices in an Ethical Manner to Promote Safer Use - A Literature Review	literature review only. No description of a step by step framework

Appendix 5.3 – List of excluded studies (from CA scoping review search) and reasons for exclusion

All the scoping review studies excluded (n=37) did not provide sufficient (if any) information on the design and development of the CA and no information was provided on how to improve the CA for the future. These 37 studies are listed below.

Study ID	Title
Ballati <i>et al.</i> (2018)	"Hey Siri, do you understand me?"
Bickmore <i>et al.</i> (2018)	Patient and consumer safety risks when using conversational assistant for medical information: an observational study of Siri, Alexa and Google assistant
Chaix <i>et al.</i> (2018)	When Chatbots Meet Patients: One-Year Prospective Study of Conversations Between Patients With Breast Cancer and a Chatbot
Cheng <i>et al.</i> (2018)	Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes
Clinical trial NCT03384550 (2017)	Can a Smartphone App That Includes a Chatbot-based Coaching and Incentives Increase Physical Activity in Healthy Adults?
Clinical trial NCT (2018)	Artificial Intelligence vs Physicians for Breast Cancer Patients' Information
Comendador <i>et al.</i> (2015)	Pharmabot: a pediatric generic medicine consultant chatbot'
Crutzen <i>et al.</i> (2011)	An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: An exploratory study
Danda <i>et al.</i> (2016)	Vaidya: A Spoken Dialog System for Health Domain
Elmasri <i>et al.</i> (2016)	A conversational agent for an online mental health intervention
Fitzpatrick <i>et al.</i> (2017)	Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial
Fulmer <i>et al.</i> (2018)	Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial
Gaffney <i>et al.</i> (2013)	Manage Your Life Online (MYLO): A Pilot Trial of a Conversational Computer-Based Intervention for Problem Solving in a Student Sample
Galescu <i>et al.</i> (2009)	Speech recognition in a dialog system for patient health monitoring
Ghandeharioun <i>et al.</i> (2018)	EMMA: An Emotionally Intelligent Personal Assistant for Improving Wellbeing
Ghosh <i>et al.</i> (2018)	Quro: Facilitating User Symptom Check Using a Personalised Chatbot-Oriented Dialogue System
Griol <i>et al.</i> (2015)	Towards Emotionally Sensitive Conversational Interfaces for E-therapy
Griol and Callejas (2016)	Mobile Conversational Agents for Context-Aware Care Applications
Huang <i>et al.</i> (2015)	Teenchat: A chatterbot system for sensing and releasing adolescents stress
Inkster <i>et al.</i> (2018)	An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study
Joerin <i>et al.</i> (2019)	Psychological Artificial Intelligence Service, Tess: Delivering On-demand Support to Patients and Their Caregivers: Technical Report
Kobori <i>et al.</i> (2018)	Novel application for sexual transmitted infection screening with an AI chatbot
Kocielnik <i>et al.</i> (2018)	Reflection companion: a conversational system for engaging users in reflection on physical activity
Liu <i>et al.</i> (2018)	Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot
Lobo <i>et al.</i> (2017)	CARMIE: A Conversational Medication Assistant for Heart Failure
Middleton <i>et al.</i> (2016)	Sorting out symptoms: design and evaluation of the 'babylon check' automated triage system
Miner <i>et al.</i> (2016)	Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health
Mujeeb <i>et al.</i> (2017)	Aquabot: A Diagnostic Chatbot for Achluophobia and Autism

Nadarzynski <i>et al.</i> (2018)	The acceptability of artificial intelligence (AI)-led chatbot, an automated advice system for sexual health
Ni <i>et al.</i> (2017)	Mandy: Towards a smart primary care chatbot application
Razzaki <i>et al.</i> (2018)	A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis
Rhee <i>et al.</i> (2014)	Mobile phone-based asthma self-management aid for adolescents (mASMAA): a feasibility study
Shaikh <i>et al.</i> (2018)	Real-world effectiveness of a digital therapeutic in improving glycaemic control in south Asians living with type 2 diabetes
Stein <i>et al.</i> (2017)	A Fully Automated Conversational Artificial Intelligence for Weight Loss: Longitudinal Observational Study Among Overweight and Obese Adults
Van Heerden <i>et al.</i> (2017)	The potential of conversational agents to provide a rapid HIV counselling and testing services
Vita <i>et al.</i> (2018)	The "doctor Apollo" chatbot: A digital health tool to improve engagement of people living with HIV
Wilson <i>et al.</i> (2017)	In bed with Siri and Google Assistant: a comparison of sexual health advice

Appendix 5.4 – How each paper contributed to a relevant phase of the conceptual framework.

Study ID	(1) Defining the goal	(2) Creating an identity	(3) Developing the content	(4) Conversation flow building	(5) Piloting and evaluating
Carfora 2019			Types of messages - employ different messages for different purposes eg. Reminder, educational - Emotional messages were more effective in inducing anticipated regret regarding meat consumption compared with informational messages.		
Casas 2018		- Personality The aim was to support the user in achieving their healthy eating goals. Hence, the CA showed empathy, was motivating and did not judge. He also used familiar language and emoticons. - Appropriate tone and language The CA was administered in French to make the users comfortable by conversing in their native language.	Consider relevant techniques to change behaviour - Motivation and coaching were used, along with short daily interactions which evolved into longer guided conversations to prevent the user from giving up on their diet goals too soon.		
Chaix 2019			- Sufficient and appropriate pre-defined answer options. Also, an engagement strategy.		

			Informative answer options. The user doesn't have to enter a sentence to view information of interest. CA initiating conversation, removes burden from user which enhances engagement.		
Denecke 2018 (a)	Text supplementations - Users thought there was too much text on the page and would prefer a video or animation - Text was too small for the older people. Select appropriate font size to suit target population.	Personality - CA motivated users to provide responses and made the conversation fun. Users felt comfortable sharing things with the CA they wouldn't share with their doctor. Tone and language - Language used was linguistically correct and understandable			
Echeazarra 2021	Delivery mode - The research team selected a platform based on target population's familiarity with Telegram.		Error management - The CA was equipped with a default method to deal with typing errors. The user was asked to double check if the value entered is outside the normal BP range. If incorrect, they could correct this. If input value was correct, the HCP would be informed.		
Gabrielli 2020	Text supplementations - Users particularly liked the video cartoons shared by the CA	Personality - Participants appreciated that the CA felt like an old friend but thought that responses were	Co-design - To collect their needs and preferences on how the CA should look and feel, the type of		

	- Too much text was sometimes sent which required the user to scroll to read it. Text could be shortened.	sometimes missing relevance or empathy.	content to include, how long each session should last and improvements on the prototype. The co-design helped inform the coaching dialogue for adolescence needing life skills training.		
Greer 2019	Text supplementations - Videos and other content produced by young adults treated for cancer were included along with text	- Personality Participants commented on the non-judgemental nature of the CA. Others appreciated the ability to “vent” indicating that the CA provided a safe space for them to express themselves.	Co-design - Focus groups with prospective user population (adolescents treated for cancer) to inform text and videos around psychosocial well-being based on their suggestions.		
Hauser-Ulrich 2020		Personality - Each interaction commences with a warm greeting from the CA where they enquire about the user’s mood in an empathetic manner. Language - The conversational style was aimed to imitate human interactions by incorporating emojis and some humour.		Personalisation - Personalised messages daily or every other day to initiate conversation. Engagement strategies - CA to initiate conversations and send notifications	
Heldt 2018		Personality - Male and female versions of the CA		Timing & frequency - Daily challenges implemented to help	

		available. They had culture specific names to make them familiar to the user.		users achieve their weight loss and relaxation goals.	
Kamita 2019	Delivery mode - The LINE messaging platform was selected as it is an avidly used chat tool in Japan.				Human involvement/Evaluation - They suggested to test the CA, in the future, in an environment where it can provide the user's mental health condition to counsellors or doctors for collaborative support
Kowatsch 2017 (a)	Multidisciplinary team - The team included computer scientists, physicians, a psychotherapist, diet and sport experts for a CA targeting childhood obesity	Personality - Friendly. Gender neutral (ie male and female versions available). Culture specific names were given to the Cas. They had a comic profile image and used emoticons to give empathic feedback. Engaged in small talk and expressed happiness to see the patient and chat with him or her.			
Kowatsch 2017 (b)					
Kowatsch 2021	Multidisciplinary team - The team comprised: computer scientists, asthma experts, and young patients with their parents for asthma	Personality - CA was given a name and hence, an identity. The CA was empathetic and asked users about their emotional state and gave appropriate feedback. Language	Evidence based content - The content was informed by the latest evidence, collated from a literature review	Suitable answer options - A mixture of pre-defined options and free text were used. where appropriate) Engagement strategies - Gamification, incentives, and notifications were effective strategies to keep users engaged and precluded attrition	Human involvement - Test in a hybrid environment where users can reach out for help beyond CA and HCPs can deliver feedback regarding inhalation technique to patients

	care support chatbot	- Emojis and some humour was used by the CA to build a social relationship with the user and nurture the working alliance.			
L'Allemand 2018					Human involvement - Study team suggested testing in a hybrid condition where virtual coach chatbot for overweight adolescents is supplemented with regular onsite visits with HCPs.
Ly 2017	Multidisciplinary team - Dialogues to promote mental well-being pre-written by professionals trained in psychology			Engagement strategies - The study team observed that initiation of the interactions by the CA was preferred as this removed the burden from user	
Stasinaki 2021			Consider appropriate behaviour change interventions (BCI) - The study team employed multi-component BCI following Swiss guidelines, with educational handouts on nutrition and physical activity to help overweight children.	Engagement strategies - Gamification was employed. Users would receive virtual rewards when they completed their challenges, performed the breathing exercises and shared pictures of their meals.	
Wang 2018	Delivery channel - WeChat was used in this study based	Language - The intervention was delivered in Chinese	Types of messages - 4 types of messages were sent by the CA:		

	on its wide popularity and favourable user interface	as applicable to the target population.	announcements (general rules and purpose of the group), sharing (health information related to smoking), reminders (for group members to share their positive experiences regarding smoking cessation) and responses (by the CA to certain keywords written by users eg "SOS").		

Appendix 6.1 – COREQ checklist for needs assessment

Item	Guide questions/description	Reported in Section # of thesis
Domain 1: Research team and reflexivity		
1. Interviewer/facilitator	Which author/s conducted the interview?	Chapter 6 – Section 6.2.2
2. Credentials	What were the researchers' credentials?	Chapter 6 – Section 6.2.2
3. Occupation	What was their occupation at the time of the study?	Chapter 6 – Section 6.2.2
4. Gender	Was the researcher male or female?	Chapter 6 – Section 6.2.2
5. Experience and training	What experience or training did the researcher have?	Chapter 6 – Section 6.2.2
6. Relationship with participants established	Was a relationship established prior to study commencement?	No
7. Participant knowledge of the interviewer	What did the participants know about the researcher?	Chapter 6 – Section 6.2.2
8. Interviewer characteristics	What characteristics were reported about the interviewer/facilitator?	Chapter 6 – Section 6.2.2
Domain 2: Study Design		
9. Methodological orientation and theory	What methodology orientation was started to underpin the study?	Chapter 6 – Section 6.2
10. Sampling	How were participants selected?	Chapter 6 – Section 6.3
11. Method of approach	How were participants approached?	Chapter 6 – Section 6.3
12. Sample size	How many participants were in the study?	Chapter 6 – Section 6.3
13. Non-participation	How many people refused to participate or dropped out? Reasons?	None
14. Setting of data collection	Where was the data collected?	Chapter 6 – Section 6.2.2
15. Presence of non-participants	Was anyone else present besides the participants and researchers?	No.
16. Description of sample	What are the important characteristics of the sample?	Chapter 6 – Section 6.3.1
17. Interview guide	Were questions, prompts, guides provided by the authors?	Yes. Appendix 6.2
18. Repeat interviews	Were repeat interviews carried out?	No.
19. Audio/visual recording	Did the research use audio or visual recording to collect the data?	Chapter 6 – Section 6.2.2
20. Field notes	Were field notes made during and/or after the interview?	Chapter 6 – Section 6.2.2
21. Duration	What was the duration of the interviews?	Chapter 6 – Section 6.2.2
22. Data saturation	Was data saturation discussed?	Chapter 6 – Section 6.2.2
23. Transcripts returned	Were transcripts returned to participants for comment and/or correction?	No.
Domain 3: analysis and findings		
24. Number of data coders	How many data coders coded the data?	Chapter 6 – Section 6.2.3

25. Description of the coding tree	Did authors provide a description of the coding tree?	Appendix 6.3
26. Derivation of themes	Were themes identified in advance or derived from the data?	Chapter 6 – Section 6.3.2
27. Software	What software, if applicable, was used to manage the data?	NA
28. Participant checking	Did participants provide feedback on the findings?	NA
29. Quotations presented	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified?	Chapter 6 – Section 6.3.2
30. Data and findings consistent	Was there consistency between the data presented and the findings?	Chapter 6 – Sections 6.3.2 and 6.4.1
31. Clarity of major themes	Were major themes clearly presented in the findings?	Chapter 6 – Section 6.3.2
32. Clarity of minor themes	Is there a description of diverse cases or discussion of minor themes?	Chapter 6 – Sections 6.3.2 and 6.4.1

Appendix 6.2 – Needs assessment interview guide

Venue: Telephone interview

Language: English

Aim: To elicit participants views on pre/diabetes-related information needs, preferences and the use of mobile phone messaging to encourage behavioural change

Format: a semi-structured interview & survey after participants are recruited and have signed an informed consent

1. Questionnaire (interviewer-administered)

Can I first ask you a few questions about yourself? Thank you.

Gender: ☐ Male ☐ Female

Age: ☐ 25 – 30 years old
☐ 31 – 35 years old
☐ 36 – 40 years old
☐ 41 – 45 years old
☐ 46 – 50 years old
☐ 51 – 55 years old
☐ 56 – 60 years old
☐ >60 years old

Ethnicity: ☐ Chinese
☐ Malay
☐ Indian
☐ Others (Please specify): _____

Marital Status: ☐ Single
☐ Married
☐ Others (Please specify): _____

Monthly Household Income: ☐ below \$4,000
☐ \$4,000 – \$6,000
☐ \$6,001 – \$8,000
☐ \$8,001 – \$10,000
☐ \$10,001 – \$12,000
☐ \$12,001 – \$14,000
☐ >\$14,000

Family Structure
No. of children & age(s): _____
No. of dependents: _____
Type of dependency: _____
Partner's Job profession: _____

What do you do for living?

Do you have prediabetes?

If yes, how did you learn you had prediabetes?

Does anyone else in your family have prediabetes or diabetes? If yes, who?

Do you have any questions on prediabetes or diabetes?

Do you use messaging apps such as Facebook Messenger? Which ones? What is your opinion on them?

Do you use any mobile applications on prediabetes or diabetes? If yes, which ones? What is your opinion on them?

Are you familiar with chatbots? Please explain what a chatbot is.

2. Questions on care gaps as well as barriers and facilitators to a healthy lifestyle

I will now proceed to ask you a few questions in relation to healthy lifestyle and your views on it.

(Pre)Diabetes knowledge/awareness

Diabetes is a serious condition that affects many people in Singapore

- a. Suppose you had one minute to explain to someone what diabetes is, what would you say?
- b. What do you think is the most important topic for you to learn more about to prevent diabetes?
- c. Pre-diabetes affects more than 20% of Singaporeans. How would you describe pre-diabetes?
- d. If you had pre-diabetes how do you think it would affect your life? If you had diabetes how do you think it would affect your life?

Eating habits

- a. How would you describe healthy eating habits?
- b. Sometimes we do not always eat the way we would like to. Let's talk about some difficulties you experience with eating healthy. What influences the foods that you decide to eat?
- c. What could help you have/maintain healthier eating habits?
- d. If someone asked you to suggest how to have a healthier diet, what would you suggest?

Physical Activity

- a. When you hear the word physical activity what comes to mind?
- b. What influences how physically active you are each day?
- c. If you do feel you need to get more physical activity, how would you add more physical activity into your daily life?

Weight management

- a. How do you feel about your body weight?
- b. What are some things that you feel would help you achieve or maintain a healthy body weight?

Stress/sleep

- a. What role does stress have in your life?
- b. What would help reduce stress levels in your life?
- c. How are your sleeping habits?
- d. What would help you improve your sleeping habits?

mHealth intervention

I would now like to talk to you about our intervention. We are putting together a diabetes prevention program for people at risk of diabetes. This intervention will be delivered via mobile phones using messages.

1. What are some topics you would want us to include in this interventions?
2. What would this messages need to look like? What type of information would you prefer – text, videos, images etc.? Would you prefer formal or informal tone in the communication? Would you prefer one-way messages (i.e. only receiving messages) or two-way communication (e.g. having to or preferring to respond)?

Closing

Allow me to do a brief recap of the main points. Please do let me know if you think that something has been left out.

Is there anything else you would like to add on the topics we have discussed today?

Thank you very much for participating in today's discussion and for your valuable input. Your comments will be very helpful to us in developing an intervention for prevention of diabetes in people at risk of diabetes.

Appendix 6.3 – Needs assessment codebook from qualitative analysis

Themes	Sub-themes	Codes	Quotes
Theme 1: Use of conversational agents			
	Defining conversational agents	AI programme/ Virtual call centre	P20 "A chatbot is an AI programme" P14 "Chatbot is someone, is a virtual call centre that they always like to ask you how can I help you?"
		Readily available	P12 "The fact there is readily available information that's feedback to me in a very short period of time."
		Unfamiliarity	P11 "Chat bots, not really in particular." P19 "No, never heard of that."
		Usage applications	car insurance, public transport, online shopping, banking and airline services
	Beneficial features & suggestions for future interventions	[Delivery format:] • Interactive • Visual aids • Text	P14 "it's very interactive. And everybody, by talking to them, we get some point, really. So, we enjoy it." P12 "I think the use of multimedia here will be best, if there can be images and videos supported by some further text is the user chooses to click on and read more, that will help."
		Pre-prepared options	P20 "was using the IBM Watson kind of AI technology. But it wasn't very, very smart, to say. So I guess if, in this case, it's more of an MCQ kind of thing, like it gives you fixed responses to give"
		Group component	P05 "If I did it with other people, like with friends and stuff yea, you could make it like a group effort. Then it'll be easier because then there are other people to monitor you, just for yourself."
		Novelty	P15 "But that one was not very good because they cannot... It becomes a pattern. Once they get used to it, then people will try to run away from the pattern."
		Appropriate tone to suit function	P19 "I'd prefer a bit of informal, but not too informal, probably a balance."
		Favourable Frequency and duration of interaction	P19 "I'd say two to three times a week. Probably 15, 20 minutes."
		Bidirectional communication	P19 "Preferably two way so that you can at least be assured that your questions are answered."
	Unfavourable traits	Incapable of answering questions	P09 "When I had questions, they were not answered appropriately. So, I had to resort to calling them."

		Irrelevant answers	P14 “not useful in answer to our questions. they can't answer specifically to our needs.”
Theme 2: Ubiquity of smartphone application			
	Usage of smartphone apps	Messaging apps	WhatsApp, Facebook Messenger, Telegram, WeChat and LINE
		Health apps	Step trackers, Healthy 365, my fitness pal, AIA vitality, Fitbit, Samsung Health, HPB (Health Promotion Board) app, Nike Run Club, Chronometer (monitor nutrition) and Stride (habit tracker)
	Beneficial features	[Positive user experience:] • ease of use • engaging • multifunctional	P17 “Generally, they are quite convenient “ P18 “Telegram, because of the you can use stickers. The fun part if that it’s engaging, when we message our friends.” P19 “You get to communicate with someone. You get to video call them as well, so I’d say it’s a very good platform.”
		Fulfil a necessity	P15 “I have to use them.”
		Provide incentives	P17 “And then for the Healthy 365 app by HPB, I use it to redeem the rewards and participate in their promotional activities and programmes.”
	Limitations of smartphone apps	Questionable accuracy of content	P07 “I think it’s like a good indicator at least you have some number that you can see. But the accuracy is still debatable.”
		Overwhelming frequency	P06 “It can be useful. But it also can be it can also be an harassment (laughs) turns out how you look at it” “As in because the messages comes so easily so sometimes there’s like a lot of messages to see a day yep.”
Theme 3: Understanding of diabetes			
	Defining type 2 diabetes	• elevated blood sugar, • insulin, • reduce sugar intake, • no cure, • medication	P20 “diabetes is a condition in which your sugar in your body is elevated.
		Knowledge of its implications on one’s life	P14 “Having good health to be drained away. And then, how does this affect families concerned” “... I was told that people with diabetes may not be able to exercise actively. Have to change the

			type of exercise that they do. Sometimes breathing will be affected, also. The diets, will have to need a repair. What they like to eat has to change to new diet requirement”
		Knowledge of its implications on one’s health	P20 “It affects your vessels in your eyes, your feet and your kidneys, yes, all parts of your body.”
	Defining prediabetes	• borderline diabetes risk, • reversible, • unfamiliarity with the term	P15 “what I understand that prediabetes is, is that they are closing towards the diabetic range when they do their tests. So they are the borderline testing.” P17 “It’s very much preventable and treatable before it goes into the diabetes stage.” P19” To be honest I'm not too sure what that term is.”
		Recognition of unequal awareness compared with diabetes	P07 “I think diabetes is more commonly spoken about I think.”
	Sources of information	Personal experience	P08 “For diabetes right, for my mum, she doesn’t need those injection like insulin. So far its just purely on medication so when I did the test you know like my dietary and also my wellness and so forth right”
		Government influence	P14 “The government is trying to educate the citizens about the pre-diabetic issue.”
		Lack of information available (in general, to older population)	P12 “not a lot of information readily available” P09 “Yea, I wish I knew more.”
		Additional suggestions (media, internet, hawker centres	P08 “and also there could be more awareness like you know when we go to restaurant or maybe hawkers more information could share like the calories. Because yea I know some hawker centres they do show but it didn’t exactly show the number of calories for all the food.” P17 “And also, maybe through TV or some of the activities that have been going on, for example, the 10k walk.”
	Knowledge gaps	How to make healthy lifestyle changes	P11 “In terms of exercise, regulating diet, maybe sleep and health in general.”
		Symptoms and prevention of diabetes	P18 “Maybe share with us on symptoms of prediabetes and also give us advice on how to prevent that.”
		Awareness of being at risk of diabetes	P09 “What are the risk factors that we have to look out for? Be aware of?”
		Application of knowledge gained	P20 “I wouldn’t say you need to learn about it. You just need to practice it and incorporate it into your life.”

Theme 4: Barriers and facilitators to a healthy living in Singapore			
	What influences lifestyle choices	Convenience	P20 "Convenience. For example, if I'm studying here then I've one hour to go eat lunch, then usually the places I choose is they have to be nearby."
		Cost	P17 "The biggest problem we have here is the price of healthy food. It's just infinitely more costly. For example, some salad can cost you six, \$7, whereas chicken rice can cost you \$3. You cannot possibly eat salad every day even if you are earning okay. But if you are one of those that are low income, definitely you will go for three-dollar chicken rice every day. And in the long run, it just causes a lot of problems."
		Personal preferences	P07 "I mean, I think naturally it will be. I think people naturally gravitate towards like your favourite food"
		Social activities	P06 "So sometimes it's a social setup. Like you go out with friends then and you go for a buffet, so that needs a lot of self-control"
		Advertising & Marketing	P17 "It just depends on what I feel like eating or maybe a bit of marketing as well. For example, if McDonald's has a new menu, then I might just go for it."
		Nutritional value of food	P20 "then the dietary value of the food also. So how healthy is it and whether it's nutritious.
		State of mind	P09 "Sometimes you just forget! Sometimes you just forget and you know, okay lah tomorrow" "Mood, cravings."
	Barriers to healthy living	Dining out	P18 "It's the availability of the ingredients. If you go to hawker centres, I don't think they really have brown rice. The food is oily, so sometimes you've no choice. Or for drinks, we can't reduce the sugar intake"
		Lack of appeal	P06 "Of course like sometimes healthy food don't really looks the most palatable"
		Health condition	P15 "For me, in the past, because of my brain tumour, I have compulsive eating problems"
		Fear of injury	P14 ", but there is also a fear that at a certain age. there is some wear and tear in the body"

		Lack of free time	P08 “whether I have time to dock off for a run or not. It really depends because most of the time, I’m actually working quite late.... I think for myself would definitely be work stress.... So myself sometimes I work 15 to 16 hours a day”
		Weather limitations	P08 “especially with the haze now even if I do wish I could go run, I can’t.”
		Specific events	P14 “where there's some unexpected incidents happen, the stress can pop up.”
		Lack of knowledge	P12 “What would help, I don’t really know. I guess if I know I would be able to reduce it.”
		Affordability	P18 “Maybe I'd join Zumba. I like that kind of thing but I'm thinking about affordability and the convenience.”
	Facilitators to healthy living	Early intervention + knowledge building	P05 “It's best to not just educate the older people, because we tend to think that the older people that they will be diabetic easier, so it's best to teach younger ones. “
		Empowering oneself	P05 - “So I have a treadmill at home and I use that. Like I walk as I’m doing some work on my laptop. So I try to I try to incorporate that. Or I might do maybe like 10/15 minutes of of yoga or cardio or something just around the house yea wherever I can.” – (P05)
		Moderation	P15 “Stop binge eating. not to have too much of every type.”
		Discipline	P19 “I definitely would agree, especially when in Singapore it is food paradise, you actually get good food. I would say it's definitely very hard to maintain that, but with a bit of discipline it should be fine.”
		Calorie counting	P17 “I think it does. If I look at the menu item and then it looks like this is 600 calories, which is top of the list out of the whole list of things that I’m looking at, I would think twice about eating it”
		Internal motivators	P18 “I just think of wanting to be healthy, so tend to discipline myself on my food intake, so that’s how I do it.”
		Reminders	P14 - “So, maybe when it comes to before lunchtime, two hours, tell them this is a suggested meal, then it’s a guideline. Because everybody like to hold their phone, and then the message pops up, and they say, oh, yes, why not, I go and try”
		Reduce sitting time	P17 “I will just stand up for longer after a meal rather than sitting down.”
		Group support	P16 “I join an exercise group.”
		Efficient exercising (less time, more yield)	P20 - “So, yes, I only exercise 15 minutes at each time. But I try to keep it more vigorous. So, yes, I do HIIT kind of training. So I do that

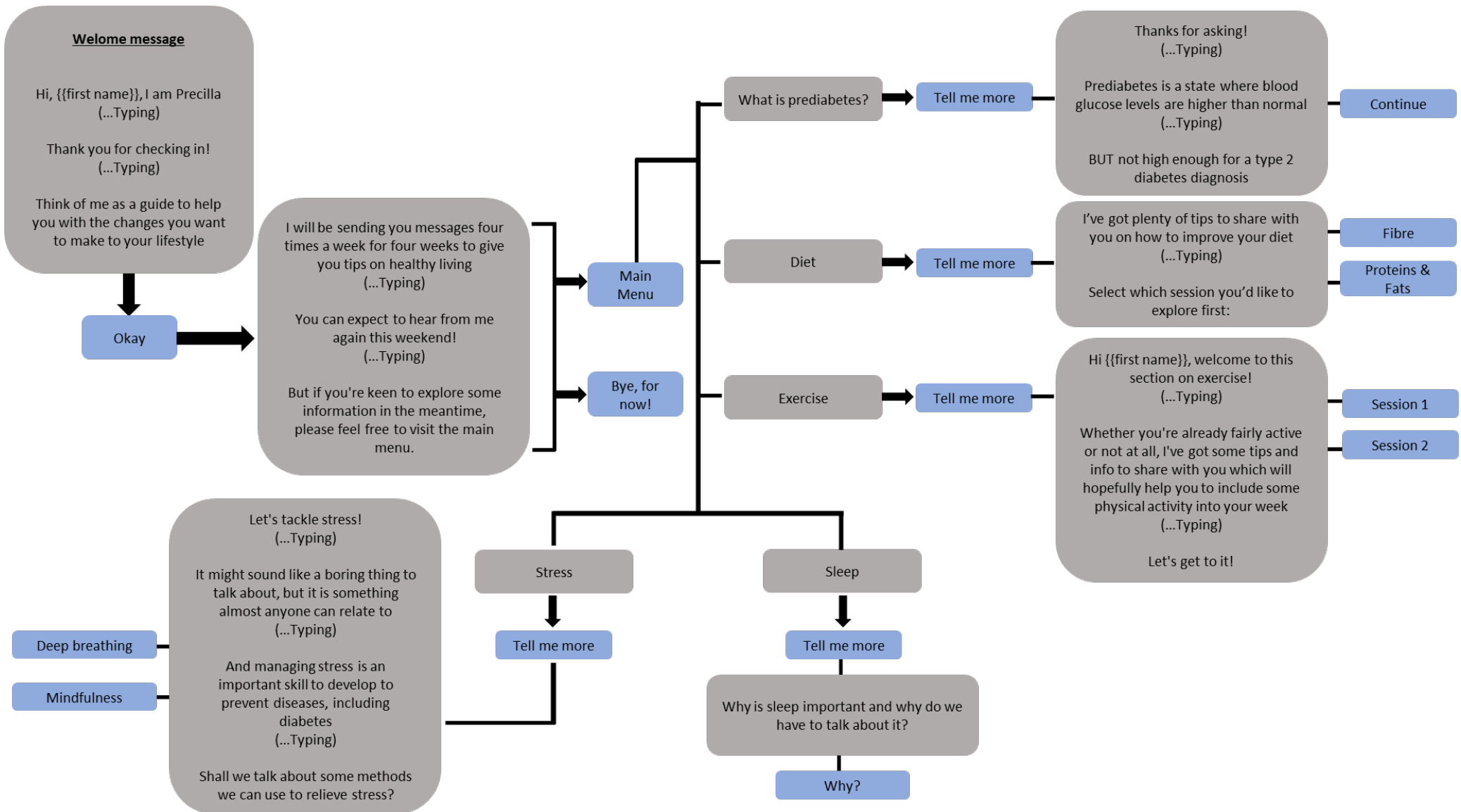
			about four or five times a week maybe.” “I think that there's more yield in doing vigorous exercise over moderate exercise. So in terms of the efficiency, the amount of time you need to spend, so I try to do more of HIIT.”

Appendix 7.1 – Sources of information for CA Precilla

Conversational domain	agent	Weekly components derived	Sources
Prediabetes & diabetes		First week introduction	War on diabetes: Singapore to host global conference on tackling epidemic [373] What is pre-diabetes [374] Prediabetes [375] Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance [376] Lean diabetes mellitus: An emerging entity in the era of obesity [377]
Diet		Week 1: Fibre Week 2: Proteins & Fats Week 3: Recap week 1* Week 4: Recap week 2	Walnut Consumption Is Associated with Lower Risk of Type 2 Diabetes in Women [378] Dairy Products and Prevention of Type 2 Diabetes: Implications for Research and Practice [379] Popcorn is more satiating than potato chips in normal-weight adults [380] Fruit and vegetable intake and incidence of type 2 diabetes mellitus: systematic review and meta-analysis [381] Intake of Fruit, Vegetables, and Fruit Juices and Risk of Diabetes in Women [382] The role of diet in the prevention of type 2 diabetes [383]
Sleep		Week 1: Bedtime routine Week 2: Stimulus control Week 3: Recap and knowledge recall of week 1 Week 4: Recap and knowledge recall of week 2	Sleepless no more: Techniques and Interventions for Sleep Disorders [384] The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis [313] Can music improve sleep quality in adults with primary insomnia? A systematic review and network meta-analysis [385] Insomnia: Relaxation techniques and sleeping habits [386] Too early to get up, too late to get back to sleep [387] The evidence base of sleep restriction therapy for treating insomnia disorder [388]
Exercise		Week 1: Aerobic & Anaerobic activity Week 2: Exercise options matched to preferences Week 3: Recap and knowledge recall of week 1 Week 4: Recap of exercise options to choose from (week 2)	Health hub SG: Aim for 150 minutes of physical activity every week [389] Physical Activity/Exercise and Diabetes: A Position Statement of the American Diabetes Association [390] Exercise for diabetes and get a leg up [391] Benefits of exercise and physical activity [392] Reversing Diabetes Through Weight Loss: How Much Should You Lose? [393]

		Different types of physical activity [394]
Stress	Week 1: Deep breathing Week 2: Mindfulness Week 3: Recall week 1 & relaxing music Week 4: Mindfulness recap	Stress Management Apps With Regard to Emotion-Focused Coping and Behavior Change Techniques: A Content Analysis [395] Managing and Controlling Stress Using mHealth: Systematic Search in App Stores [396] Stress Management: Breathing Exercises for Relaxation [397] Getting started with Mindfulness [398] Stress Management Techniques: evidence-based procedures that reduce stress and promote health [399]

Appendix 7.2 – Sample dialogue tree for CA Precilla (Week 1)



Appendix 7.3 – Questionnaires for Feasibility study

Quality of life (QOL) questionnaire - 14 item quality of life questionnaire. Each item was rated on a 5-point scale.

Taking everything into consideration, during the past week how would you rate your	Work	Family relationships	Sexual drive, interest and/or performance	ability to get around physically without feeling dizzy or unsteady or falling
Physical health	Household activities	Leisure time activities	Economic status	ability to do work or hobbies
mood	Social relationships	Ability to function	Living/housing situation	overall sense of well being

Knowledge questionnaire - Section 1.

For each item below, let us know the response that BEST DESCRIBES YOUR OPINION of diabetes and prediabetes	Scoring for Q1-6: Yes = 1 No = 2 I don't know = 3 Q7 answer options: Almost no chance, slight chance, moderate chance, high chance
Q1	Diabetes is a disease that occurs when your blood sugar is too high
Q2	Prediabetes is a condition in which someone's blood sugar levels are higher than normal but not high enough to be diagnosed a diabetic
Q3	People with diabetes have a higher chance of getting diabetes
Q4	Regular exercise may help prevent diabetes from developing
Q5	A healthy diet may help prevent diabetes from developing
Q6	Weight loss may help prevent diabetes from developing
Q7	What do you think your risk or chance is for getting diabetes over the next 10 years?

Knowledge questionnaire - Section 2

What statement best reflects your opinion for each?	Scoring: Strongly agree = 1 Strongly disagree = 4
Q1. I feel that I have little control over risks to my health	Q7. Over the past year I made conscious food choices to improve my diet
Q2. If I am going to get diabetes, there is not much I can do about it	Q8. Over the past year I chose to be active to improve my health
Q3. I think that my personal efforts will help control my risks of getting diabetes	Q9. Over the past year I lost weight through conscious efforts (ie. Not due to sickness or stress)

Q4. People who make a good effort to control the risks of getting diabetes are much less likely to get diabetes	Q10. In the coming year I will make food choices to improve my diet
Q5. Compared to other people of my same age, I am less likely that they are to get diabetes	Q11. In the coming year I chose to be active to improve my health
Q6. Compared to other people of my same age, I am less likely that they are to get a serious disease	Q12. In the coming year I plan to lose weight

Knowledge questionnaire - Section 3

We would like you to think about members of the general public here. Check the box for each statement that best describes your opinion.	Scoring: Increases or raises risk = 1 Has no effect on risk = 2 Decreases or lowers the risk = 3 Don't know = 4
Q1. Having a sibling or a parent with diabetes	Q5. Having high blood pressure
Q2. Being 40 years old and above	Q6. Having abnormal blood cholesterol or lipid levels
Q3. Being overweight or obese	Q7. Having a history of gestational diabetes
Q4. Leading an inactive lifestyle	Q8. Having prediabetes

Diet – Food frequency questionnaire

The next couple of questions will be about your diet.

We would like to know how often you consume each of the following food items per day.

Please select the option which best represents your daily intake of 1 serving of the following food and drinks.

Vegetables

- Almost never
- Once
- Twice
- Three times
- More than 3 times

(same options below)

Fruit (exclude juices)

Sugar-free sweetened beverages

Fried food/snacks

Physical activity Questionnaire

1. In the last 7 days, can you please indicate how many days you spent doing:

- vigorous physical activity like heavy lifting, aerobics, fast bicycling?
- moderate intensity activity like carrying light loads, regular paced cycling, doubles tennis?
- low intensity physical activity like walking?

- 0 days
- 1-3 days
- 5-7 days

2. How much time did you usually spend doing:

- vigorous physical activities on one of those days?
- moderate intensity physical activity on one of those days?

- 10-20 mins per session
- 20-30 mins per session
- 30-60 mins per session
- >60 mins per session

3. If you were physically active, could you please indicate the length of each exercise session? (in minutes)

- 10-20 mins per session
- 20-30 mins per session
- 30-60 mins per session
- >60 mins per session

During the last 7 days, how much time (in hours per day) did you spend sitting on a weekday?

[Short answer text]

Stress Questionnaire

The questions in this scale ask you about your feelings, thoughts and stress levels during the last month. Please do feel free to be honest. Your responses will be kept confidential amongst the study team members, please do not worry about your data being shared publicly.

For each question, please indicate how often you felt or thought a certain way in the past month.

Been upset because of something that happened unexpectedly?

- Never
- Almost never
- Sometimes

- Fairly often
- Very often

(same options throughout)

Felt that you were unable to control the important things in your life?

Felt nervous and "stressed"?

Felt confident about your ability to handle your personal problems?

Felt that things were going your way?

Felt that you could not cope with all the things you had to do?

Been able to control irritations in your life?

Felt that you were on top of things?

Been angered because of things that were outside of your control?

Felt difficulties were piling up so high that you could not overcome them?

Sleep Questionnaire

The following questions relate to your usual sleep habits during the past month only. Your answers should indicate the most accurate reply for the majority of days and nights in the past month.

- 1. What time have you usually gone to bed at night? (Please use 12 hr times with am/pm (eg. 10.30pm))**

[Short answer text]

- 2. How long has it usually taken you to fall asleep each night?**

- <30 mins
- 30-60 mins
- 1-2 hrs
- >2 hrs

- 3. What time have you usually gotten up in the morning? (Please use 12 hr times with am/pm (eg. 9.15am))**

[Short answer text]

- 4. How many hours of actual sleep did you get at night on weekdays?**

- <6 hrs
- 6-7 hours
- 8 hours

- >8 hours

5. How many hours of actual sleep did you get at night on weekends?

- <6 hrs
- 6-7 hours
- 8 hours
- >8 hours

6. During the past month, how often have you had trouble sleeping?

- Yes
- No

7. (if yes above) How often have you had trouble sleeping?

- Not during the past month
- Less than once a week
- Once or twice a week
- Three or more times a week

8. Could you please select from the list provided, the potential reasons for your sleeping problems? You may choose multiple answers.

- I Cannot get to sleep within 30 minutes
- I Wake up in the middle of the night or early morning
- I Get up to use the bathroom
- I Cannot breathe comfortably
- I Cough or snore loudly
- I Feel too cold
- I Feel too hot
- I Had bad dreams
- I Have pain
- Other

9. If you have reasons besides those stated, please describe it here (short answer text)

10. During the past month, how would you rate your sleep quality overall?

- Very bad
- Bad
- Decent
- Good
- Very good

How often have you taken medicine to help you sleep (prescribed or over the counter)?

- Not during the past month

- Less than once a week
- Once or twice a week
- Three or more times a week

How often have you had trouble staying awake while driving, eating meals, or engaging in social activity?

- Not during the past month
- Less than once a week
- Once or twice a week
- Three or more times a week

How much of a problem has it been for you to keep up enough enthusiasm to get things done?

- No problem at all
- Only a very slight problem
- Somewhat of a problem
- A very big problem

Satisfaction questionnaire

The four specific questions were:

Q1 - How would you rate your overall satisfaction with the healthy lifestyle program delivered by Precilla?

(Answer options: Very satisfied, moderately satisfied, dissatisfied)

Q2 - How likely are you to recommend "Precilla" to others?

(Answer options: Extremely likely, highly likely, somewhat likely, unlikely, very unlikely)

Q3 - In the future if a healthy lifestyle program is available through Precilla, how likely would you want to use Precilla again?

(Answer options: Extremely likely, highly likely, somewhat likely, unlikely, very unlikely)

Q4 - As a result of the help you received from Precilla, would you say your health is...

(Answer options: neither better nor worse/somewhat better than before/much better than before.)

Appendix 7.4 – User-CA engagement data in feasibility study

	No. of completed interactions	No. of immediate interactions	No. of delayed interactions	Average duration of each interaction (mins)	Number of incomplete interactions	Number of absent interactions	Total (complete + incomplete + absent)	Reasons for incomplete interactions (if any)
Week 1								
Diet	41	28	14	2.25	1	13	55	(1) free text (can you tell me more?), (2) possible tech difficulty. "Fibre" flow appeared twice. Left and returned to convo 30 mins later, (3) possible tech difficulty. Multiple repetitions of same block. Keep pressing the button?
Exercise	40	21	27	3.54	6	9	55	(1) free texting (eg. <i>Can I please redirect you to the main menu?</i> That's fine, no need. <i>Are you ready to get started?</i> Sure. <i>do you exercise?</i>
Sleep	42	21	29	3.67	6	8	56	(1) Detour to main menu caused by free text entry
Stress	47	27	26	2.59	3	6	56	(1) possible tech difficulties
Week 2								
Diet	38	18	28	1.97	7	11	56	(1) Free text = conversation detoured to main menu, (2) possible tech difficulties
Exercise	38	11	34	4.11	6	12	56	(1) free text = conversation detoured to main menu
Sleep	43	21	30	2.44	7	6	56	(1) possible tech difficulties. Multiple pressing of the same button.
Stress	39	25	22	1.97	6	11	56	(1) Possible tech difficulties. Same button recurring. (2)

								Free text = conversation detoured to main menu
Week 3								
Diet	41	22	23	2.10	1	14	56	NA
Exercise	30	25	15	4.00	7	19	56	Possible tech difficulties? Same button recurring.
Sleep	40	13	32	1.80	5	11	56	(1) free text (thumbs up) conversation detoured to main menu and halted
Stress	38	26	18	2.11	4	14	56	NA
Week 4								
Diet	35	15	28	2.34	5	16	56	NA
Exercise	40	19	24	3.15	3	13	56	(1) free text " <i>grapeseed</i> . <i>Not rapeseed</i> ". Conversation halted at this point.
Sleep	41	17	24	1.56	1	14	56	NA
Stress	39	18	27	2.26	4	13	56	NA

Appendix 8.1 – COREQ checklist for follow up interviews

Item	Guide questions/description	Reported in Section # of thesis
Domain 1: Research team and reflexivity		
1. Interviewer/facilitator	Which author/s conducted the interview?	Chapter 8 – Section 8.2.2
2. Credentials	What were the researchers' credentials?	Chapter 8 – Section 8.2.2
3. Occupation	What was their occupation at the time of the study?	Chapter 8 – Section 8.2.2
4. Gender	Was the researcher male or female?	Chapter 8 – Section 8.2.2
5. Experience and training	What experience or training did the researcher have?	Chapter 8 – Section 8.2.2
6. Relationship with participants established	Was a relationship established prior to study commencement?	No
7. Participant knowledge of the interviewer	What did the participants know about the researcher?	Chapter 8 – Section 8.2.2
8. Interviewer characteristics	What characteristics were reported about the interviewer/facilitator?	Chapter 8 – Section 8.2.2
Domain 2: Study Design		
9. Methodological orientation and theory	What methodology orientation was started to underpin the study?	Chapter 8 – Section 8.2
10. Sampling	How were participants selected?	Chapter 8 – Section 8.2.1
11. Method of approach	How were participants approached?	Chapter 8 – Section 8.2.1
12. Sample size	How many participants were in the study?	Chapter 8 – Section 8.3
13. Non-participation	How many people refused to participate or dropped out? Reasons?	None
14. Setting of data collection	Where was the data collected?	Chapter 8 – Section 8.2.2
15. Presence of non-participants	Was anyone else present besides the participants and researchers?	No.
16. Description of sample	What are the important characteristics of the sample?	Chapter 8 – Section 8.3.1
17. Interview guide	Were questions, prompts, guides provided by the authors?	Yes. Appendix 8.2
18. Repeat interviews	Were repeat interviews carried out?	No.
19. Audio/visual recording	Did the research use audio or visual recording to collect the data?	Chapter 8 – Section 8.2.2
20. Field notes	Were field notes made during and/or after the interview?	Chapter 8 – Section 8.2.2
21. Duration	What was the duration of the interviews?	Chapter 8 – Section 8.2.2
22. Data saturation	Was data saturation discussed?	Chapter 8 – Section 8.2.2
23. Transcripts returned	Were transcripts returned to participants for comment and/or correction?	No.
Domain 3: analysis and findings		
24. Number of data coders	How many data coders coded the data?	Chapter 8 – Section 8.2.3

25. Description of the coding tree	Did authors provide a description of the coding tree?	Appendix 8.3
26. Derivation of themes	Were themes identified in advance or derived from the data?	Chapter 8 – Section 8.3.2
27. Software	What software, if applicable, was used to manage the data?	NA
28. Participant checking	Did participants provide feedback on the findings?	NA
29. Quotations presented	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified?	Chapter 8 – Section 8.3.2
30. Data and findings consistent	Was there consistency between the data presented and the findings?	Chapter 8 – Sections 8.3.2 and 8.4.1
31. Clarity of major themes	Were major themes clearly presented in the findings?	Chapter 8 – Section 8.3.2
32. Clarity of minor themes	Is there a description of diverse cases or discussion of minor themes?	Chapter 8 – Sections 8.3.2 and 8.4.1

Appendix 8.2 – Interview guide for follow up interviews

Duration: 20 mins

Participants: 20 participants who took part in our pilot study

Language: English

Aim: To elicit participants views on mHealth lifestyle modification intervention they have received over the previous 4 weeks

Format: a telephone delivered interview & survey

1. Introduction

Thank you for agreeing to this call. My name is [NAME], I am a researcher at LKCMedicine and the member of the mHealth study research team and I would like to ask you several questions in relation to the study you recently took part. Would that be ok with you?

Today we will be discussing your views in relation to the study you took part in over the last 4 weeks. This was a lifestyle modification focused intervention delivered via a mobile phone and I would be grateful for your views on it. There are no right or wrong answers. All comments are welcome.

This call should not last more than 20 minutes. Everything you tell us will be kept confidential. To protect your privacy, we won't connect your name with anything that you say. At any time during our conversation, please feel free to let me know if you have any questions or if you would rather not answer any specific question. You can also stop the interview at any time for any reason.

Do you have any questions? Thank you.

2. Demographics

Can I first ask you a few questions about yourself? Thank you.

Age: ☐ 25 – 30 years old
☐ 31 – 35 years old
☐ 36 – 40 years old
☐ 41 – 45 years old
☐ 46 – 50 years old
☐ 51 – 55 years old
☐ 56 – 60 years old
☐ >60 years old

Ethnicity: ☐ Chinese
☐ Malay
☐ Indian
☐ Others (Please specify): _____

Marital Status: ☐ Single
☐ Married
☐ Others (Please specify): _____

**Monthly Household
Income:**

- ☐ below \$4,000
☐ \$4,000 – \$6,000
☐ \$6,001 – \$8,000
☐ \$8,001 – \$10,000
☐ \$10,001 – \$12,000
☐ \$12,001 – \$14,000
☐ >\$14,000

Family Structure

No. of children & age(s): _____
No. of dependents: _____
Type of dependency: _____
Partner's Job profession: _____

What do you do for living?

Do you have prediabetes?

(If yes) How did you learn you had prediabetes?

Does anyone else in your family have prediabetes or diabetes? If yes, who?

In Singapore, do you feel like you receive enough information about diabetes and prediabetes? Or do you have a lot of questions which remain unanswered?

3. Thank you for answering these questions. Study-related questions

I will now proceed to ask you a few questions in relation to the study itself.

How did you feel about the study? Can I ask if you currently use other sources of information and advice for healthy living? (eg. Apps/physical activity tracker/healthy 365/ health hub etc.)

Did you prefer this chatbot to other sources providing information and advice on healthy living and diabetes prevention? Eg. The HPB website/other health apps in Singapore?

CONTENT: What are your views on the content of the intervention? What about the focus of the messages – diet, physical activity, weight loss etc.? Was some information missing? Was some information unclear?

FORMAT: What about the format of the messages? What did you prefer: text, animations or images? Why?

What did you think of Facebook messenger as a delivery platform? Would you have preferred a different interface? (eg. Whatsapp, telegram or a standalone app?)

TONE: What about the tone of the messages? Was it too (in)formal or (in)personal?

FREQUENCY: What about the frequency of the messages? Would you have liked to receive messages more frequently or less frequently?

TIMING: What about the timing? Morning, evening, afternoon? Weekday vs weekend?

INTERACTIVITY: The messages were conversational. What is your view on this? Would you have preferred one-way communication? Why? What did you think of the response options provided? Were they accurate in terms of how you would reply generally? Did you feel restricted by the buttons provided? If you could have typed free text, what would you have liked to write? Ask? (eg. Questions, comments etc.)

DURATION: What are your views on the study duration? Would you have preferred to receive messages for a longer period of time? Stop earlier?

IMPACT: Do you feel that this intervention had any effect on your lifestyle? Could you share more? Why (not)? Would you be able to maintain the changes in your lifestyle (if any have been made) after the intervention?

SENDER: Who did you imagine was sending you the text messages? What was their intention? Did talking to Precilla feel like you were speaking to another person? b. Did you feel that anyone was looking at the messages? How would you characterize the person(s) sending and/or reading the messages? Did this change over time? Did you ever feel like you might let someone down or be judged if you failed to

meet your physical activity goals?

At some points in the conversations, you were provided with links to external webpages. Did you find this useful or not so useful? Would you have liked to have access to more links? What sorts of links would you be interested in?

GENERAL COMMENTS: What did you like/dislike? What did you find (un)helpful? Have your views changes throughout the study? What types of messages did you find to be the most helpful? Unhelpful? Surprising? Confusing? Would you recommend it to your friends and family?

Closing

Is there anything else you would like to add on the topics we have discussed today?

Thank you very much for participating in today's discussion and for your valuable input. Your comments will be very helpful to us in developing an intervention for prevention of diabetes in people at risk of diabetes.

Appendix 8.3 – Follow up interview codebook from qualitative analysis

Theme	1 st Level code	2 nd Level code	Quotes
Enjoyable and acceptable experience with Precilla			
	Precilla's positive traits		
		Easy to understand	P147 "Most helpful I would say it's the way the information was raised, it was very understandable and simple to understand"
		Informative	P035 "It is informative. It's quite clear because it comes from general, then they go into more detailed information later, so it's quite clear."
		Aesthetically Pleasing	P078 – "It looks nice," P062 "It did not become too much images, which could be a bit jarring because it's only images, and neither was it too text or too wordy, so it presented a good blend."
		Reliable/Accurate	P080 "I feel like it was a healthcare professional."
		Personality	P124 "I liked that she was approachable. I liked the fact that the conversation was friendly."
		Visual aids	P061 "I like how you guys show pictures, not just text, along with it. That's a plus point. And it's very easy to be digested. It's quite good."
		Readily available information	P061 "But what I liked about it is there is the information that she feeds, I may or may not actually read it, but I know that the information is there. If I want to, I can always refer back" P140 "What I like about it is that I can check it on my own time"
		Appropriate Tone (friendly, casual, slow with pauses)	P147 "Leaning towards formal, but it's quite appropriate." P063 "It came across very friendly, casual, which I think is great." P102 ""But those pauses are good to let the message sink in"
	Preference for Precilla over other sources		P108 - I prefer Precilla. P062 "Yes, I definitely preferred using Precilla."
		Ministry of health	P077 "Actually, I've been receiving more information about prediabetes, but that's maybe because that's what the Health Hub app is pushing out."
		Exposure to health information through work	P140 "The thing is, I'm in this field like you, so I think I already have some information, prior."
		Health tracking apps	P108 - I do use this Healthy 365, it's a Singapore app. I use it to monitor my diet and everything.
	Positive effects on lifestyle		

		Sleep improvements	P062 “Yes, definitely. Me, personally, from all this information, especially for sleep in particular, just like the example, there were some tips and tricks that Priscilla shared. I'm actively trying to use them in my day to day because sleep, personally, is one of my issues.”
		More active	P130 “I usually procrastinate to go exercising, so it kind of spurred me on that I should.”
		Less stress	P080 “It did impact on me, especially the stress management.
		Healthier diet	P147 “I was more mindful about my diet and exercise after that.”
		Regular reminders	P125 “I think it gave good reminders. I liked that there were reminders on healthy living.”
	Helpful content		
		General	P124 “Generally it was good, the information was good.”
		Mental health & stress management	P143 “But I think it’s great that you guys did that because I think mental health is a very important field that is not really talked about in Singapore. I think it was a great idea that you guys did it.”
		Food recommendations	P125 “I think there was a bit of emphasis on diet.”
		Proper sleep habits	P147 “but I also find that introduction of proper sleeping habits and the mindfulness practice was also a very good addition to the chatbot.”
		Physical activity advice	P124 “I liked the physical activity one. The sleep one was quite comprehensive also.
	Recommendations to friends and family		
		For sedentary people	P080 “Yes, I will. For the very sedentary people. It's a good start.”
		Those at high risk of diabetes	P102 “For Priscilla, maybe those who are in their late 20s to maybe their early 40s, those who are at a higher risk of getting prediabetes and diabetes serves as a good reminder for them especially”
		Conditional recommendation (if Precilla is more human, more personalised, translated etc.)	P098 “Not at the current stage. But when she’s a bit more human I suppose and perhaps a tad more personalised, that will be great.”
Sub-optimal experience(s) with Precilla			
	Minimal changes to lifestyle		
		No intention to change lifestyle	P140 “It presented some new information, but I wasn’t planning to make any changes to my lifestyle to begin with.”
		Lack of long-term guidance (to maintain changes)	P126 “Not really. I don’t think it’s sustainable because it’s just a tip for the four weeks.”
	Unhelpful content		

		High prior knowledge (nothing new learnt)	P080 "The least helpful for me would be the exercise component" "Maybe because I'm quite well versed with the content, so it's more like I'm checking the content than to use the content in a sense"
		Non-actionable content (how to put the knowledge into practice?)	P111 "The least helpful for me is what I mentioned, it's just informative but it doesn't really push you to put the information into real work."
		Not relatable to one's current life circumstances	P125 "Least helpful, I think is about sleep because I have a six-month-old so I can't sleep even if I want to. I think that wasn't useful to me."
		Need for physical instruction of some concepts (eg. mindfulness)	P143 "I think mindfulness is possibly something that requires more, I guess, someone to really teach you properly. I think it's a difficult concept to grasp"
		Some repetitive content	P143 "It can even be longer, but I guess because at some points in time during the study, itself I found the content to be quite repetitive sometimes, if there can be more content added in, then I think it's good for it to be a longer time."
	Unfavourable user experience		
		Lengthy interactions (too slow/boring)	P078 "Too much text. Each message is a bit long." P125 "I thought just that they were not very short and sweet [?], but I guess it's because it has good information to share that's why." "It was okay. But that was a bit lengthy."
		Insufficient answer options	P078 "Sometimes we want to interact with it, so we don't want to follow exactly what it gives us."
Comparing Precilla to other sources			
	Existing platforms more familiar		P077 "I wouldn't prefer the chatbot over other sources." "In terms of finding things online, it's something that we are habituated to doing already, so you can look for specific queries pretty quickly. With the chatbot, also because I wasn't too familiar with the flow"
	Similar to existing conversational agents		P124 "It was fine. It's no different from what I have seen in some standalone chatbots like Wysa or Woebot, those mental health."
Favourable aspects of the intervention			
	Timing		
		Stratified timings for specific messages (eg. healthy eating message before mealtime)	P078 "Yes, if it's about sleep time, maybe eight or nine pm, and then we can just follow the suggestions."

		After work/ when at home (alone + no need to use data)	P102 “. I think it will be better if it's after work maybe right before dinner. That's the time where either you are eating or you're going to go for exercise, gyms or wherever, right after work.” “Because during lunchtime, everybody's lunchtime is not fixed.”
		Mornings before work	P125 “Maybe morning would be a good time also when you're starting work and you're just about to get started, so you are trying to clear your messages on your phone or something and you can have a quick glance.”
	Bi-directional communication		
		One way communication is okay sometimes	P125 “I'm okay to just skim through rather than actively participate, actually. I barely reply to messages anyway even from friends and stuff. I'm okay with just a one-way conversation.” “... I feel that if I were to ask question, I can just google.” P038 “Because it's only through a bot, it's not a human interaction so based on a bot, that's sufficient enough.”
		More answer options	P124 “More options would have been okay. You didn't really need to have the free text option, maybe it's good to have it also but with some more options.”
		Free text option	P147 “If it would be possible to introduce maybe an open-ended question option, that would be good” Making it conversational helps the information stick better, because if it's just one way text then you are essentially reading from a very long paragraph, so you're forced to concentrate and focus on what each message is saying before picking the next.”
	Delivery platform		
		Facebook messenger acceptability	P140 “Well, for me, I think Facebook is fine, because I'm primarily on Facebook, but the younger people seem to be more Instagramming. I'm not sure how that works, so Facebook is fine for me.”
		Preference for standalone app	P080 “A standalone app would be great”. P035 “Maybe can start with an app.”
		Preference for other messaging app	P062 “Probably Telegram or WhatsApp would have been a better option because, for instance, me, personally, if I did not actually log in or switch on my notifications for Facebook Messenger, I would miss out that that Precilla has an update for me on that particular day.”
	Conversational agents can be supported		
		Integrate with an online course	P077 “If that was the case, wouldn't it be easier to just have it as a wiki page, or in a more, quote, unquote, traditional online course format rather than someone I have to converse with.”

		Combine with HCP consultations	P038 "Because I had done another research probably a year ago, one that actually has got a dietician. A one-off, they have a bot also, but separately they also have a dietician so the chatting was more interactive."
Suggestions for the future			
	Different message types		
		Voice messages	P078 "Maybe a voice. We always find that listening is easier than just reading the text."
		Video messages	P098 "It's like when you try to find things on how to do certain exercises. When they have either drawings or when it's just descriptive, it's usually a little bit hard. I think what can be improved possibly is videos."
		Better default/error message (outline conversational agent's limits and how the user can get help)	P077 "On a more basic level, to go back on the main menu, like, because when I tried typing like main menu and stuff, there isn't a command, or at least to my knowledge, in using the chatbot. So the only way to trigger Precilla to take you back to the main menu was to type something wrong or type anything, so that was slightly annoying because you didn't know."
	Personalised content		
		Age group specific content	P125 "Missing? I'm not sure whether there was age related content for different age groups." "And also different age groups have different requirements."
		Options for more detailed content (for selected interest areas)	P126 " I think the app can be more specialised because it's quite a wide topic on general health. So let's say you want to delve into a topic, you can actually choose it and it will be more better for people who want to know more about certain topics, not just everything in general."
		Search function to retrieve specific information	P061 "but I'm not sure if there is actually a search function where you can look through the conversation that you had. Because she sends quite a lot of messages, sometimes you want to look back into certain things that she mentioned before."
		Follow up notifications for unread messages	P098 "Because if the notification shade is not there, then I will not remember it." P080 "I'm not sure if there's such a thing whereby if I've not clicked on it, then send a reminder. Whereas if I click on it or read, no reminder."
		Content translated into languages relevant to Singapore	P098 "But the only thing I suppose that would be of concern is if let's say you are looking to broaden your target audience, then you might have to send it in other languages like Chinese and Malay. Because there are definitely, especially the older generations, they are not going to be able to read English. That's for sure. And then otherwise, you have people

			who are, say, from Malaysia who also will not be too good with reading English.”
	Suggestions for website links		
		Helplines	P080 “It will be great if there is help links that are provided, so what if I tried and I'm still not good at it, how I can get better help, and because I will trust that, the chatbot will give me more update information, so the links would have helped.” “For me it's more like links to helplines.”
		Education talks/courses to attend	P080 ““Yes, or maybe there are talks or courses that are around in some medical institution, in the hospitals, poly clinics, anything like that. So then I know if I want to learn more about stress management somebody is giving a talk somewhere.”
		Links to more in-depth information	P063 “but more informational type links because the information Precilla provided was very simple.”
		Doctors/HCPs giving interviews on healthy living	P108 “I was thinking that maybe you can add some reports or interviews from a real doctor speaking about their experience or their views or their advice on diabetes or prediabetes. I think it will be more relevant and we will feel more secure”
	Novel/additional content		
		More on prediabetes (DM more commonly talked about)	P108 “I think we know some information about diabetes but not really prediabetes.”
		Signs and symptoms for diabetes and prediabetes	P038 “. I don't recall, was there advice to look up what the symptoms are if you're prediabetic? The signs or symptoms to look up for.”
		Context/race specific food recommendations (esp. for those at high risk)	P143 “Yes, especially since I think Singapore, the bigger problem for diabetes in particular is, e.g., for Indians, so, yes, it would be good to... It can bring in race specific food.”
		difference between healthy foods and healthier options	P102 “They promote like, this milk tea is a healthier choice, but it's not healthy. This difference between healthier choice and whether something is healthier or not, it should be the emphasis for the group who are really educated”

Appendix 9.1 – Search terms for systematic apps assessment

Search Term	iOS app store	Android play store
1. Healthy lifestyle chatbot	3	3
2. Healthy lifestyle conversational agent	0	0
3. Fitness chatbot	11	13
4. Fitness conversational agent	0	0
5. Good health chatbot	9	9
6. Good health conversational agent	0	0
7. Health companion	844	710
8. Medical companion	291	286
9. Medical buddy	49	50
10. Health friend	451	467
11. Medical coach	310	207
12. Health buddy	185	169
13. Health chat companion	43	42
14. Chatbot	158	128
15. AI chatbot	50	46
16. Chat assistant	81	107
17. Stress chatbot	30	19
18. Mental health chatbot	35	25
19. Sleep chatbot	25	18
20. Healthy diet chatbot	2	3
21. Healthy eating chatbot	3	4
22. Exercise chatbot	7	11
23. Physical activity chatbot	4	2
24. Health artificial intelligence	277	234
25. Artificial intelligence chatbot	6	2
26. Artificial intelligent chatbot	24	18
TOTAL:	2898	2573

Appendix 9.2 – Apps assessment data extraction sheet

Field	Description	Options
App Description (from App Store)		
App ID	A unique ID number generated for each app for easy identification and merging of data.	Alpha-numeric
App Name	The given name/title of the app.	Free text
Developer name	The company of the app developer. If there is no name, note the developer's name.	Free text
Platform	Is this app tested on Android or iOS?	Android/iOS
App version number	List the latest version of the app during assessment	Alpha-numeric
User rating	State user ratings if applicable on a scale of 5.	Numeric
Number of people who rated	State the number of people who provided ratings for the app at the point of assessment.	Numeric
Cost (App or in-app purchases)	Cost to download the app or extra features.	Numeric
Accessible region	Is this app limited to certain locations/ countries/ regions?	Free text
Target user	Who is the target user of the app?	Adult Patient/ Caregiver/HCP/ others
Aspect of healthy living	Which aspects of healthy living does the chatbot target?	Diabetes prevention/general healthy living (free text)
CA specific information		
Intelligence of CA	Is the CA app rule based or artificially intelligent?	Rule based/AI/Both
Input modalities (input all that apply)	Text/ Voice/ images/ Video	

Output modalities (input all that apply)	Text/ Voice/ images/ Video	
Default message	Is there a default message that pops up when the user enters information that the CA cannot understand or recognise?	Yes/No (what is the message?)
Avatar	Is the CA embodied?	ECA/Non-ECA
	What does the avatar look like?	Free text
	Can the user make changes to the appearance?	Yes/no
Personality	What is the personality of the CA. (options from our scoping review)	Coach-like, CA identity, culture specific, factual, gender specific, healthcare professional like, human-like, informal, knowledgeable
Personalization	Does the app have any features which promote personalisation? (eg. Timings, record a history of user's preferences/health status, goals etc.) *Addressing user by name alone is not personalisation?	Yes/No
	If yes, does it use personalised timings?	
	If yes, does it achieve personalisation by recording a history of the user's preferences?	
	If yes, does it record a history of the user's health status?	
	If yes, does it keep a record of user's goals?	
	If yes, what other method of personalisation does the app employ?	
Health tracking	Does the app have any function related to tracking of health (e.g. physical activity, food intake, water consumption etc.).	Yes/No
Information & education	Does the app provide information on healthy living in general?	Yes/No

	Does the app provide more specific information to educate the user on a particular health condition of focus? (risk factors (psychological or environmental), safety plan, emergency management?)	
Evidence	Is the information in the app backed by evidence (could be a scientific paper, or a reputable website, e.g. NHS Health A to Z, Medline Plus, etc)?	Yes/No
	If yes, include citation for the research paper(s)	
	Has the CA been evaluated in any research studies?	Yes/No
	If yes, include citation for the research paper(s)	
	Have the developers published any scientific articles related to the design and development of the CA?	Yes/no
	If yes, include citation for the research paper(s)	
Human involvement	Does the app have a separate chat channel to allow the user to share information or contact people in her/his support network? (e.g. family members, friends, support groups, others)	Yes/No
	Does the app include a health professional who is accessible to the user (without users inputting any health provider/ health professional information)?	Yes/No
Emergency safety netting	Does the CA allow the user to contact emergency services directly through the app?	Yes/No
Other attributes		
	Can the app create user profiles?	Yes/No
	Does the app allow for user to remain anonymous while using it?	Yes/No

	Is the account password protected?	Yes/No
	Does the app provide user with rewards for completing tasks/ education/ condition monitoring?	Yes/No
	If yes, elaborate on what is being rewarded	Free text
	If yes, elaborate on the type of reward (e.g. Stars, motivational messages, etc.)	Free text
	Does the app contain any games?	Yes/No
	If yes, what is/are the purpose of the game(s)?	Free text
	Does the app gets stuck, crash or frozen more than once when using? (Answer "Yes" if app hangs for more than 3s in two occasions)	Yes/No
	Does the app allow recorded data of the patient to be shared on social media (e.g. Facebook/twitter?)	Yes/No
	Does the app allow data to be exported (e.g. PDF/Excel, email, cloud, etc.)?	Yes/No
	Please record your overall impressions of this app and any other observations from this app testing.	Free text
mHON code principles		
Authority	Does the app indicate the qualifications of specific individuals who developed the app or contributed to the information provided?	Yes/no
Complementarity	Is there a disclaimer stating or which implies that the information provided and/or app functions do not replace the healthcare provider's advice?	Yes/No/Advice given by healthcare provider
Confidentiality	Is there a privacy and confidentiality clause in the app?	Yes/No
	Is the privacy policy easily accessible within the app?	Yes/No
	Is consent to data collection required at the first launch of the application?	yes/no

	Where the data is stored (on the application, on the SD card, etc.)	Free text
	Is the data transmitted to third parties?	yes/no
	If yes, to which third parties are they transmitted?	
Validity	Is the date of last general update provided?	Yes/No
Justifiability	Does the health information have references, is it complete and provided in an objective manner?	Yes (claims are backed up by evidence)/ Inconsistent/No (claims not backed up by evidence)/ NA (no claims)
User's practice	Is the target audience clearly stated? (eg. Patients, health professionals etc.)	Yes/No (elaborate on audience)
	If the application is prohibited for use by minors, is it clearly indicated and designed in a way that use by minors is impossible?	Yes/No (how?)
	Are the developers contactable by email? (the email may be found in the app, the app store, app developer's website, etc)	Yes/No
	Are there instructions provided so that the application is easy to use without any bugs?	Yes/No
Financial disclosure	Does the app indicate any funding sources?	Yes (the app is managed by a registered commercial company) No (the app is managed by an individual developer without funding information)
	If Yes, which body created/ commissioned/ partnered the app?	Free text
	If no, who created/ commissioned/ partnered the app?	Free text
Advertising Policy	Are advertorials distinguishable from content of the app?	Yes/No/No advertising
	Do the adverts interfere with the functioning of the app (if any?)	Yes/No/No adverts

