

Exploring the Potential of Virtual Agents in Atrial Fibrillation Management: Insights from a Randomized Trial

Mina Fallah
fallah.m@northeastern.edu
Northeastern University
Boston, Massachusetts, USA

Timothy Bickmore
t.bickmore@northeastern.edu
Northeastern University
Boston, Massachusetts, USA

Stefan Olafsson
stefanola@ru.is
Reykjavik University
Reykjavik, Iceland

Michael Paasche-Orlow
michael.paasche-
orlow@tuftsmc.edu
Tufts University School of Medicine
Boston, Massachusetts, USA

Andrew Joseph Mrkva
jm240@pitt.edu
University of Pittsburgh
Pittsburgh, Pennsylvania, USA

Jared W. Magnani
magnanij@pitt.edu
University of Pittsburgh
Pittsburgh, Pennsylvania, USA

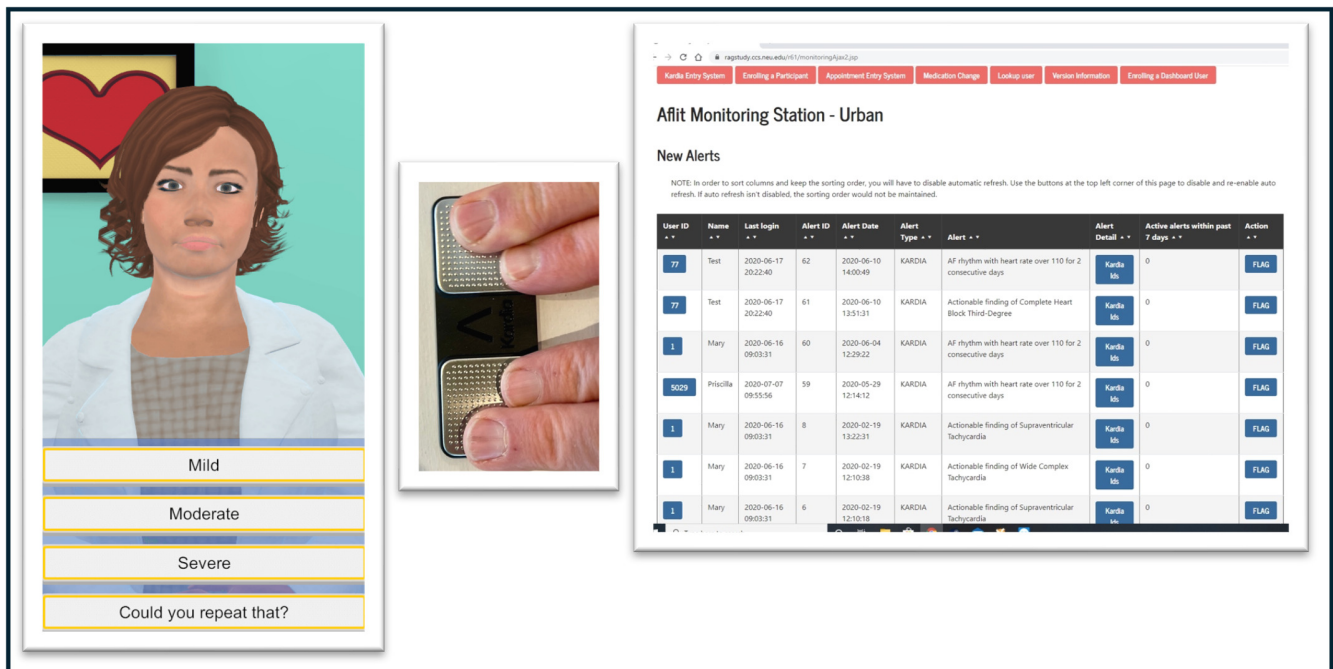


Figure 1: Virtual Agent System for Atrial Fibrillation Management, comprised of Smartphone-based Virtual Agent (left), Mobile ECG sensor (middle), and Clinician monitoring workstation (right)

ABSTRACT

Smartphone-based conversational agents offer a convenient way to deliver health education, particularly for managing complex health conditions such as chronic diseases. The present study explores using a smartphone-based virtual agent system to aid patients with a chronic heart condition—atrial fibrillation—to manage their health

by encouraging the use of a heart rhythm sensor integrated with their smartphones. We report the results of a randomized clinical trial with 240 patients experiencing atrial fibrillation who were provided with smartphones and heart rhythm sensors for 4 months. Participants in the intervention group interacted with a virtual agent, while the control group received general health education via the WebMD app. Intervention participants completed a median 91 interactions with the agent over the 4 months of the study period, and agent features designed to increase engagement—such as storytelling—were effective at increasing use. The agent’s promotion of heart rhythm sensor use was effective, with intervention participants taking significantly more heart rhythm readings compared to those in the control group. Participants were followed



This work is licensed under a [Creative Commons Attribution International 4.0 License](https://creativecommons.org/licenses/by/4.0/).

IVA '24, September 16–19, 2024, GLASGOW, United Kingdom
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0625-7/24/09
<https://doi.org/10.1145/3652988.3673955>

for 8 months thereafter to assess the sustainability of the effects, specifically focusing on medication adherence.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; *Empirical studies in HCI*.

KEYWORDS

Virtual Agents, Mobile Applications

ACM Reference Format:

Mina Fallah, Timothy Bickmore, Stefan Olafsson, Michael Paasche-Orlow, Andrew Joseph Mrkva, and Jared W. Magnani. 2024. Exploring the Potential of Virtual Agents in Atrial Fibrillation Management: Insights from a Randomized Trial. In *ACM International Conference on Intelligent Virtual Agents (IVA '24)*, September 16–19, 2024, GLASGOW, United Kingdom. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3652988.3673955>

1 INTRODUCTION

Chronic diseases pose a significant public health challenge worldwide, affecting millions of individuals and placing a substantial burden on healthcare systems. In the US, six in ten adults have a chronic condition, and 90% of the \$4.1 trillion in annual healthcare costs are for people with chronic health conditions. [9] Among chronic conditions, atrial fibrillation (AF) stands out due to its complexity and prevalence. AF is a common arrhythmia characterized by an irregular and often rapid heart rate, leading to an increased risk of stroke, heart failure, and other cardiovascular complications. The prevalence of AF in the US is estimated to increase from approximately 5.2 million in 2010 to 12.1 million in 2030, reflecting a significant rise over the years as the population ages [11].

The management of AF is multifaceted, encompassing lifestyle modifications, medication adherence, and regular monitoring of heart rhythm [24]. Anticoagulants and medications are commonly prescribed to reduce the risk of stroke and control heart rate and rhythm, respectively. However, the effectiveness of these treatments is heavily dependent on patient adherence to medication regimens and consistent monitoring of heart rhythm and symptoms [36]. Nonadherence to prescribed medications is a prevalent issue in AF management, leading to suboptimal outcomes and increased healthcare costs [15]. In addition, the irregular nature of AF episodes necessitates regular assessments to detect the arrhythmia promptly [37]. Traditional methods of monitoring, such as periodic in-clinic electrocardiograms (ECGs) may not capture intermittent episodes, highlighting the need for solutions that enable more continuous monitoring and real-time detection of AF episodes.

We describe a virtual-agent based automated system for helping patients with AF manage their condition. Patients are provided with a smartphone and a portable ECG sensor that patients can use to take readings at any time. The virtual agent AF coach on the smartphone instructs patients about AF and the importance of medication adherence and taking frequent ECG readings, and allows them to report medication taking and AF-related symptoms. While there are numerous apps designed for the management of chronic diseases, discontinuation is a universal problem with all voluntary-use health applications, and continued use is a prerequisite for patients to obtain benefit from these applications [39]. To

address this challenge, the agent uses a variety of techniques, such as social chat and serial storytelling, to promote engagement with patients. Virtual agents utilize these techniques to establish rapport and trust with users, achieving both task-oriented and interpersonal objectives, showcasing the agent's expertise [4].

Information from these conversations with the agent is posted to a workstation monitored by clinicians, in addition to ECG readings, so that the patient can be contacted if problems arise (Figure 1).

We report the preliminary results from a randomized clinical trial of this system, comparing AF patients who use it for four months (AGENT group) to patients who are provided with the same smartphone and ECG sensor, but with a medical app installed on their smartphones instead of the virtual agent (CONTROL group). We hypothesize:

- **H1:** Patients in the AGENT group will use the virtual agent regularly throughout the four months of the intervention.
- **H2:** Patients in the AGENT group who interact with the engagement content (storytelling) will have more interactions with the agent overall.
- **H3:** Patients in the AGENT group will take more ECG readings compared to those in the CONTROL group.
- **H4:** Patients in the AGENT group will rate the system significantly higher on key attributes such as satisfaction, trustworthiness, likability and perceived knowledge, compared to how patients in the CONTROL group rate their system.
- **H5:** Patients in the AGENT group will report having a significantly greater quality of life compared to those in the CONTROL group, based on a validated AF-related quality of life measure (AFEQT) [35].

This virtual agent is unique in that it is designed to support patients for daily use over an extended period of time and promotes the use of a heart rhythm sensor important for management of a chronic disease. It is evaluated with a large sample of actual atrial fibrillation patients with for four months.

2 RELATED WORK

2.1 Mobile Interventions for Chronic Disease Management

Management of many chronic diseases requires the regular measurement of physiological metrics [17]. Monitoring devices such as biometric sensors in smartphones, smartwatches, and other mobile devices play an essential role in tracking and recording vital physiological data. These include tools like smartphone apps for tracking blood sugar levels and wearable devices that monitor heart rhythms [27, 29]. Pamungkas et al. investigated the effects of a mobile app-based diabetes coaching intervention on self-management behaviors and clinical outcomes in individuals with type 2 diabetes [31]. Results revealed improvements in dietary control, exercise, blood glucose monitoring, medication adherence, and complication screening. Mamykina et al. conducted a qualitative study on a mobile application, MoDD, that facilitates reflection and problem-solving in diabetes self-management [25]. The study revealed that MoDD facilitated problem-solving by guiding participants through the steps of the problem-solving process, including identifying problematic blood glucose patterns, exploring behavioral triggers

contributing to these patterns, selecting alternative behaviors, and implementing these behaviors while monitoring improvements in glycemic control.

Several mobile technologies have been developed for the detection and management of AF [14]. Brandes et al. delved into the complexities of consumer-led screening for AF [8]. They noted that users of single-lead ECG devices generally express high satisfaction. The study emphasizes the significance of offering training and clear guidance to empower patients in effectively using these tools. Ding et al. examined older patients' perceptions regarding using smart devices for heart rhythm monitoring, finding that the majority of participants (78%) considered the devices user-friendly and rated them as more convenient than traditional cardiovascular monitors [13].

2.2 Mobile Virtual Agents

Jaber et al., in their survey on conversational user interfaces on mobile devices, emphasized the effectiveness of animated agents in delivering nonverbal cues and enhancing engagement with users [18]. They highlighted that animated agents can provide nonverbal expressions like facial expressions, head nods, and eye gazes, contributing to more engaging social interactions. Bickmore et al. through a design study, described how virtual agents on a handheld device successfully build social connections and convey health information effectively [5]. In a similar study, Kang et al. explored the impact of an animated virtual agent in a mobile chat interaction [20]. They found that users engaged in long conversations with a 3D virtual agent compared to a static image or an audio-only interface.

2.3 Virtual Agent Health Interventions

Virtual Agents have been developed to counsel patients on health problems and chronic disease self-care across diverse populations and conditions, including mental health, diabetes, heart failure, and HIV/AIDS [7, 16, 34]. A recent scoping review highlighted the increasing use of Virtual Agents for chronic disease management, but only some have been deployed on mobile devices [19]. For instance, Martínez-Miranda et al. discuss developing and evaluating a mobile-based agent that monitors and supports individuals with mental health problems remotely, offering cognitive behavior therapy-based activities, and detecting factors associated with suicidality risk [26]. Kowatsch et al. described a Virtual Agent designed as a mediating social actor involving health care professionals, patients, and family members in chronic disease management [23], with users accepting the agent as a member of their healthcare team.

2.4 Automated Interventions for Atrial Fibrillation

The literature on enhancing adherence and self-management skills in AF through automated interventions is limited. One intervention, known as "Health Buddies", aimed to improve medication adherence among older adults with AF [12]. This intervention involved an agreement where patients and their grandchildren engaged in daily challenges. Users rated the system highly on acceptance and

usability, although only 13% of eligible patients agreed to participate due to a lack of interest or discomfort with technology.

Laranjo et al. evaluated the use of conversational artificial intelligence intervention to support patients with AF [38]. This intervention used automated telephone calls, texts, emails, and an education website. The digital health intervention effectively addressed gaps in post-hospital discharge AF care by improving patients' knowledge of AF, aiding in lifestyle adjustments, and supporting self-management. However, participants noted several barriers to engagement, including limited conversational flexibility, inadequate personalization of content and communication, and the need for continuous intervention adaptation to accommodate the episodic nature of AF symptoms.

3 ATRIAL FIBRILLATION COACH DESIGN

Our virtual AF coach system is designed to help patients by providing both education about AF and promotion of long-term treatment regimen adherence. The system consists of the Virtual Agent app and a heart rhythm sensor (AliveCor Kardia ECG) on users' smartphones, a central database, and a clinician workstation used to remotely monitor patient status (Figure 1).

Individuals with AF often require frequent heart rhythm monitoring. Traditional management of AF relies on regular hospital visits for electrocardiogram (ECG) monitoring, posing challenges for patients and healthcare systems. The AliveCor Kardia ECG monitor is small and mobile and transmits heart rhythm data to user smartphones via Bluetooth. The Kardia ECG consists of two sensors that users place their fingers on to take a reading. Data is uploaded to the AliveCor server which classifies the reading as indicative of atrial fibrillation or not, and also provides current heart rate. This information is provided back to the user and to our clinician workstation (Figure 1, right).

Our Virtual Agent is a 3D-animated character that simulates face-to-face dialogue with users (Figure 1, left), displaying a range of conversational nonverbal behavior, including head nods, facial displays, and hand gestures that are generated using BEAT [10]. Conversations between users and the agent are scripted using a hierarchical transition network-based scripting language with template-based text generation. User interaction with the agent primarily involves selecting utterances from a multiple-choice menu, updated at each turn of the conversation. Dialogue data is persistently stored locally on the smartphone and synchronized with the central database at the start and end of each conversation, enabling multi-session interactions and remote clinical monitoring.

The agent provides AF education, including potential causes and associated complications. Patients are informed about current treatment strategies and their effectiveness. They also learn about self-monitoring skills and the importance of self-care in AF management.

The agent also talks about medications, their side-effects, and the importance of adhering to prescribed anticoagulant medications for stroke prevention. Common challenges to medication adherence are discussed, and the agent shares recommendations for managing these challenges. The agent also explains common side effects from anticoagulant medications, and encourages users to be active participants in their care by reporting AF symptoms and medication

side effects to the agent. Users can report that they took their AF medication either to the agent or via a graphical user interface (GUI) in the app.

The agent educates users on the Kardia ECG, to provide the knowledge and skills to use the Kardia device properly. The ECG education facilitates the differentiation of a heart rhythm in AF from a regular heart rhythm. Importantly, the agent promotes adherence to consistent heart rhythm monitoring by emphasizing the clinical importance of regular readings. The agent also provides support to address potential barriers to using the ECG, and motivates patients to set a goal to use the ECG every day, following up on their progress after two weeks. The agent encourages users to use the ECG whenever three days have passed since the last reading. The agent acknowledges regular ECG usage by congratulating users after five consecutive days of taking a reading in which no other ECG concerns are found.

The agent informs users when either a high heart rate, AF, or other adverse event was detected in an ECG reading they have taken. Upon receiving this information, the agent leads users through a symptom assessment where the symptom they were experiencing at the time of the adverse event is logged in the database for clinicians to review, and then suggests that users inform their clinician. Following this, users have the opportunity to discuss their symptoms in more detail with the agent, log the severity of their symptoms, and are ultimately encouraged to take a ECG reading and contact their health provider.

After the second week of the study and once the agent has delivered all education material, it helps users set goals for healthy activities related to AF. These goals include using the ECG every day, taking anticoagulant medication as prescribed, exercising, keeping a symptom diary, and preparing for a doctor's visit. The agent checks in on users' progress two weeks after each goal is set and updates its status. If a user did not achieve their goal, the agent uses empathic language to encourage them to try again. If they met their goal they are congratulated and asked if they would like to continue pursuing the goal for another two weeks.

The agent also engages in episodic storytelling about her life throughout the study to promote engagement with users. A dialog module was developed that allows the agent to say one episode of an ongoing story per day, such as interactions with her neighbors, her friends, and their lives. The events of these story episodes were inspired by content from an existing long-running television soap opera. Plot summaries of hundreds of episodes were web-scraped from the a fan-wiki page containing plot summaries of all episodes¹. A script writer was hired to write short versions of 119 of these episodes and ensure they transition well from one day to the next, preserving the main characters and plot threads from one installment to the next. This module is activated when the user chooses the option to chat with the agent from the main menu of possible daily interactions and they can only experience one episode per day.

This virtual agent system goes significantly beyond prior work [6] by supporting longer term interactions (4 months of daily use), using storytelling and other mechanisms to promote sustained

engagement, enabling users to report medication taking and symptoms to either the agent or via GUI widgets in the app, and explicitly motivating AF management behaviors—such as taking regular ECG readings—through goal-setting dialog.

4 RANDOMIZED CLINICAL TRIAL

We conducted a randomized clinical trial to evaluate the AF management system described above over four months of use.

We examined the impact of the AF virtual coach promoting the use of ECG on the actual number of ECG readings performed by participants. Participants were assigned to either the AGENT group, which was provided with the AF virtual coach in addition to the ECG, or a CONTROL group, which received the ECG along with a medical information app (WebMD) installed on their smartphone.

4.1 Participants

Participants with AF were recruited from socioeconomically disadvantaged areas in the Pittsburgh, PA, metropolitan region. The inclusion criteria comprised adults aged 21 and above with a confirmed AF diagnosis who were prescribed either warfarin or Direct Oral Anticoagulant for stroke prevention, and proficient in English communication. Exclusion criteria included previous electrophysiologic interventions for AF treatment (as such individuals are likely to have high engagement with specialty care for AF), having non-AF conditions requiring anticoagulation, cardiac surgery or hospitalization within 3 months of enrollment, and other health issues that could affect participation.

Participants were recruited via several methods, including screening electronic health records and sending direct mail to potentially eligible patients. We also used alerts in clinical settings to notify health providers when a participant might be eligible for the study. We distributed patient-centered brochures and posters in clinic areas and waiting rooms, and also used a study website and a study Facebook page accessible to potential participants. After initial contact with potential participants, our research assistants scheduled online visits for screening and enrollment.

4.2 Measures

At enrollment, we collected demographics, medical history, experience with AF, current medications, and socioeconomic details including employment status, marital status, income level, and education level.

During the four-month intervention period, we tracked the usage of the AF virtual coach system and the ECG. This data includes usage timestamps (date and time of use), agent dialogue topics used by participants, the use of stories told by the agent to maintain participant engagement, various AF-related goals set by participants, and symptoms and medication adherence reported to the AF coach.

ECG data were uploaded to the AliveCor server, including the time and date of each ECG reading, the classification of the reading as indicative of atrial fibrillation or not, and user heart rate.

We measured AF-specific health-related quality of life using the AF Effect on Quality of Life (AFEQT) questionnaire [35], administered at enrollment and again at the end of four months. We evaluated the sustainability of the intervention's effect on AFEQT

¹https://coronationstreet.fandom.com/wiki/Category:Episodes_by_year

through longer-term follow-ups at 8 and 12 months after randomization.

Health literacy is assessed with both the Newest Vital Sign (NVS). Those scoring below 4 on the NVS indicate limited health literacy, while 4 or higher signifies adequate health literacy [42].

Participant attitudes towards either the AF coach or WebMD were evaluated using several single-item ten-point Likert scale measures, including satisfaction with the system, their level of trust in the information provided by the system, their overall liking of the system, perceived knowledge level of the system, perceived value of the system, and how well the system's information addressed their needs (Table 1).

4.3 Methods

Eligible participants were randomly assigned to either the AGENT or CONTROL groups. Participants attended an educational session focused on AF and Kardia, alongside receiving an informational brochure from the American Heart Association detailing AF and the importance of anticoagulation. Each participant was provided with a study smartphone that would be returned after the study. Those in the AGENT group were provided with the agent system described above. Participants in the CONTROL group were also given a smartphone with pre-installed applications for health-related information, WebMD, and Kardia apps.

Both the AGENT and CONTROL groups engaged with their assigned media for 4-months (120 days). At the end of the intervention period, outcome measures, including AFEQT and attitude towards the intervention media were collected, and a brief semi-structured interview conducted to probe patient experiences with the intervention. Follow-up interactions were conducted remotely due to the COVID-19 pandemic.

4.4 Results

In the following, we clustered data into 17 weeks to facilitate longitudinal analyses.

4.4.1 Participants and Sociodemographics. Between January 2020 and April 2022, we enrolled 240 participants; 121 were assigned to the AGENT group, and 119 were assigned to the CONTROL group. Participants were predominantly older adults, with a mean age of 70.73 (SD = 9.8), and majority female (64.0%). The racial composition was primarily Non-Hispanic White (66.3%), followed by Black (30.5%), Asian (0.4%), and Multiple/Other (2.1%). Most participants were retired (73.7%), with a smaller portion employed (16.0%). Nearly half were married or living as married (49.8%), while a significant portion were widowed (18.5%). In terms of education, 48.1% held an associate degree or higher, 22% had a high school diploma or less, 17.3% had some college experience but no degree, and 7.4% had attended a vocational or trade school.

4.4.2 Longitudinal Agent Use. AGENT group participants held an average of 76.7 (SD = 41.9) sessions with the AF virtual coach (range 1 to 176), a median of 91 sessions, and IQR = 63.

The correlation between participants' ages and the number of agent sessions was positive but not significant, Pearson $r(118) = 0.13$, $p = 0.15$. There were also no significant differences in agent system use between those with adequate and inadequate health

literacy, Wilcoxon rank-sum $W=166$, $p=0.66$. This suggests that older participants interacted with the AF agent as effectively as younger ones, and participants with low health literacy showed the same level of interaction as those with high health literacy. Participants used the agent intervention for a duration ranging from 1 to 17 weeks, with an average active period of 14 weeks (SD = 4.9). The median active period is 17 weeks, indicating that half of the participants maintained engagement with the AF agent throughout the entire study duration.

We investigated discontinued use of the virtual agent by defining an inactive user as someone who ceases interaction with the agent before the end of the intervention period and does not resume any sessions afterward. Out of the 121 participants, 28 (23.1%) were identified as inactive. One of these inactive users passed away. We did not find any correlations between participant characteristics (age, gender, education level) and whether they discontinued use or not. The analysis conducted for active users revealed a positive correlation between the sessions and time (Week), as indicated by Spearman's rank correlation coefficient ($\rho = 0.12$) with a significant p -value of less than 0.001 (Figure 2).

The number of participant self-reported medication adherence reports ranged from 0 to 228 during the 120-day study, indicating that some participants reported multiple times daily, with a mean of 78.7 (SD = 41.9). Participants could report that they either took their medication as prescribed ('Yes'), did not take it ('No'), or would take it at a later time ('Later'). The number of 'Yes' responses ranged from 0 to 209, with a mean of 73.1 (SD = 39), 'Later' responses ranged from 0 to 112, with a mean of 5 (SD = 13.6), and 'No' responses had a mean of 0.4 (SD = 1.0). Medication taking was primarily reported via the agent (number of reports range: 0-228, M: 65.4), but also through graphical user interface widgets (range: 0-114, M: 14.9). The Wilcoxon rank-sum test revealed a significant difference between reporting through these different modalities ($W = 11782$, $p < 0.001$).

Participants reported their symptoms to the agent between 1 and 128 times during the intervention, mean 42.4 (SD = 38.0). The analysis of participant-reported symptoms during the intervention revealed diverse experiences. Fast Heartbeat had an average occurrence of 2.7 times (SD = 4.4), Palpitation was reported around 3.8 times (SD = 8.9), and Shortness of breath occurred with a mean of 3.2 times (SD = 7.7). Symptoms like Chest Pain and Lack of appetite were less prevalent, with mean occurrences of 0.3 (SD = 0.6) and 0.06 (SD = 0.3) times, respectively. Anxiety and Sadness were reported relatively infrequently, with mean frequencies of 0.3 (SD = 1.0) and 0.1 (SD = 0.6) times, respectively.

4.4.3 Impact of Engagement Mechanisms on Agent Use. There was significant variation in the number of times participants listened to the AF coach's stories, ranging from 0 to 182 times, with an average frequency of 23.6 times (SD = 40.1) and a median of 4 times over the 17 weeks of the intervention. In order to determine whether story listening drove engagement, we analyzed the number of stories heard during the first eight weeks of the intervention, and determined whether these predicted the number of agent sessions in the remaining nine weeks. Participants listened to an average 2.4 stories during first eight weeks (SD = 7.9, ranging from 0 to 55), with the median, Q1, and Q3 all at 0. Based on the median, participants were split into two groups, those who heard any stories during the

Question	Anchor 1	Anchor 10	AGENT	CONTROL	p
How satisfied are you with Tanya/WebMD?	Not at all	Very satisfied	6.7 (3.1)	6.7 (2.8)	0.783
How much do you trust information from Tanya/WebMD?	Not at all	Very much	8.2 (2.4)	7.5 (2.5)	0.004
How much do you like Tanya/WebMD?	Not at all	Very much	7.1 (3.2)	6.4 (2.8)	0.014
How knowledgeable is Tanya/WebMD?	Not at all	Very	8.1 (2.6)	7.4 (2.6)	0.01
How valuable is Tanya/WebMD?	Not at all	Very valuable	7.0 (2.9)	6.4 (2.9)	0.103
Did Tanya/WebMD provide information that addressed your needs?	Not at all	Very much	6.9 (2.9)	6.5 (2.9)	0.302
How much would you like to continue working with Tanya/WebMD?	Not at all	Very much	5.7 (2.9)	5.9 (3.6)	0.898
How natural were your conversations with the agent?	Very unnatural	Very natural	6.5 (3.2)	-	-
How would you characterize your relationship with the agent?	Stranger	Close friend	6.5 (3.0)	-	-
How much would you like Tanya to be part of your healthcare team?	Dislike	Like	5.7 (3.3)	-	-

Table 1: Self-Report Attitudes towards Agent and WebMD. Wilcoxon significance tests on single items.

first eight weeks and those who did not. Participants who heard stories during the first eight weeks had a higher median of 64 sessions in the second nine weeks, compared to participants who did not hear any stories, who had a median of 44 sessions. There was a trending difference between these groups on subsequent sessions, Wilcoxon rank sum, ($W = 1379.5$, $p = 0.06$).

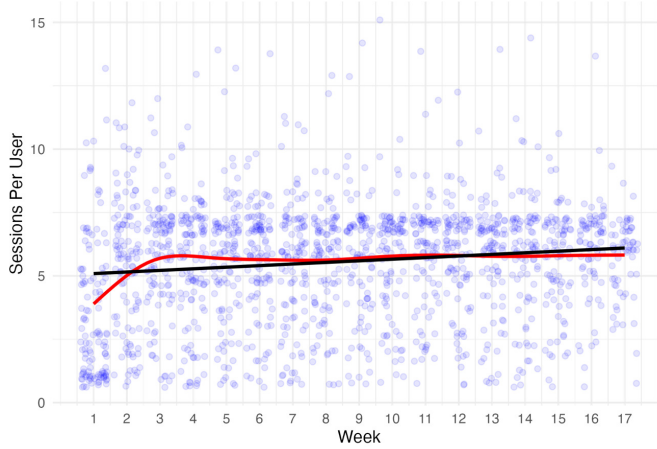


Figure 2: Sessions per User Over Time: Red line indicates actual weekly averages including initial week. Black line is a linear fit, excluding initial week.

4.4.4 Heart Rhythm Sensor Readings. Results from the ECG classification showed that participants had an average of 26.5% (SD = 36.0%) of their readings classified as being in AF.

A longitudinal linear mixed model analysis of the data showed there was a significant negative effect of time (weeks) on the number of ECG readings (-0.2 , SE = 0.03, $t(238) = -5.9$, $p < 0.001$), suggesting a significant decrease in the number of ECG readings over the 17 week intervention for both the CONTROL and the AGENT groups. Study condition had a significant main effect on the number of readings (1.5 , SE = 0.6, $t(238) = 2.2$, $p = 0.03$), indicating that participants in

the AGENT group had, on average, 1.5 more ECG readings over the entire 17 weeks compared to those in the CONTROL group (Figure 3).

We found a significant positive correlation between the number of agent sessions a participant in the AGENT group started and the number of ECG readings they took over during the intervention period, Spearman $\rho = 0.24$, $p < 0.001$. There was also a significant positive correlation between the number of goals a participant endorsed with the agent for “using the Kardia every day” and the number of ECG readings they took, Pearson $r = 0.32$, $p < 0.001$.

4.4.5 AF Quality of Life. We compared changes across study groups in AFEQT between the beginning and end of the intervention period. The baseline mean AFEQT global score for the AGENT group was 73.2 (SD = 18.9), slightly higher than that of the WebMD group, which was 73.0 (SD = 20.1). By the 4-month mark, both groups saw increases, with AGENT at 74.8 (SD = 20.4) and WebMD at 74.7 (SD = 20.6). However, the mean difference between the two groups was not significant ($p = 0.92$). At 8 months, the mean AFEQT scores continued to show slight increases: AGENT group $M = 77.3$ (SD = 17.7), WebMD group $M = 75.6$ (SD = 19.1). By month 12, the scores were AGENT group $M = 76.5$ (SD = 17.6) and WebMD group $M = 76.0$ (SD = 20.7), with no significant difference between them at this time point ($p = 0.87$).

4.4.6 Self-Report Ratings of Attitudes Towards Media. The mean and standard deviations of the media satisfaction ratings are in Table 1.

We Used Wilcoxon rank-sum tests to compare various aspects between the AGENT and CONTROL groups, including satisfaction, trust in the information, overall system liking, perceived knowledge level, perceived value of the system, and the system’s effectiveness in addressing users’ needs. The results indicated that the AGENT group exhibited significantly higher trust scores ($W = 4543$, $p = 0.004$), a greater liking of the system ($W = 4706.5$, $p = 0.014$), and perceived a higher level of knowledge about the system ($W = 4669.5$, $p = 0.010$) compared to the CONTROL group. However, there were no significant differences between the two groups in terms of satisfaction with the system, perceived value of the system, fulfillment of information needs, and intention to continue using media.

Spearman's non-parametric correlation analyses revealed several significant positive correlations between various attitudes towards the agent and the number of sessions started in the AGENT group. There were correlations between sessions and: satisfaction ($\rho = 0.35, p < 0.001$), trust in the agent ($\rho = 0.30, p < 0.001$), overall liking of the agent ($\rho = 0.32, p < 0.001$), how knowledgeable participants felt the agent was ($\rho = 0.27, p = 0.004$), how valuable participants felt the agent was, ($\rho = 0.25, p = 0.007$) and the degree they felt the agent provided information ($\rho = 0.25, p = 0.008$).

Non-parametric Wilcoxon Signed Rank tests showed that participants' perceptions of the virtual agent's naturalness were significantly higher than the neutral value of 5 on these scales ($Z = 0.81, p < 0.001$). Additionally, ratings for perceived relational closeness with the agent ($Z = -0.85, p < 0.001$) and the desire for the agent to be part of their healthcare team ($Z = -1.13, p = 0.015$) were significantly different from neutral.

4.4.7 Qualitative Results. Participants in the AGENT group described the AF agent, Tanya, as **useful** resource: "[I] learned so much about health and found it to be so useful." [P6119]. They also highlighted that Tanya was **informative** "She had great information about Afib" [P6051]. Moreover, participants mentioned Tanya's **friendly** nature and noted that "She was friendly and told stories" [P6153] and expressed **enjoyment** in interacting with the agent, "I got used to talking to her, and I enjoyed speaking with her" [P6009].

Participants also mentioned that the agent was able to **tell stories** during interactions: "[I] Enjoyed being able to speak to her and listen to her stories" [P6050]. Few described the agent as a real friend, saying, "She almost seems real, like a friend." [P6068] and some participants reported that "Tanya was someone to talk to about Afib" [P6198].

Based on the interview results, 84 out of 111 participants mentioned that they used the application to report their medication adherence and appreciated the **medication reminder** features. They noted that "She is helpful to remind you to take your medication and pay more attention to your health" [P6051]. Participants also mentioned that Tanya **can answer questions** "she gave good explanations to questions" [P6002]. However, some of the participants mentioned that they were not able to ask their own questions "I couldn't ask questions of my own." [P6066]. Some of the participants also mentioned that the agent was **repetitive** "At the beginning it was interesting but then it got repetitious and annoying" [P6094].

Most of the participants reported that they preferred to interact with the agents in the morning, the majority of them interacted with Tanya at home, and only two participants reported that they used it at work. Five participants used other health management applications, and they found Tanya different from these applications: "She is more personal; she goes through a series of questions that the other applications don't do." [P6196]. A small number of participants reported using the application to record questions for their doctor.

4.5 Discussion

Participants in the AGENT group interacted with the virtual agent an average of 76.7 times over the duration of the intervention, with only 23.1% discontinuing use before the end of the 120 days, providing support for **H1**. This underscores the potential of virtual

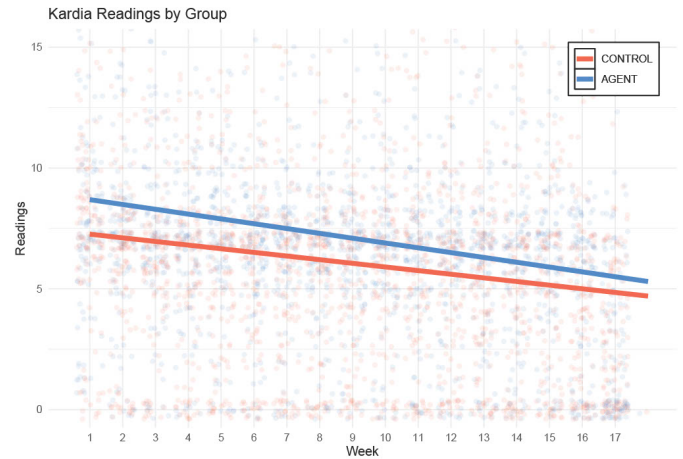


Figure 3: ECG Readings per Week: The AGENT group had a consistently higher number of Kardia readings than the Control group. The plot depicts a significant decrease in readings over time for both groups.

agents in long-term care, particularly in chronic disease management. Research has consistently shown that chronic conditions require ongoing support and guidance [1]. Despite challenges faced by many e-health applications that see a decline in usage after a few weeks [44], our results provide evidence that participants can sustain long-term engagement with the AF virtual coach.

Previous studies have identified correlations between specific demographic factors and higher attrition rates in digital health interventions, including male gender [22, 33, 44], younger age [21, 33, 40], lower socioeconomic status [33], and lack of a partner [32]. However, our results indicate that the AF virtual coach design effectively interacts with individuals across various demographics. Older participants interacted with the AF agent as effectively as younger ones, and there was no difference in interaction level between participants with low health literacy and those with high health literacy.

Participants who actively engaged in storytelling with the agent during the initial weeks of the intervention conducted a higher number of sessions in the second half of the intervention period. Furthermore, several participants highlighted storytelling as one of the most appealing features of the AF coach, providing some evidence for **H2**. This demonstrates a correlation between storytelling and participants' sustained engagement in interactions with the AF coach. This result aligns with the findings of previous studies, suggesting that conversational agents can use storytelling techniques to maintain user engagement within a single interaction and across multiple interactions [3].

Patients in the AGENT group took significantly more ECG readings compared to those in the CONTROL group, supporting **H3**. While participants in both groups received training on using the Kardia ECG sensor, the agent encouraged participants to take readings every day and to set goals related to ECG use, resulting in higher use. We also found a significant positive association between the number of goals set for using the ECG daily and the

number of ECG readings among participants in the AGENT group. Behavioral science and HCI researchers have described goal setting as a highly motivational tool that increases user engagement in longitudinal interventions [43].

Patients in the AGENT group reported significantly more medication-taking events via conversation with the agent than via graphical user interface widgets. This result is in contrast to a previous study in which user interface widgets were preferred over a conversational agent for transactional health interactions, such as reporting medication-taking or symptoms [30]. One reason for this could be that patients in our study initiated interactions with the agent for a range of reasons and medication reporting was embedded in a larger conversation, whereas in the earlier study health event reporting was the only activity performed with the agent.

Participants in the AGENT group rated the virtual coach significantly higher on trust, liking, and how knowledgeable the information source was compared to ratings of WebMD by CONTROL group participants. However, there were no significant differences on satisfaction ratings, providing only partial support for H4. Positive correlations were observed between session frequency and participant attitudes which suggested that increased user interaction with the AF coach correlates with enhanced positive experiences. Participants found the virtual agent to be natural, felt a sense of closeness to the agent, and desired the agent to be part of their healthcare team.

There were no significant differences between AGENT and CONTROL group ratings on AF-related Quality of Life (AFEQT), at either the end of the intervention at 4 months or at 8 month follow up. Thus H5 was not supported. The slight increases in AFEQT scores over the intervention period observed in both groups suggest that the ECG sensor coupled with educational content on AF—from any media—had a small positive impact on the participant’s ability to manage their chronic conditions. However, use of the virtual coach and increased frequency of ECG readings did not translate into improved quality of life, as was found in earlier studies [6].

5 CONCLUSION

We demonstrated that patients with atrial fibrillation, a serious and potentially debilitating chronic condition, can and will use a virtual coach agent on their smartphone for an extended period of time to help manage their disease. Patients across the age and health literacy spectra talked to the agent regularly for four months.

We found that serial storytelling by the agent—even though the stories were unrelated to the user’s task—served as an effective engagement mechanism, contributing to participants’ sustained engagement in interactions with the AF coach.

We found that participants trusted the AF-related information provided by the agent significantly more than the WebMD app, and that they liked the agent more than the WebMD app.

Finally, we demonstrated that participants used a mobile ECG sensor significantly more when a virtual coach motivated its use, compared to participants who were just provided with educational information about the device. The effectiveness of personal health devices in continuously monitoring health parameters is closely tied to how frequently they are used [41]. To the best of our knowledge,

our work is the first to develop and evaluate virtual agents that successfully promote the use of health devices.

Researchers have highlighted the role of conversational agents in facilitating tasks such as self-monitoring and overcoming obstacles for self-management, particularly in chronic disease management and public health crises such as pandemics [16, 28]. In our study, the AF virtual coach played an important role in assisting participants with various tasks specific to their chronic condition, including medication adherence and symptom reporting. The insights gathered from participant interviews revealed their preference for features related to medication reminders and symptom reporting within the AF virtual coach application.

5.1 Limitations

There are several important limitations of our study. Although all of our participants were diagnosed with AF, they were all recruited from one metropolitan area and our findings may not hold for all regions of the US or beyond. While four months is a long intervention by HCI standards, individuals with chronic conditions generally have to live the rest of their lives with their conditions (by definition). In this context, four months is a relatively brief period, and our results may not generalize to longer time periods.

5.2 Future Work

Although our agent negotiates health behavior goal setting with users, it does not follow up to see whether goals have been achieved or whether they need to be re-negotiated. Thus, longitudinal follow up on behavioral goals would be important to explore in future systems. This finding is supported by prior research, which suggests that simply setting goals in health behavior change interventions is not effective; instead, regularly checking progress and providing feedback are crucial aspects [2].

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Heart Lung and Blood Institute of the National Institutes of Health under award R33HL144669.

REFERENCES

- [1] Jordi Alonso, Montserrat Ferrer, Barbara Gandek, John E Ware, Neil K Aaronson, Paola Mosconi, Niels K Rasmussen, Monika Bullinger, Shunichi Fukuhara, Stein Kaasa, et al. 2004. Health-related quality of life associated with chronic conditions in eight countries: results from the International Quality of Life Assessment (IQOLA) Project. *Quality of life research* 13 (2004), 283–298.
- [2] Ryan R Bailey. 2019. Goal setting and action planning for health behavior change. *American journal of lifestyle medicine* 13, 6 (2019), 615–618.
- [3] Cristina Battaglino and Timothy Bickmore. 2015. Increasing the engagement of conversational agents through co-constructed storytelling. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 11. 9–15.
- [4] Timothy Bickmore and Justine Cassell. 1999. Small talk and conversational storytelling in embodied conversational interface agents. In *AAAI fall symposium on narrative intelligence*. 87–92.
- [5] Timothy Bickmore and Daniel Mauer. 2006. Modalities for building relationships with handheld computer agents. In *CHI’06 extended abstracts on Human factors in computing systems*. 544–549.
- [6] Timothy W Bickmore, Everlyne Kimani, Ha Trinh, Alexandra Pusateri, Michael K Paasche-Orlow, and Jared W Magnani. 2018. Managing chronic conditions with a smartphone-based conversational virtual agent. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 119–124.
- [7] Abdullah Bin Sawad, Bhuva Narayan, Ahlam Alnefaie, Ashwaq Maqbool, Indira Mckie, Jemma Smith, Berkan Yuksel, Deepak Puthal, Mukesh Prasad, and

- A Baki Kocaballi. 2022. A systematic review on healthcare artificial intelligent conversational agents for chronic conditions. *Sensors* 22, 7 (2022), 2625.
- [8] Axel Brandes, Stavros Stavrakis, Ben Freedman, Sotiris Antoniou, Giuseppe Boriani, A John Camm, Clara K Chow, Eric Ding, Johan Engdahl, Michael M Gibson, et al. 2022. Consumer-led screening for atrial fibrillation: frontier review of the AF-SCREEN international collaboration. *Circulation* 146, 19 (2022), 1461–1474.
 - [9] C Buttorf, T Ruder, and M Bauman. 2017. *Multiple Chronic Conditions in the United States*. Report. Rand Corp. https://www.rand.org/content/dam/rand/pubs/tools/TL200/TL221/RAND_TL221.pdf
 - [10] Justine Cassell, Hannes Högni Vilhjálmsón, and Timothy Bickmore. 2001. Beat: the behavior expression animation toolkit. In *SIGGRAPH*. ACM.
 - [11] Susan Colilla, Ann Crow, William Petkun, Daniel E Singer, Teresa Simon, and Xianchen Liu. 2013. Estimates of current and future incidence and prevalence of atrial fibrillation in the US adult population. *The American journal of cardiology* 112, 8 (2013), 1142–1147.
 - [12] Lien Desteghe, Kiki Kluts, Johan Vijgen, Pieter Koopman, Dagmara Dilling-Boer, Joris Schurmans, Paul Dendale, Hein Heidebuchel, et al. 2017. The health buddies app as a novel tool to improve adherence and knowledge in atrial fibrillation patients: a pilot study. *JMIR mHealth and uHealth* 5, 7 (2017), e7420.
 - [13] Eric Ding, Dongqi Liu, Apurv Soni, Oluwaseun Adaramola, Dong Han, Syed Khairul Bashar, Yeonsik Noh, Ki H Chon, and David D McManus. 2017. Impressions of older patients with cardiovascular diseases to smart devices for heart rhythm monitoring. In *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 270–271.
 - [14] Eric Y Ding, Gregory M Marcus, and David D McManus. 2020. Emerging technologies for identifying atrial fibrillation. *Circulation research* 127, 1 (2020), 128–142.
 - [15] Stacie B Dusetzina, Robert J Besaw, Christine C Whitmore, T Joseph Mattingly, Anna D Sinaiko, Nancy L Keating, and Jordan Everson. 2023. Cost-related medication nonadherence and desire for medication cost information among adults aged 65 years and older in the US in 2022. *JAMA Network Open* 6, 5 (2023), e2314211–e2314211.
 - [16] Ashley C Griffin, Zhaopeng Xing, Saif Khairat, Yue Wang, Stacy Bailey, Jaime Arguello, and Arlene E Chung. 2020. Conversational agents for chronic disease self-management: a systematic review. In *AMIA Annual Symposium Proceedings*, Vol. 2020. American Medical Informatics Association, 504.
 - [17] Martine WJ Huygens, Ilse CS Swinkels, Judith D de Jong, Monique JWM Heijmans, Roland D Friele, Onno CP van Schayck, and Luc P de Witte. 2017. Self-monitoring of health data by patients with a chronic disease: does disease controllability matter? *BMC family practice* 18 (2017), 1–10.
 - [18] Razan Jaber and Donald McMillan. 2020. Conversational user interfaces on mobile devices: Survey. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–11.
 - [19] Zhili Jiang, Xiting Huang, Zhiqian Wang, Yang Liu, Lihua Huang, and Xiaolin Luo. 2024. Embodied Conversational Agents for Chronic Diseases: Scoping Review. *Journal of Medical Internet Research* 26 (2024), e47134.
 - [20] Sin-Hwa Kang, Andrew W Feng, Anton Leuski, Dan Casas, and Ari Shapiro. 2015. The effect of an animated virtual character on mobile chat interactions. In *Proceedings of the 3rd International Conference on Human-Agent Interaction*. 105–112.
 - [21] Saskia M Kelders, Julia EWC Van Gemert-Pijnen, Andrea Werkman, Nicol Nijland, and Erwin R Seydel. 2011. Effectiveness of a Web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. *Journal of medical Internet research* 13, 2 (2011), e1624.
 - [22] Leonie FM Kohl, Rik Crutzen, and Nanne K de Vries. 2013. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. *Journal of medical Internet research* 15, 7 (2013), e146.
 - [23] Tobias Kowatsch, Theresa Schachner, Samira Harperink, Filipe Barata, Ullrich Dittler, Grace Xiao, Catherine Stanger, Florian v Wangenheim, Elgar Fleisch, Helmut Oswald, et al. 2021. Conversational agents as mediating social actors in chronic disease management involving health care professionals, patients, and family members: multisite single-arm feasibility study. *Journal of medical Internet research* 23, 2 (2021), e25060.
 - [24] Dennis H Lau, Stanley Nattel, Jonathan M Kalman, and Prashanthan Sanders. 2017. Modifiable risk factors and atrial fibrillation. *Circulation* 136, 6 (2017), 583–596.
 - [25] Lena Mamykina, Elizabeth M Heitkemper, Arlene M Smaldone, Rita Kukafka, Heather Cole-Lewis, Patricia G Davidson, Elizabeth D Mynatt, Jonathan N Tobin, Andrea Cassells, Carrie Goodman, et al. 2016. Structured scaffolding for reflection and problem solving in diabetes self-management: qualitative study of mobile diabetes detective. *Journal of the American Medical Informatics Association* 23, 1 (2016), 129–136.
 - [26] Juan Martínez-Miranda, Ariadna Martínez, Roberto Ramos, Héctor Aguilar, Liliana Jiménez, Hodwar Arias, Giovanni Rosales, and Elizabeth Valencia. 2019. Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. *Journal of medical systems* 43, 8 (2019), 246.
 - [27] Graeme Mattison, Oliver Canfell, Doug Forrester, Chelsea Dobbins, Daniel Smith, Juha Töyräs, and Clair Sullivan. 2022. The influence of wearables on health care outcomes in chronic disease: systematic review. *Journal of Medical Internet Research* 24, 7 (2022), e36690.
 - [28] Adam S Miner, Liliana Laranjo, and A Baki Kocaballi. 2020. Chatbots in the fight against the COVID-19 pandemic. *NPJ digital medicine* 3, 1 (2020), 65.
 - [29] Vasileios Nittas, Chiara Zecca, Christian P Kamm, Jens Kuhle, Andrew Chan, and Viktor von Wyl. 2023. Digital health for chronic disease management: An exploratory method to investigating technology adoption potential. *Plos one* 18, 4 (2023), e0284477.
 - [30] S Olafsson, D Parmar, E Kimani, T O'Leary, and T Bickmore. 2021. 'More like a person than reading text in a machine': Predicting Choice of Embodied Agents over Conventional GUIs on Smartphones. In *CHI*.
 - [31] Rian Adi Pamungkas, Andi Mayasari Usman, Kaniththa Chamroonsawasdi, et al. 2022. A smartphone application of diabetes coaching intervention to prevent the onset of complications and to improve diabetes self-management: A randomized control trial. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 16, 7 (2022), 102537.
 - [32] Dominique A Reinwand, Rik Crutzen, Iman Elfeddali, Francine Schneider, Daniela Nadine Schulz, Eline Suzanne Smit, Nicola Esther Stanczyk, Huibert Tange, Viola Voncken-Brewster, Michel Jean Louis Walthouwer, et al. 2015. Impact of educational level on study attrition and evaluation of web-based computer-tailored interventions: results from seven randomized controlled trials. *Journal of medical Internet research* 17, 10 (2015), e228.
 - [33] Dominique A Reinwand, Daniela N Schulz, Rik Crutzen, Stef PJ Kremers, and Hein de Vries. 2015. Who follows eHealth interventions as recommended? A study of participants' personal characteristics from the experimental arm of a randomized controlled trial. *Journal of medical Internet research* 17, 5 (2015), e3932.
 - [34] Theresa Schachner, Roman Keller, and Florian v Wangenheim. 2020. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. *Journal of medical Internet research* 22, 9 (2020), e20701.
 - [35] John Spertus, Paul Dorian, Rosemary Bubien, Steve Lewis, Donna Godejohn, Matthew R Reynolds, Dhanunjaya R Lakkireddy, Alan P Wimmer, Anil Bhandari, and Caroline Burk. 2011. Development and validation of the Atrial Fibrillation Effect on Quality-of-Life (AFEQT) Questionnaire in patients with atrial fibrillation. *Circulation: Arrhythmia and Electrophysiology* 4, 1 (2011), 15–25.
 - [36] Stavros Stavrakis, Julie A Stoner, Joel Kardokus, Paul J Garabelli, Sunny S Po, and Ralph Lazzara. 2017. Intermittent vs. Continuous Anticoagulation therapy in patients with Atrial Fibrillation (iCARE-AF): a randomized pilot study. *Journal of Interventional Cardiac Electrophysiology* 48 (2017), 51–60.
 - [37] Megan Streur. 2019. Atrial fibrillation symptom perception. *The Journal for Nurse Practitioners* 15, 1 (2019), 60–64.
 - [38] R Trivedi, T Shaw, CK Chow, and L Laranjo. 2023. Conversational artificial intelligence intervention to support patients with atrial fibrillation: process evaluation of a randomised controlled trial. *European Heart Journal* 44, Supplement_2 (2023), ehad655–2995.
 - [39] Isaac Vaghefi, Bengisu Tulu, et al. 2019. The continued use of mobile health apps: insights from a longitudinal study. *JMIR mHealth and uHealth* 7, 8 (2019), e12983.
 - [40] Celien Van der Mispel, Louise Poppe, Geert Crombez, Maïté Verloigne, and Ilse De Bourdeaudhuij. 2017. A self-regulation-based eHealth intervention to promote a healthy lifestyle: investigating user and website characteristics related to attrition. *Journal of medical Internet research* 19, 7 (2017), e241.
 - [41] Sudip Vhaduri and Temiloluwa Prioleau. 2020. Adherence to personal health devices: A case study in diabetes management. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 62–72.
 - [42] Barry D Weiss, Mary Z Mays, William Martz, Kelley Merriam Castro, Darren A DeWalt, Michael P Pignone, Joy Mockbee, and Frank A Hale. 2005. Quick assessment of literacy in primary care: the newest vital sign. *The Annals of Family Medicine* 3, 6 (2005), 514–522.
 - [43] Anna Weston, Leanne Morrison, Lucy Yardley, Max Van Kleek, and Mark Weal. 2015. Measurements of engagement in mobile behavioural interventions? (2015).
 - [44] Lucy Yardley, Bonnie J Spring, Heleen Riper, Leanne G Morrison, David H Crane, Kristina Curtis, Gina C Merchant, Felix Naughton, and Ann Blandford. 2016. Understanding and promoting effective engagement with digital behavior change interventions. *American journal of preventive medicine* 51, 5 (2016), 833–842.