



## Review

# In Search of Embodied Conversational and Explainable Agents for Health Behaviour Change and Adherence

Amal Abdulrahman  and Deborah Richards \* 

Department of Computing, Faculty of Science and Engineering, Macquarie University,  
Sydney, NSW 2109, Australia; amal.abdulrahman@students.mq.edu.au

\* Correspondence: deborah.richards@mq.edu.au; Tel.: +61-2-9850-9567

**Abstract:** Conversational agents offer promise to provide an alternative to costly and scarce access to human health providers. Particularly in the context of adherence to treatment advice and health behavior change, they can provide an ongoing coaching role to motivate and keep the health consumer on track. Due to the recognized importance of face-to-face communication and establishment of a therapist-patient working alliance as the biggest single predictor of adherence, our review focuses on embodied conversational agents (ECAs) and their use in health and well-being interventions. The article also introduces ECAs who provide explanations of their recommendations, known as explainable agents (XAs), as a way to build trust and enhance the working alliance towards improved behavior change. Of particular promise, is work in which XAs are able to engage in conversation to learn about their user and personalize their recommendations based on their knowledge of the user and then tailor their explanations to the beliefs and goals of the user to increase relevancy and motivation and address possible barriers to increase intention to perform the healthy behavior.



**Citation:** Abdulrahman, A.; Richards, D. In Search of Embodied Conversational and Explainable Agents for Health Behaviour Change and Adherence. *Multimodal Technol. Interact.* **2021**, *5*, 56. <https://doi.org/10.3390/mti5090056>

Academic Editor: Cristina Portalés Ricart

Received: 18 August 2021  
Accepted: 15 September 2021  
Published: 18 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** health behaviour change; embodied conversational agents; explainable agents; working alliance

## 1. Introduction

Chronic diseases, including non-communicable diseases and mental health conditions, are the main causes of death around the world. Reports by the World Health Organization [1] continue to emphasise that the onset of most chronic diseases is linked to four preventable risk factors: inadequate physical activity, unhealthy diet, excessive alcohol consumption and tobacco use. While governments continue to put more human and financial resources into dealing with and preventing chronic diseases, these resources are likely to be ineffective if individuals continue to practice unhealthy behaviours, particularly those associated with the four risk factors above. In other words, individuals' adherence to the healthcare giver's recommendation/s is the cornerstone for building a successful intervention.

Non-adherence to the recommended actions by a healthcare giver could be unintentional or intentional behaviour. Unintentional non-adherence is usually a result of the patient's forgetfulness or misunderstanding, whereas intentional non-adherence is a result of the patient's choice not to take action as recommended [2]. Reminders, simplification and education have been found to be successful interventions to deal with unintentional non-adherence [3,4]. The reach of such interventions may be limited depending on whether they can address issues around low health literacy, which is one of the major factors that influence health outcomes [5]. However, intentional non-adherence is a more complex problem, and psychological interventions that target the patient's cognition and emotions are more successful in motivating the patients to sustain a behaviour for a longer time [6].

Technology-based interventions to encourage health behaviour change have been introduced for some time now, including web-based and mobile-based interventions [7,8]. Their main advantage is their availability whenever the user needs support and at their

pace [9–11]. The success of these interventions in improving adherence increases when they are built on the grounds of a psychological theory such as cognitive behavioural therapy [9] or when the intervention system includes persuasive elements such as dialogue support and social support [12]. Hence, conversational agents (CAs) built on top of appropriate theories are potential interventions in persuading the users to adhere to the treatment.

The therapist-patient relationship known as a working alliance (WA) or therapeutic alliance, is regarded as a good predictor of patient adherence [13] and fosters behaviour change [14–16]. According to [14], developing a WA between the therapist and patient is characterised by three elements: agreement on the *goal/s* of the treatment, the *task/s* required to reach those *goal/s* and a trusting *bond*. This positive relationship is not restricted to human-human contexts only, as there is evidence that humans can unconsciously develop an empathetic and personal relationship with computers as well, particularly when the computerised intervention systems are endowed with social/emotional cues [16].

CAs for interventions can be equipped with a wealth of features to simulate face-to-face therapy, particularly when a CA is an embodied CA (ECA), such as appearance, facial expressions, head nods, body and hand gestures, and voice. These features, when synched appropriately to signal therapist skills, can boost the user-agent WA and, consequently, the adherence [17–19]. How these features could be utilized successfully to build WA that eventually improves the users' adherence is still ongoing research. While some researchers focus on the emotional/empathic features to build rapport rather than a WA [20], others argue that development of WA and empathy in health are cognitive processes and effective communication to build mutual understanding through explanation is essential [21,22]. We refer to CAs that provide explanations as Explainable Agents (XAs).

This paper reviews how CAs have been used to change health behavior and improve adherence, focusing on whether ECAs build a WA towards improved health outcomes. The article is structured as follows. Section 2 provides essential background to introduce the concepts of Working Alliance (Section 2.1), Conversational Agents (Section 2.2) and Explainable Artificial Intelligence (AI) and Agents (Section 2.3). Section 3 reviews ECAs (Section 3.1) and XAs (Section 3.2) in health behaviour change and then evaluates in Section 3.3 the current ability of the presented ECA/XAs to support the essential elements of the WA. Section 4 concludes the paper with a summary of the current state of the art and future important considerations.

## 2. Background

### 2.1. Working Alliance

The term working alliance refers to the positive collaboration between the therapist and the patient. Mathieu, et al. [23] defined it as the motivation behind the patient's acceptance and subsequent adherence to the therapist's recommended treatment. A large body of research has reported that a positive WA correlates with a good treatment outcome in psychotherapy, pharmacotherapy and placebo-therapy [24–26]. Other terms that have been used to refer to this relationship include 'alliance', 'therapeutic alliance', 'therapeutic relationship' and 'helping alliance' [27]. While researchers have used these terms with different conceptualisations and measures, Bordin [14] provided the most integrative model to describe this relationship, which he termed WA.

Bordin [28] stated that the WA is 'what makes it possible for the patient to accept and follow treatment faithfully'. He introduced the three main components/dimensions of the WA: tasks, goals and bond. 'Tasks' refer to the behaviours or actions of the treatment plans that are required of the patient to achieve the treatment's goal. 'Goals' refer to the objectives or outcomes of the counselling/intervention. The 'bond' is the trust the patient forms with the therapist and their confidence in the appropriateness of the treatment's tasks in achieving the treatment's goals.

According to Bordin's model, the therapist-patient WA is a result of the collaborative work to develop the patient's sense of involvement and the central focus that eventually motivates the treatment adherence [29]. Building on this principle, to develop a WA, the

therapist should plan how to build mutual understanding with the patient and acquire their agreement on the treatment tasks and goals by meeting their needs, considering their perspective and promoting positivity throughout the communication.

Planning therapist-patient mutual understanding includes communicating the patient's intentions towards the treatment tasks [30]. According to Bratman [31], in the belief-desire-intentions theory, having a plan does not mean knowing the actions but having the intention to do the actions and replanning when required. Grosz and Kraus [32] further emphasised the importance of intentions in collaborative planning and introduced the concept of SharedPlans, which refers to the plan constructed by members of a team based on their intentions. Such a plan could be of two types: partial or complete. Partial plans are initial plans containing uncertainty owing to a lack of knowledge about other team members' intentions. Complete plans have agreed-upon goals and tasks, resulting from positive team member collaboration. The team members collaborate, sharing their beliefs, goals and intentions, so that a member, with a level of certainty, can build a plan according to their mutual understanding. While the concept of SharedPlans was originally introduced to model agent communication and collaboration in achieving common goals in a shared environment with other agents or human users, it still emphasises the importance of discussing team members' beliefs, goals and intentions in mutual planning, which is also a cornerstone in treatment planning [30]. In harmonious environments where the team members are humans only, all the members have the same cognitive structures/mental models. This is also true of teams of artificial agents designed using the same/similar cognitive architecture. Thus, members can easily understand each other as they use the same underlying decision-making and reasoning process [33]. However, in environments with heterogeneous mental models, effective collaboration and planning becomes more challenging especially when a human is in the loop due to the fact that humans' behaviours are unpredictable compared to programmed agents [34].

These concepts of shared planning and mutual understanding are closely connected with the use of dialogue for human-agent interaction and communication [35]. There is evidence that human-agent shared planning and collaborative decision-making can help a human-agent dyad develop a WA [19]. Several types of software have been introduced successfully to implement the idea of shared planning, such as Collagen [36], where the dialogue is designed as a state-based tree to negotiate the actions/subtasks to complete a plan. Later, a task-based framework called Dtask [37] was developed based on ANSI/CEA-2018, where each task is coupled with a dialogue segment to facilitate the human-agent communication to perform a particular task. This research was inspired by the idea of utilising an adaptive dialogue manager to facilitate user-agent shared planning and thus build mutual understanding that eventually promotes a WA and health-related behavioural change.

## 2.2. An Introduction to Conversational Agents

With the increasing use of mobile devices, voice-activated CAs built into mobile devices are being utilised to help to accomplish various tasks such as searching the web for goods and services or answering enquires about different aspects, including health-related issues [38]. Miner, et al. [39] evaluated four mobile-based CAs on their ability to recognise crises related to mental health (suicidal thoughts), violence (rape) and physical health (heart attack). The four CAs: Siri (Apple), S Voice (Samsung), Google Now and Cortana (Microsoft) showed none to very limited abilities to recommend helpful advice or connect the user to helplines. While one can argue these CAs were mainly designed as personal assistants to search the internet and help in general tasks, it would be helpful if designers could take advantage of their popularity to advance their capabilities to run longer conversations with users with health concerns or motivate them to adopt healthier behaviours. Further, the safety of such open input CAs is a serious issue that has not been properly evaluated yet [40]. Hodgson, et al. [41] reported a significant increase in the error rate for completing clinical documentation tasks using a speech recognition method

compared to the traditional method using a mouse and keyboard. This indicates that this technology is still not ready to be used in a critical application such as health.

Designing CAs requires a collective effort from different disciplines including the social, cognitive and computer sciences [42]. Combining all these knowledge domains and features in one CA is challenging, which suggests the applicability of designing domain-dependent CAs that address a particular problem in a particular context [43]. Several taxonomies have been proposed to guide the design of CAs [43–45]. For example, Isbister and Doyle [43] distinguished four categories of features: believability, sociability, domain application design and agent design. Another example is the taxonomy introduced by Bittner, et al. [44], who defined several design features for collaborative CAs, including the agent's role, the number of human users in the team, the collaboration goal, communication mode and the agent's socio-emotional behaviour. However, the impact of these design features on the user's perception and their effectiveness and applicability in the various domains have not been well researched and validated [46,47]. Focusing on CAs designed for the health domain, ter Stal, et al. [47] concluded that not all of the investigated design features have proven their effectiveness in achieving health-related goals and that agent relational/empathic behaviour is the most promising feature towards improving the user-agent relationship.

The earliest CA for healthcare can be dated to 1966 when Weizenbaum [48] proposed Eliza, a computer program that simulated the role of a Rogerian therapist and utilised natural language conversation with the user. Parry is another early chatbot designed by Colby [49], which simulates a paranoid patient to facilitate the study of paranoia by psychiatrists and psychologists. Both chatbots, Eliza and Parry, were text-based chatbots; however, Eliza analysed the input text to do pattern-matching with the available agent's responses and Parry used symbolic representation to understand the context and deliver a response according to the meaning.

CAs like Eliza and Parry are chatterbots able to keep chatting with a user for a long time due to their large databases of diverse responses. However, they were not designed to achieve a task, to plan, or to build a relationship with a human user. These agents are more useful in delivering information to the user. More recent examples of educational CAs are Bzz [50]—designed to answer adolescents' text-based typed open questions about sex, drugs and alcohol—and Harlie [51], which was designed as a speech-based agent that chats about general topics with patients with neurological conditions (e.g., autism or Parkinson's disease), who could have a problem in using a text-based CA.

An early attempt to include a theory-driven CA was introduced by Colby [52] to deliver cognitive behavioural therapy for depressed patients. More recently, Fitzpatrick, et al. [53] implemented the principles of cognitive behavioural therapy to build Webot as a text-based CA. Webot has been evaluated in a randomised control experiment with 70 university students aged between 18 and 28. Compared to the control group ( $N = 36$ ), who received an e-book with information about how to deal with depression, Woebot ( $N = 34$ ) significantly reduced the patients' depression. However, the relationship between it and the patients was not evaluated.

Grudin and Jacques [54] proposed a taxonomy of CAs based on their conversation focus identifying (a) *virtual companions* that can have a broad or deep focus, engage in 10 to 100 s of exchanges as in the examples of ELIZA, Cleverbot, Tay, XiaoIce, Zo, Hugging Face; (b) *intelligent assistants* with typically broad and shallow focus and one to 3 exchanges as we see in Siri, Cortana, Alexa, Google Assistant and Bixby; and (c) *task-focused CAs* that have narrow or shallow focus and 3–7 exchanges as in Dom the Domino Pizza bot, customer service bots, Russian trolls and non-player characters.

ECAs have been introduced as an extension to the text-based or speech-based CAs to include more conversational features and imitate natural face-to-face conversation. Although patients can form an interpersonal relationship with simple CAs such as a phone-based virtual health counsellor [16], the combination of verbal (speech) and non-verbal (e.g., eye gaze, gesture and facial expressions) conversation features facilitate effective,

meaningful communication and mutual understanding [55]. ECAs for health behavior fit best into a task-focused CA as defined by Grudin and Jacques [54]. However, what distinguishes them from customer service bot or trolls, which seek to provide (dis) information, is the development of a relationship, somewhat similar to a caring human relationship and WA.

Much effort is being devoted to increase the ECAs' acceptance by simulating the natural appearance and behaviour of a human being (i.e., anthropomorphism) [56]. To build believable ECAs towards fruitful interactions and establishing relationships, the design of ECAs requires more effort compared to CAs that are text- or speech-based only [45,57] to deliver more sophisticated capabilities, such as emotion modelling [58,59]. The implications of a leap from text-only and speech-only CAs to ECAs are higher user expectations as they unconsciously perceive ECAs as social entities and interact accordingly [60]. Further, users may prefer to interact with ECAs rather than humans in some contexts [61]. Simply adding a face to a CA may not be adequate and could distract and deviate the users from the main goal of the application. An early meta-analysis by Yee, et al. [62] concluded that the effect of adding a face to the CAs is small and subjective (users' rating) rather than objective (task performance), and adding a face is not as important as making the interaction believable.

A recent systematic review of design features for building trust with ECAs found that the efficiency of ECAs' design features largely depends on the application context and goal [63]. While researchers have striven to endow ECAs with more human features such as personality [64,65], emotions, and body and hand gestures [66,67], ECAs' facial appearance and spoken or textual speech are the most studied design features, which have been identified as the most effective features in building the human-agent relationship [47], such as the WA in the health domain.

### 2.3. Explainable AI & Explainable Agents

According to the *Oxford Dictionary*, explanation is 'A mutual declaration of the sense of spoken words, motives of actions, etc., with a view to adjusting a misunderstanding and reconciling differences. Hence: a mutual understanding or reconciliation of parties who have been at variance.' [68] (para. 3). Walton [22] states, 'the purpose of an explanation is for the agent to verbally transfer understanding to the other' (p. 1). Explanation can be seen as a shortcut to establishing trust through allowing the user to interrogate the reasoning and boundaries of the artificial systems knowledge [69]. Through conversation answering the *why* question and being relevant to the explainee Hilton [70], explanation becomes a social process that bridges the knowledge or inference gap between the explainer and the explainee's understanding [71].

The field of explainable AI (XAI) emerged to improve the human understanding of the machine decision-making process. It can be dated to the 1970s when researchers attempted to explain expert systems using logical rules [72,73]. More recently, XAI regained attention due to advancements in automated systems, especially in the field of machine learning, which is still a black box for the designers themselves [74]. The need for explainability further increased when automated systems showed some bias and fairness violations, leading to ethical concerns. For instance, gender recognition systems that are trained to identify gender based on physical appearance are perceived by vulnerable users (e.g., transgender) as inappropriate and disturbing [75]. In such a situation, XAI can help the designers to eliminate or consider more features to increase the fairness of the technology and limit the bias.

From the perspective of the end-user rather than the system designers, the growing interest in XAI stems from the current underutilisation of AI systems [76]. Human use of the advice provided by automated systems can include: misuse, an overreliance that fails to question the advice provided; disuse, ignoring or underutilisation of the guidance and advice provided; or abuse, use without concern for the consequences [77]. Dzindolet, et al. [78] found a correlation between automated system reliability and trust and that



explanation improves user trust, especially when unexpected behaviour occurs. Thus, the need for XAI stems from the need for trust in automation and the proper use of technology.

In explainable artificial intelligence (XAI), the explanation can be of two types: data-driven or goal-directed [79]. Data-driven explanation is the interpretation of machine-learning models, while goal-directed explanation, also called explainable agency, is the justification of the agent's actions according to its mental state and reasoning process. The majority of XAI research concerns explanations of the outputs of machine-learning, and thus is data-driven [80]. In goal-directed agency, the agent may use different explanation patterns according to the purpose such as to justify its action or for knowledge transfer. Different users may have specific requirements and expectations for the level of transparency and verification provided by an agent before they are willing to trust its recommendations and actions [69]. Explanation personalised to the user's context, preferences and beliefs will be more meaningful and persuasive [81].

There is a small but growing body of work in the area of explainable agents (XAs), particularly, goal-driven agents [79]. Intelligent agents in their various forms (i.e., text-based chatbots, ECAs, and physical robots) are being used in health [82], education [83], and marketing [84]. Explanation provides transparency towards building human-agent trust, which is important for their acceptance [69] and towards achieving the system goals (e.g., behaviour change). However, XAs are still in their developmental stage [79,85].

Theories and findings from the social sciences are commonly used as the basis for the design of explainable agents (XAs) that seek to simulate the ways in which humans use and provide explanation. Following Dennett [86], agents can simulate humans by providing three stances: physical, design and intentional, where the first two stances are related to the hardware and software used to construct the XA while the intentional stance is based on the agent's rational cognitive representation allowing the agent to reason about and explain its past, current and future actions [87]. Although this aspect could be seen as part of the design stance, changes in the agent's environment also impact the intentional stance as they require continuous updates to the agent's beliefs and desires.

Communication of an XA's reasoning is essential to creation of XAs that are believable and acceptable to users [85]. Believability hinges on people's perception of computers as social entities which they respond to similarly to the ways they respond socially to a human [88]. As part of the interaction, humans attribute the virtual agent with reasoning capabilities and expect that the agent can explain their reasoning like humans can [89]. During human-human social interactions, people seek explanations to reconcile knowledge discrepancies and build a common understanding (the main principle of a WA). Explanation in human-agent interaction serves a similar function [76].

Malle [90] introduced the folk explanation model, which is grounded on the theory of mind (ToM) (attributing mental state to oneself and others), to distinguish between intentional and unintentional behaviours. The model defines four modes of behaviour explanations: causes, reasons, causal history of reasons and enabling factors. On one hand, unintentional behaviours can only be explained by the causes (causal explanation), which include mental and physical states but without awareness or intentionality to do the behaviour. On the other hand, intentional behaviours can be explained by one of the other three modes. The causal history of reasons cites the causal background of the reasons such as one's traits, history and culture (beliefs and desires). The enabling factors refer to the factors that facilitate the behaviour to be done, such as the agent's skills or the opportunities provided by the environment. Finally, the reason explanation involves the agent's mental state (beliefs and desires) that the agent uses in the behaviour decision-making, which means the agent was fully aware and intended to do it. This is the only explanation mode characterised by rationality. The theory of mind does not only serve to explain the behaviour but also facilitates social functions such as coordination and understanding [91]. XAs can build this mutual understanding by utilising ToM reasoning in their explanation patterns [92].

An agent's behaviour is driven by the agent's belief and goals. For some decisions this reasoning process may be lengthy with a corresponding lengthy explanation that is likely to contain irrelevant content [76,93]. Understanding what content is relevant to include in a reason explanation using a BDI agent's beliefs and goals has been explored by Harbers, et al. [94] who used Goal Hierarchy Theory (GHT) to evaluate different explanation patterns including the agent's beliefs and goals behind conducting an action. A study with twenty novice firefighting trainers evaluated four types of explanations. Belief-based explanations were preferred by trainers for single action/goals, but goal-based explanations were preferred when the agent had to perform sequences of actions. Differences have been found in the explanation preferences of children and adults [95].

To persuade someone to change their behaviour, Wheeler, et al. [96] found that the explainer needs to deliver a message that provides a link between the recommendation and the explainee's beliefs. In the attentional systems theory [87], the behaviour of another (human beings or artefacts) are explained by their intentional stance: beliefs and desires. The intentionality to do the behaviour is volumized when adopted beliefs are perceived as useful. Malle [91] posits that a behaviour could be considered intentional only when there is a desire to achieve a goal through that particular behaviour. For example, a user may say 'I want to eat healthy (desire) because I think I will live longer (belief)'. The belief is that of the user, not the agent, and if the user does not adopt this belief, their intention may not change towards achieving the desired goal. From this point of view, we conclude that to increase a user's intentionality, explanation should include the beliefs of the user. Through alignment to the user's cognitive state, the motivation to perform the recommended action becomes internalised [97]. To this end, an ECA must store the user's mental state and connect its recommendations to it. This can be implemented using the ToM model, where an agent stores another's mental state (in what is usually called user model) besides its own and uses them in its decision-making and to produce an appropriate explanation.

### 3. Conversational and Explainable Agents for Health Behaviour Change

#### 3.1. Embodied Conversational Agents for Health Behaviour Change

Many ECAs have been introduced to address the problem of behaviour change. Some of these ECAs are designed to build the user-agent relationship through the use of empathy [98–100] and others focus on endowing the ECAs with therapeutic knowledge, following counselling strategies such as motivational interviewing [101–103].

Towards building an expressive ECA that mimics human verbal and nonverbal empathy, Greta was proposed by De Rosi, et al. [66]. Greta is a talking head endowed with personality, and social and emotional capabilities. Greta's face is designed with detailed facial animation parameters (FAPs) that convey its beliefs, intentions, affective state and metacognitive information. Greta has been evaluated in several scenarios. In a comparison of 3 different Gretas (neutral/non-expressive, consistent expressive or inconsistent expressive) to deliver a healthy eating message to 44 participants, users perceived neutral Greta as more convincing and its message as trustworthy. They were more likely to eat healthier after interacting with the neutral Greta, followed by the consistent expressive Greta and then the inconsistent one.

Another example of empathic agents for health applications is Simsensei [104], a virtual interviewer designed to engage the user through endowing the ECA, Ellie, with the appropriate verbal and non-verbal emotional behaviours to encourage patients to feel comfortable and disclose more information. Ellie is designed to detect the users' distress in patients with post-traumatic stress disorder, mainly from their non-verbal and verbal behaviours using natural language processing. Participants did not perceive Ellie's ability to understand their nonverbal behaviour but did build rapport with Ellie at the same level they did with a human therapist. Interestingly, the users built higher rapport with Ellie with an accurate verbal response using Wizard-of-OZ (WoZ) than they did with Ellie with an automated response or with human therapist. However, when they believed the same WoZ agent was controlled by a human rather than being automated, this rapport decreases

and the users fear of disclosing information increases [105]. The findings indicate, first, the value of the verbal content of the message rather than the non-verbal cues in building the user-agent relationship and, second, the preference for automated over human agents by users when discussing health topics, as confirmed by other studies [103,106].

Users showed greater self-disclosure to the empathic agent when it starts to self-disclose to them [107]. In a study with 57 participants, Kang and Gratch [107] used an empathic agent that expresses nonverbal behaviours (e.g., nodding and gestures), with three self-disclosure settings: non-disclosure, low-disclosure and high disclosure. They reported that this simple verbal empathy condition (high disclosure only) motivated the participants to disclose more information with the agent.

The above studies did not seek to build or measure a human-agent WA. In the case of Greta, this is possibly because Greta was not originally designed for the purpose of therapy; the use of empathy by Greta and Ellie was to improve their ability to actively listen, rather than to provide advice. Ellie, described as a virtual interviewer, does not provide recommendations and thus is also not interested in adherence or behaviour change. In contrast, the ECA Dr Evie was designed to build a WA with paediatric patients and their families to encourage adherence to treatment advice. By adding Dr Evie to an interactive website that provided tailored advice, adherence significantly improved and led to improved health outcomes [108].

Bickmore and Picard [109] argue that including social-emotional cues (relational cues) in the agent's dialogue improves the user-agent relationship, namely the bond element of the WA, and consequently improves the patients' behaviour change. They implemented these ten cues to build Laura to encourage daily physical activity. The 101 young adults who participated in the study were motivated to interact with the ECA daily during the intervention. The relational agent Laura (i.e., with relational cues), compared to non-relational Laura, was significantly more successful in building a WA with the participants after 7 and 27 days; however, the participants reported a similar daily number of steps using the MIT FitTrack system. The Laura with relational cues was also successful in encouraging 16 patients with schizophrenia to adhere to their medication and do physical activity [110]. The same relational agent was then used in several clinical studies by the group [98].

In another study, Bickmore [60] investigated the nonverbal behaviour of 12 participants who interacted either with a human therapist or an ECA to receive advice about diet and exercise. They reported that participants used the same nonverbal behaviours (i.e., eye, eyebrow, head and hand movements) when they interacted with the ECA just as they did with the human. Bickmore, et al. [111] also found that patients with adequate health literacy ( $N = 19$ ), compared with those with low health literacy ( $N = 11$ ), felt more comfortable talking with the ECA that explained the discharge instructions rather than a human nurse because the ECA was not judgemental and could repeat the instructions as often as the patients needed without making them feel inadequate. Similarly, in another study with 44 adults aged between 55 to 82, some participants reported their preference to discuss their spiritual matters and talk about death with the ECA rather than a human being, as they believed the human could be opinionated [61].

Schulman and Bickmore [112] built Elizabeth with social dialogue to persuade users to do regular exercise. The user could speak a selected utterance from the answers designed for the user or say something else. The researchers used the WoZ approach to control Elizabeth's responses. To measure changes in the user's attitude towards exercise, Schulman and Bickmore [112] provided 47 university student participants with a list of positive and negative statements that the user had to rank from most to least important. They compared Elizabeth with social dialogue to another ECA with neutral dialogue (no social cues), text-based social dialogue and text-based dialogue with no social cues. Social dialogue had a significant impact in changing the users' attitude towards exercise, but not in building a bond. Although the established bond correlated significantly with the



perception of the agent and persuasive message, it did not correlate with the user's attitude towards exercise.

The concept of a WA underpins all of the abovementioned work of Bickmore and colleagues involving their relational agents, including Laura and Elizabeth. They have used the WAI in their studies to measure the existence of a WA [113]. The relational cues (such as, social dialogue, humour, empathic feedback) used in the dialogues, however, seem to have greatest impact on the bond subscale and minimal impact on goal and task subscales [113]. This suggests that improvements can be made to the WA developed.

Bickmore, et al. [101] proposed a reusable framework that connects the various models and theories required to build an ECA for behaviour change interventions. They applied the proposed framework incorporating the Transtheoretical Model (TTM) [114], social-cognitive theory and motivational intervention theories to build an ECA called Karen for encouraging physical activity and fruit and vegetable consumption [115]. The framework follows Shared Plans theory by treating "dialogue as a collaboration in which participants coordinate their action towards achieving a shared goal". The conversations seek to develop mutual understanding. Over a two month period, 122 participants were randomly assigned to the control group, or to interact with Karen to receive diet counselling, physical activity counselling, or both. Compared to the control group, participants showed significant increase in their fruit and vegetable consumption after interacting with Karen to receive diet counselling only, and in their number of daily steps after interacting with Karen to receive physical activity counselling only. However, participants who received both did not show a significant difference compared to the control group, which could be because they received half of the content of the counselling for every piece of advice. The relationship was not measured in the study, so the effect of a WA could not be determined.

The findings of the abovementioned studies indicate the role of an ECA's verbal and nonverbal empathy to encourage behaviour change, but they are not firm enough for a solid conclusion [47]. Other researchers have also investigated the use of therapeutic strategies such as the Transtheoretical Model and motivational interviewing. For example, Fiorella, et al. [116] proposed an ECA called Valentina to promote healthy eating. The dialogue is built to detect and adapt to the user's stage of change following the Transtheoretical Model [114]. The agent architecture includes a user model that stores the user's static information (e.g., name, age and background) and dynamic information (e.g., state of change). To establish a relationship with the user, the dialogue includes emotional and social cues such as sympathy, appreciation and disappointment. The study did not involve an implemented agent and the evaluation did not include standard measures such as questionnaires to evaluate the user-agent relationship, acceptance or behaviour change. In the WoZ study, users showed progress in their readiness to change their behaviour during the conversation.

Lisetti [117] identified 10 advantages of interacting with ECAs over humans in health contexts which includes increased accessibility, increased confidentiality and divulgence, tailored information, diminished variability, avoidance of righting reflex with infinite patience, addresses low literacy, lower attrition rates, allows patient-physician concordance/matching, provide working alliance and express empathy. Drawing on these advantages, Lisetti, et al. [118] introduced the On-Demand Virtual Counselor (ODVIC) to help patients with excessive drinking behaviour. The ECA's counselling dialogue was crafted following a motivational interviewing strategy, with its responses tailored to the patient's emotional state. The ODVIC agent Amy was evaluated with 51 patients assigned to three experimental groups to interact with an empathic ECA, a non-empathic ECA or text-based agents, and several evaluation measures were utilised, such as anthropomorphism and trust. However, changes in the patients' behaviours were not evaluated [119].

Olafsson, et al. [120] proposed an automated counselling framework to help substance users. They extracted the labels and therapist dialogue acts from real face-to-face counselling sessions and built a machine-learning model to automatically recognise and respond to the patients following a motivational interviewing strategy. The framework

was evaluated with 23 opioid users under treatment, who reported being more comfortable talking to the ECA about their substance problem rather than to a human therapist. They reported a neutral relationship with the ECA called Tinna, although they trusted its advice. Olafsson, et al. [120] also reported that substance users preferred to discuss their problems with the ECA more than the human therapist.

Conversely, there is evidence that the content of the delivered message could be a more important determinant in fostering health-related behavioural change than the way the ECA interacts with the user. For example, the abovementioned Schulman and Bickmore [112] study reported a slight to no effect of relational cues in building the user-agent bond and that a persuasive message is adequate to foster health behaviour change. Murali, et al. [121] assigned 40 Indians who moved to the US after the age of 16 to interact with a virtual coach of Indian or American appearance to promote physical activity with Indian or American culturally tailored argumentation. Compared to the appearance, the argumentation played a significant role in participant satisfaction and in persuading them to change their decision balance and self-efficacy. We suggest that the benefit of the argumentation in this study was due to cultural tailoring which made the recommendation relevant to the individual. We next look at the use of explanation to build a bond and mutual understanding, consistent with WA, through clarifying the relevance of a recommended behaviour.

### 3.2. Explainable Agents in Healthcare

Concerning the patient's adherence, providing recommendations to a patient is not enough to ensure adherence, the recommendation requires personalised and relevant explanation [122]. The facilitating role of explanation in the health domain was investigated early in the 1980s to understand why a diagnosis was made by an expert system like MYCIN and PUFF with emphasis on its important role as the major requirement to be met in a healthcare application [123,124]. Later, interest in explanation diminished significantly in line with the dramatic change in the type of health applications from diagnostic to reminder systems, preventive and educational applications [125]. However, researchers reported a low acceptance of decision support systems in the health domain and attributed it to a lack of user understanding and decision relevance [126].

Nevertheless, researchers have continued to introduce diagnostic systems with explanation to increase the acceptance of decision-making systems by expert users (i.e., healthcare giver). For example, Gage, et al. [127], Lip, et al. [128], and Letham, et al. [129] introduced three applications to calculate the probability of having a stroke in the future for a patient with atrial fibrillation. The applications can interpret the resultant score based on the probability of a pre-crafted set of contributing features.

Differences have been found in the explanation preferences of children and adults. Using a Nao-robot to educate children with Type 1 diabetes, Kaptein, et al. [95] used GHT to provide goal and belief-based explanations to 19 children and parents. Both groups preferred goal-based explanations, but adults' preferences were significantly stronger than the childrens' preferences. These findings confirm that appropriateness of an explanation highly depends on the situation and the receiver [130]. No data was captured to measure mutual understanding, relevance or perceived relationship with the Nao-robot.

The study by Kaptein, et al. [95] is one of few investigating the role of XAs in health. Where explanations are provided by ECAs, they tend to be in the form of general education and guidance to users with little sense of personalisation. An example of an ECA providing explanations in the form of educating and guiding the users is the study by Bickmore, et al. [37], with 29 participants. The study found users were more likely to consent to participate in a study after receiving an explanation about the study by an agent rather than a human, particularly those with low health literacy. Similar results have been reported by Zhou, et al. [131], with 149 patients who interacted with a virtual discharge nurse to receive instructions and explanation on the discharge procedure and what to do after. Two conclusions could be drawn from these studies. First, explanation plays a vital role in the user's decision-making process concerning their health. Second, the use

of an ECA could be more beneficial than a human in contexts where patients feel more comfortable disclosing information or discussing sensitive topics with ECAs rather than a human. We do not classify these ECAs as XAs because they do not explain their reasoning or the human's reasoning.

While the literature reports the need to modify the belief and goal-based explanations according to some features in the user's profile, to-date only approximately 8% of current work in XAs include personalisation and user models [79]. The majority of that work, however, uses the agent's beliefs and goals in the explanation. In the context of behavior change to manage study stress, where the human is recommended to perform an action, the work of Abdulrahman, et al. [132] suggests that for the recommendation and explanation to be relevant and persuasive, the reason explanation needs to be based on the beliefs and goals of the user. The work of [132,133] aims to increase the user's behaviour intention using an ECA, named Sarah, that provides recommendations with explanations to consider the user's beliefs or mental state, rather than the ECA's. That work distinguishes between the behaviours of the agent and the user, and investigates the following question: how does a user perceive the agent's explanations that refer to the the user's own beliefs or goals in terms of their intention to change their behaviours as recommended by the agent? A study with 91 participants who received recommendations with belief-based, goal-based or belief & goal-based explanation rated the explainable virtual advisors (XVAs) similarly concerning trustworthiness and the level of WA. Besides the user profile, this relationship was able to predict up to 40% of the change in the intention to do the recommended behaviours after interacting with an XVA along with some factors from the users' profiles. This finding aligns with the concept of WA: where adherence can be predicted by the level of therapist-patient relationship. While for most of the recommended behaviours there was a significant change in intention following the interaction with the XA, this was true for all explanation patterns (belief only, goal only, belief and goal). But for some behaviours, the work found that longer explanations, using the belief and goal pattern, was a hindrance in intention to perform that behaviour. This work shows the importance of relevance, communication and WA which is demonstrated in an explanation that includes what has been discussed between the agent and human.

### 3.3. A Comparison of Health ECAs and XAs through WA Lens

ECAs offer several features essential for a WA and which are not common in other eHealth technologies. A review by Hillier [134] exploring the nature of the WA delivered in digital interventions, identified the importance of *interactivity* and *availability*. Another recent comprehensive narrative review of working alliance in Digital Mental Health Interventions by Tremain, et al. [135], questions the ability of unsupported digital interventions, that is interventions that involve technology without human support, to provide the *bond element* in Bordin's WA definition. They cite two exceptions, the Bickmore, et al. [113] study mentioned above and a study by Holter, et al. [136] involving the CA Endre that used human strategies including humor, personalization and empathy to build a WA to encourage smoking cessation, however WA and outcomes were not measured. The review by Tremain, et al. [135] found that digitally delivered WA interventions did not directly link to improved health outcomes, in contrast to interventions involving what they call "face-to-face therapies". They conclude the need for technology to "emulate relational characteristics" and investigate how digital interventions can not only build WA but also deliver improved health outcomes. We believe that ECAs are able to tick the box for interactivity, availability, bond element, emulate relational characteristics and even deliver face-to-face therapies, if one considers that both the ECA and the human both have faces and communicate together. So how can we improve the WA and benefits ECAs deliver?

Credibility has been identified as a factor that influences WA from the client's perspective [137]. Credibility is tightly connected with trust. Trust in a mobile-based intervention affects the quality of the WA [138]. While there are many definitions and operationalisations of trust, the most common integrated model is by Mayer, et al. [139], which

operationalised trust as the perception of the three factors of trustworthiness: ability, benevolence and integrity. Ability equates to competence. Due to this some ECA researchers (e.g., [119,120,132,133]) have measured trust and its influence on the human-ECA relationship. As argued in Section 2.3, providing explanation plays an important role in development of trust. Hence, we argue the need for XAs and propose that, without explanation, an ECA will be less able to gain trust and build a WA.

Drawing together important factors for development and measurement of WA, we compare the ECAs and XAs presented in the previous two sections. Table 1 lists examples of ECAs for behaviour change. We don't include the criteria of accessibility, as this is a function and driver for all ECAs. The first column identifies the ECA/XA followed by the targeted health behaviour in column 2. Concerning the feature of interactivity, we consider the nature of user inputs and ECA outputs, columns 3 and 4. Column 5 identifies the theory underlying the ECA's design to identify whether WA or another strategy was used. Given the issues raised by Tremain, et al. [135], concerning whether digital interventions can achieve the bond element of the WA and deliver improved health outcomes, we compare the measures of the human-ECA relationship and change in behaviour.

**Table 1.** Comparison of ECA/XAs for Behaviour Change.

ECA/XA	Target Behaviour	Interactivity		Main Design Strategy/Theory	Measures	
		User Input	ECA Output		User-Agent Relationship	Behaviour Change
ECA Greta [140]	Healthy eating	Multiple choice	Facial expressions and spoken utterances (TTS)	Adaptive emotional expression	1 question each on liking and trust	Likelihood of following the advice after the interaction
ECA Laura [109]	Exercise	Multiple choice	Facial expressions, movement of body, hand and head, and spoken utterance (TTS)	Relational cues, WA	WAI, liking the agent	Daily walking steps
ECA Elizabeth [112]	Exercising	Multiple choice (spoken)	Spoken utterances (WoZ)	Relational cue, WA	Bond from WA	A change in attitude towards exercise
ECA Ellie [141]	PTSD	Video, voice, natural language	Facial expressions, movement of body, hand and head, and spoken utterance (TTS)	Empathy	Rapport	Willingness to disclose
ECA Karen [115]	Physical activity, diet	Multiple choice (spoken)	Spoken utterance (TTS)	TTM, social cognitive theory and motivational interviewing, WA	Self-report satisfaction	Number of steps, and amount of fruit and vegetables consumed
ECA Amy [119]	Excessive alcohol consumption	Facial expression	Facial expressions and spoken utterance (TTS)	Empathy	1 trust question, social presence	-
ECA Tinna [120]	Substance use	Speech	Spoken utterance (TTS)	Motivational interviewing	2 trust, 4 relationship questions	-
ECA Dr Evie [108]	Pediatric incontinence	Multiple choice (spoken)	Spoken utterance (TTS) and text	WA, empathy	WAI	Increased adherence, reduced incontinence
XA Nao [95]	Education on Type II diabetes	None—no interaction	Spoken utterance (TTS) and text	GHT to measure preference for Goal or belief based explanations	-	-
XA Sarah [132,133]	Social interaction, healthy eating, exercise, caffeine reduction	Multiple choice (spoken)	Spoken utterance (TTS) and text	WA, reason explanation, relational cues/empathy	WAI, trustworthiness (ability, benevolence and integrity)	Change in behaviour intention for the 4 behaviours

In summary, ECAs have been investigated to address the problem of health behaviour change mainly for physical activity and diet. The reported results are promising and indicate the capability of ECAs in motivating users to adopt healthier behaviours or abandon unhealthier ones. Several studies attempt to address the problem by establishing the user-agent relationship or endowing the agent with therapist practice skills. To build this

relationship, researchers designed their ECAs with verbal and/or nonverbal behaviours. While only half of the ECA/XAs were designed with WA in mind, most of those studies hypothesise the efficiency of the use of socio-emotional cues to build the user-agent relationship, their main focus in the evaluation was to measure the user's bond with or trust in the agent. These ECAs boost the sense of a user-agent WA but with conflicting results for the impact on behaviour change [47]. Empathy has been the driver for the design of some ECAs. While empathy is efficient in building the human-agent relationship, it has a different definition and operationalisation as a means to build mutual understanding, particularly in the health domain [21], and is less relevant for encouraging change in behavior intentions. Finally, we note that only 2 out of the 10 ECAs reviewed also offer explanation.

#### 4. Conclusions and Future Directions

As a technology-based intervention, ECAs could successfully play the role of the therapist and build a WA [98]. This relationship implies establishing user-agent mutual understanding which could be the result of positive emotional communication [20] or providing an appropriate explanation [22]. Affective agents that use empathic verbal and non-verbal communication behaviours have been shown to build strong rapport [20] and be more persuasive [142]. Given that socio-emotional cues may not be enough to foster a WA that can promote adherence, more attention should be paid to user-agent collaboration on the treatment tasks and goals during the interaction [112,113]. This prior finding is congruent with the principle of a WA. The WA mainly refers to the mutual understanding between the user and the therapist (the ECA, here), which is achieved by collaboration and agreement on the treatment tasks and goals, which consequently builds the bond. This agreement could be achieved through explanation that eventually fosters user-agent trust [22]. Thus, we review the use of explanation in building a WA for health behaviour change.

A key question remains concerning the need and ability of ECAs to build and maintain long-term relationships [109,143–145]. Without studies that test long-term interaction, it is not feasible for participants to realistically comment on their relationship with the agent [120]. The improved health outcomes delivered by Dr Evie were only possible due to six months access [108]. Where the ECA acts as a coach, for example for weight loss [109], one can expect an ECA to be consulted on a daily basis and to be able to keep the human engaged for that extended period [146]. The problem of repetitiveness motivated development of the DTASK framework [37]. Based on the findings of two RCTs, Bickmore, et al. [146] conclude that autobiographical and interactive, e.g., involving the user, storytelling by the agent and including references to current weather and sport events, taken from the internet, are two key strategies for maintaining user engagement over time. Cole-Lewis, et al. [147] identify the need for different definitions and measures for engagement for short and long-term interactions. They argue that for (medium to long-term) digital behavior change interventions, ECAs need to go beyond (short-term) health behavior engagement, to include features that support behavior change and frequency of use, as measured by length and depth of use.

The ECAs presented in this review each provide their own expertise. However, in reality, health consumers are likely to require advice from more than one health domain expert [148]. This can in part be solved by ECAs that provide advice for more than one domain (e.g., [149]). However, a single coach approach raises two key issues: what advice to follow first and what to do if the advice is conflicting. The Council of Coaches application [150,151] offers a holistic ECA-based solution to both of these problems by providing multiple ECA coaches that collectively manage the individual's health behaviours including two obligatory coaches one for physical activity and another for diet, and diabetes and chronic pain coaches for users with those health issues. There are also expert coaches in cognition and social activity and a peer support agent and an assistant that guides use of the application. Consistent with the suggestion to provide biographic stories [146], these coaches not only have their own roles and expertise, they each have their own life story.



Users can select which coach they want to speak with and other coaches can add in their viewpoint as needed. A 5–9 week study in the Netherlands with 51 adults aged 55 and over (mean age 65.3), found that after 5 weeks, 21 participants continued to interact with the Council of Coaches and reported minimal clinically important differences in their quality of life [151].

The findings from the literature provide successful examples of ECAs' reason explanations, using the agent's beliefs and goals to explain their behaviours, with evidence of the importance of tailoring the explanation pattern to the user profile [95,152]. However, these explainable ECAs/XAs did not take into account the human user's beliefs and goals. Initial work has begun in this space (e.g., [132,133]), but in general, CAs need greater understanding of the user including medical histories and preferences stored as user-specific models and real-time physiological and emotional signals from users to provide personalised and tailored interactions with ECAs over the long term. Interpretation of various inputs, generation of recommendations and reasoning about appropriate ECA responses will require utilisation of and further advances in other areas of artificial intelligence such as machine learning, image processing and natural language processing.

As technology advances further and CAs play important roles in health behaviour change, the ethical appropriateness of building a WA with a computer should be considered. Porra, et al. [153] argue "that feelings are the very substance of our humanness and therefore are best reserved for human interaction" (p. 533). Hudlicka [154] identifies that, in addition to common concerns such as data privacy, CAs raise concerns of affective privacy (the right to keep your feelings private), emotion induction (manipulation of one's feelings), and creation of human-agent relationships that might create dependencies, loss of touch with reality, replacement of human social relationships or other unhealthy outcomes. Recognising that learning is a social activity which can be supported by pedagogical agents and acknowledging the potential risk that students might prefer to interact with pedagogical agents instead of learning to engage with human teachers and peer-learners, Richards and Dignum [153] recommend a design for values approach to develop and deploy ethical CAs involving openness and participation following the principles of accountability, responsibility and transparency. Following the five AI4People ethical principles of beneficence, non-maleficence, justice, autonomy and explainability [154], XAs can play an important role to empower humans by making them aware of the relevant decision process and options, helping them to reflect on their current behaviours and what they hope to achieve and to educate consumers on the different choices through personal and relevant recommendations and explanation. In the health behaviour change context, CAs and XAs in particular, can fill the gap in access to costly and limited human health professionals and potentially offer some advantages [117], to allow individuals to achieve personal and societal goals in which the consumer/patient exercises self-management of their health and wellbeing behaviours.

**Author Contributions:** Conceptualization, A.A. and D.R.; methodology, A.A. and D.R.; software, A.A.; validation, A.A. and D.R.; formal analysis, A.A.; investigation, A.A. and D.R.; resources, A.A. and D.R.; data curation, A.A.; writing—original draft preparation, A.A. and D.R.; writing—review and editing, A.A. and D.R.; visualization, A.A.; supervision, D.R.; project administration, D.R.; funding acquisition, D.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by an International Macquarie University Research Training Program (iMQ RTP) scholarship.

**Institutional Review Board Statement:** This review did not involve conduct of any studies with human participants. All research studies conducted by the authors which build on this review or which are cited in this review were conducted according to the guidelines of the Declaration of Helsinki, and approved by the Human Research Ethics Committee, Macquarie University with reference number: 52020535012841 dated 25 May 2019.

**Informed Consent Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. WHO. *Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013–2020*; WHO: Geneva, Switzerland, 2013.
2. Lehane, E.; McCarthy, G. Intentional and Unintentional Medication Non-Adherence: A Comprehensive Framework for Clinical Research and Practice? A Discussion Paper. *Int. J. Nurs. Stud.* **2007**, *44*, 1468–1477. [\[CrossRef\]](#)
3. Rueda, S.; Park-Wyllie, L.Y.; Bayoumi, A.; Tynan, A.M.; Antoniou, T.; Rourke, S.; Glazier, R. Patient support and education for promoting adherence to highly active antiretroviral therapy for HIV/AIDS. *Cochrane Database Syst. Rev.* **2006**, CD001442. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Pop-Eleches, C.; Thirumurthy, H.; Habyarimana, J.P.; Zivin, J.G.; Goldstein, M.P.; De Walque, D.; Mackeen, L.; Haberer, J.; Kimaiyo, S.; Sidle, J.; et al. Mobile Phone Technologies Improve Adherence to Antiretroviral Treatment in a Resource-Limited Setting: A Randomized Controlled Trial of Text Message Reminders. *AIDS* **2011**, *25*, 825–834. [\[CrossRef\]](#)
5. Nutbeam, D. The Evolving Concept of Health Literacy. *Soc. Sci. Med.* **2008**, *67*, 2072–2078. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Atkins, L.; Fallowfield, L. Intentional and Non-intentional Non-adherence to Medication Amongst Breast Cancer Patients. *Eur. J. Cancer* **2006**, *42*, 2271–2276. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Sieverink, F.; Kelders, S.M.; van Gemert-Pijnen, J.E. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. *J. Med. Internet Res.* **2017**, *19*, e402. [\[CrossRef\]](#)
8. Navarra, A.-M.D.; Gwadz, M.V.; Whittemore, R.; Bakken, S.R.; Cleland, C.M.; Burleson, W.; Jacobs, S.K.; Melkus, G.D.E. Health Technology-Enabled Interventions for Adherence Support and Retention in Care Among US HIV-Infected Adolescents and Young Adults: An Integrative Review. *AIDS Behav.* **2017**, *21*, 3154–3171. [\[CrossRef\]](#)
9. Christensen, H.; Griffiths, K.M.; Farrer, L. Adherence in internet interventions for anxiety and depression: Systematic review. *J. Med. Internet Res.* **2009**, *11*, e13. [\[CrossRef\]](#)
10. Badawy, S.M.; Thompson, A.A.; Kuhns, L.M. Medication Adherence and Technology-Based Interventions for Adolescents With Chronic Health Conditions: A Few Key Considerations. *JMIR mHealth uHealth* **2017**, *5*, e202. [\[CrossRef\]](#)
11. Heynsbergh, N.; Heckel, L.; Botti, M.; Livingston, P.M. Feasibility, Useability and Acceptability of Technology-Based Interventions for Informal Cancer Carers: A Systematic Review. *BMC Cancer* **2018**, *18*, 244. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Kelders, S.M.; Kok, R.N.; Ossebaard, H.C.; Van Gemert-Pijnen, J.E. Persuasive system design does matter: A systematic review of adherence to web-based interventions. *J. Med. Internet Res.* **2012**, *14*, e152. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Bennett, J.K.; Fuertes, J.N.; Keitel, M.; Phillips, R. The Role of Patient Attachment and Working Alliance on Patient Adherence, Satisfaction, and Health-Related Quality of Life in Lupus Treatment. *Patient Educ. Couns.* **2011**, *85*, 53–59. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Bordin, E.S. The Generalizability of the Psychoanalytic Concept of the Working Alliance. *Psychother. Theory Res. Pract.* **1979**, *16*, 252. [\[CrossRef\]](#)
15. Harman, K.; MacRae, M.; Vallis, M.; Bassett, R.J.P.C. Working with people to make changes: A behavioural change approach used in chronic low back pain rehabilitation. *Physiother. Can.* **2014**, *66*, 82–90. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Kaplan, B.; Farzanfar, R.; Friedman, R.H. Personal Relationships with an Intelligent Interactive Telephone Health Behavior Advisor System: A Multimethod Study Using Surveys and Ethnographic Interviews. *Int. J. Med. Inform.* **2003**, *71*, 33–41. [\[CrossRef\]](#)
17. O’Keefe, D.J. Trends and Prospects in Persuasion Theory and Research. In *Readings in Persuasion, Social Influence, and Compliance Gaining*; Pearson/Allyn and Bacon: Boston, MA, USA, 2004; pp. 31–43.
18. Provoost, S.; Lau, H.M.; Ruwaard, J.; Riper, H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. *J. Med. Internet Res.* **2017**, *19*, e151. [\[CrossRef\]](#)
19. Roberts, S.; Chaboyer, W.; Gonzalez, R.; Marshall, A. Using technology to engage hospitalised patients in their care: A realist review. *BMC Health Serv. Res.* **2017**, *17*, 388. [\[CrossRef\]](#)
20. Huang, L.; Morency, L.-P.; Gratch, J. Virtual Rapport 2.0. In *International Workshop on Intelligent Virtual Agents*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 68–79.
21. Hojat, M. *Empathy in Patient Care: Antecedents, Development, Measurement, and Outcomes*; Springer Science & Business Media: Berlin, Germany, 2007.
22. Walton, D. Dialogical Models of Explanation. *ExaCt* **2007**, *2007*, 1–9.
23. Mathieu, J.E.; Heffner, T.S.; Goodwin, G.F.; Salas, E.; Cannon-Bowers, J.A. The Influence of Shared Mental Models on Team Process and Performance. *J. Appl. Psychol.* **2000**, *85*, 273–283. [\[CrossRef\]](#)
24. Graßmann, C.; Schölmerich, F.; Schermuly, C.C. The Relationship between Working Alliance and Client Outcomes in Coaching: A Meta-Analysis. *Hum. Relat.* **2020**, *73*, 35–58. [\[CrossRef\]](#)
25. Horvath, A.O.; Symonds, B.D. Relation between Working Alliance and Outcome in Psychotherapy: A Meta-Analysis. *J. Couns. Psychol.* **1991**, *38*, 139. [\[CrossRef\]](#)
26. Krupnick, J.L.; Sotsky, S.M.; Elkin, I.; Simmens, S.; Moyer, J.; Watkins, J.; Pilkonis, P.A. The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: Findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Focus* **2006**, *4*, 532–539. [\[CrossRef\]](#)
27. Catty, J. ‘The Vehicle of Success’: Theoretical and Empirical Perspectives on the Therapeutic Alliance in Psychotherapy. *Psychol. Psychother. Theory Res. Pract.* **2004**, *77*, 255–272. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Bordin, E. Of human bonds that bind or free. In *Proceedings of the Annual Meeting of the Society for Psychotherapy Research*, Pacific Grove, CA, USA, 17–20 June 1980.

29. Horvath, A.O.; Greenberg, L.S. Development and Validation of the Working Alliance Inventory. *J. Couns. Psychol.* **1989**, *36*, 223. [CrossRef]
30. Sherwood, G.; Thomas, E.; Bennett, D.S.; Lewis, P. A teamwork model to promote patient safety in critical care. *Crit. Care Nurs. Clin. N. Am.* **2002**, *14*, 333–340. [CrossRef]
31. Bratman, M. *Intention, Plans, and Practical Reason*; Harvard University Press: Cambridge, MA, USA, 1987; Volume 10.
32. Grosz, B.J.; Kraus, S. The Evolution of SharedPlans. In *Applied Logic Series*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 227–262.
33. Dorri, A.; Kanhere, S.S.; Jurdak, R. Multi-agent systems: A survey. *IEEE Access* **2018**, *6*, 28573–28593. [CrossRef]
34. Rizk, Y.; Awad, M.; Tunstel, E.W. Cooperative heterogeneous multi-robot systems: A survey. *ACM Comput. Surv. (CSUR)* **2019**, *52*, 1–31. [CrossRef]
35. Grosz, B.J.; Sidner, C.L. *Plans for Discourse*; BBN LABS Inc.: Cambridge, MA, USA, 1988.
36. Rich, C.; Sidner, C.L.; Lesh, N. Collagen: Applying collaborative discourse theory to human-computer interaction. *AI Mag.* **2001**, *22*, 15.
37. Bickmore, T.; Schulman, D.; Shaw, G. Dtask and Litebody: Open Source, Standards-Based Tools for Building Web-Deployed Embodied Conversational Agents. In *International Workshop on Intelligent Virtual Agents*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 425–431.
38. Smith, A.; Page, D.U.S. Smartphone Use in 2015. 2015. Available online: <http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015> (accessed on 1 April 2015).
39. Miner, A.S.; 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 Intern. Med.* **2016**, *176*, 619–625. [CrossRef]
40. Laranjo, L.; Dunn, A.G.; Tong, H.L.; Kocaballi, A.B.; Chen, J.; Bashir, R.; Surian, D.; Gallego, B.; Magrabi, F.; Lau, A.Y.S.; et al. Conversational agents in healthcare: A systematic review. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1248–1258. [CrossRef]
41. Hodgson, T.; Magrabi, F.; Coiera, E. Efficiency and safety of speech recognition for documentation in the electronic health record. *J. Am. Med. Inform. Assoc.* **2017**, *24*, 1127–1133. [CrossRef]
42. Kennedy, C.M.; Powell, J.; Payne, T.H.; Ainsworth, J.; Boyd, A.; Buchan, I. Active assistance technology for health-related behavior change: An interdisciplinary review. *J. Med. Internet Res.* **2012**, *14*, e80. [CrossRef]
43. Isbister, K.; Doyle, P. The blind men and the elephant revisited. In *From Brows to Trust*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 3–26.
44. Bittner, E.; Oeste-Reiß, S.; Leimeister, J.M. Where Is the Bot in Our Team? Toward a Taxonomy of Design Option Combinations for Conversational Agents in Collaborative Work. In Proceedings of the 52nd Hawaii International Conference on System Sciences, Hawaii, HI, USA, 8 January 2019.
45. Diederich, S.; Brendel, A.B.; Kolbe, L.M. Towards a Taxonomy of Platforms for Conversational Agent Design. In Proceedings of the 14th International Conference on Wirtschaftsinformatik, Siegen, Germany, 24–27 February 2019.
46. Zierau, N.; Wambsganss, T.; Janson, A.; Schöbel, S.; Leimeister, J.M. The anatomy of user experience with conversational agents: A taxonomy and propositions of service clues. In Proceedings of the International Conference on Information Systems (ICIS), Hyderabad, India, 13–16 December 2020.
47. Ter Stal, S.; Kramer, L.L.; Tabak, M.; op den Akker, H.; Hermens, H. Design Features of Embodied Conversational Agents in eHealth: A Literature Review. *Int. J. Hum.-Comput. Stud.* **2020**, *138*, 102409. [CrossRef]
48. Weizenbaum, J. ELIZA—A computer program for the study of natural language communication between man and machine. *Commun. ACM* **1966**, *9*, 36–45. [CrossRef]
49. Colby, K.M. Artificial Paranoia: A Computer Simulation of Paranoid Processes. *Behav. Ther.* **1975**, *7*. [CrossRef]
50. Crutzen, R.; Peters, G.-J.Y.; Portugal, S.D.; Fisser, E.M.; Grolleman, J.J. An Artificially Intelligent Chat Agent That Answers Adolescents' Questions Related to Sex, Drugs, and Alcohol: An Exploratory Study. *J. Adolesc. Health* **2011**, *48*, 514–519. [CrossRef]
51. Ireland, D.; Atay, C.; Liddle, J.J.; Bradford, D.; Lee, H.; Rushin, O.; Mullins, T.; Angus, D.; Wiles, J.; McBride, S. Hello harlie: Enabling speech monitoring through chat-bot conversations. *Stud. Health Technol. Inform.* **2016**, *227*, 55–60. [PubMed]
52. Colby, K.M. A Computer Program Using Cognitive Therapy to Treat Depressed Patients. *Psychiatr. Serv.* **1995**, *46*, 1223–1225. [CrossRef] [PubMed]
53. Fitzpatrick, K.K.; 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**, *4*, e19. [CrossRef]
54. Grudin, J.; Jacques, R. Chatbots, humbots, and the quest for artificial general intelligence. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–11.
55. Cassell, J. Embodied conversational agents: Representation and intelligence in user interfaces. *AI Mag.* **2001**, *22*, 67.
56. Van Pinxteren, M.M.; Pluymaekers, M.; Lemmink, J.G. Human-like communication in conversational agents: A literature review and research agenda. *J. Serv. Manag.* **2020**, *31*, 203–225. [CrossRef]
57. Isbister, K.; Nass, C. Consistency of Personality in Interactive Characters: Verbal Cues, Non-Verbal Cues, and User Characteristics. *Int. J. Hum.-Comput. Stud.* **2000**, *53*, 251–267. [CrossRef]
58. Clore, G.L.; Ortony, A. Psychological Construction in the OCC Model of Emotion. *Emotion Rev.* **2013**, *5*, 335–343. [CrossRef]
59. Picard, R.W. *Affective Computing*; MIT Press: Cambridge, MA, USA, 2000.

60. Bickmore, T. Towards the Design of Multimodal Interfaces for Handheld Conversational Characters. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2002.
61. Utami, D.; Bickmore, T.; Nikolopoulou, A.; Paasche-Orlow, M. Talk about death: End of life planning with a virtual agent. In *International Conference on Intelligent Virtual Agents*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 441–450.
62. Yee, N.; Bailenson, J.N.; Rickertsen, K. A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, CA, USA, 28 April–3 May 2007; pp. 1–10.
63. Rheu, M.; Shin, J.Y.; Peng, W.; Huh-Yoo, J. Systematic Review: Trust-Building Factors and Implications for Conversational Agent Design. *Int. J. Hum.-Comput. Interact.* **2020**, *37*, 81–96. [\[CrossRef\]](#)
64. Bian, Y.; Yang, C.; Guan, D.; Xiao, S.; Gao, F.; Shen, C.; Meng, X. Effects of pedagogical agent's personality and emotional feedback strategy on Chinese students' learning experiences and performance: A study based on virtual Tai Chi training studio. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose, CA, USA, 7–12 May 2016; pp. 433–444.
65. McRorie, M.; Sneddon, I.; McKeown, G.; Bevacqua, E.; de Sevin, E.; Pelachaud, C. Evaluation of Four Designed Virtual Agent Personalities. *IEEE Trans. Affect. Comput.* **2011**, *3*, 311–322. [\[CrossRef\]](#)
66. De Rosi, F.; Pelachaud, C.; Poggi, I.; Carofiglio, V.; De Carolis, B. From Greta's Mind to Her Face: Modelling the Dynamics of Affective States in a Conversational Embodied Agent. *Int. J. Hum.-Comput. Stud.* **2003**, *59*, 81–118. [\[CrossRef\]](#)
67. Paiva, A.; Leite, I.; Boukricha, H.; Wachsmuth, I. Empathy in virtual agents and robots: A survey. *ACM Trans. Interact. Intell. Syst. (TiiS)* **2017**, *7*, 1–40. [\[CrossRef\]](#)
68. "Explanation". Oxford English Dictionary. Available online: <https://www.oed.com/view/Entry/66604?rskey=wFVxij&result=1&isAdvanced=true> (accessed on 28 June 2021).
69. Glass, A.; McGuinness, D.L.; Wolverton, M. Toward establishing trust in adaptive agents. In *Proceedings of the 13th International Conference on Intelligent User Interfaces*, Gran Canaria, Spain, 13–16 January 2008; pp. 227–236.
70. Hilton, D.J. Conversational Processes and Causal Explanation. *Psychol. Bull.* **1990**, *107*, 65–81. [\[CrossRef\]](#)
71. Van den Bosch, K.; Harbers, M.; Heuvelink, A.; van Doesburg, W. Intelligent agents for training on-board fire fighting. In *International Conference on Digital Human Modeling*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 463–472. [\[CrossRef\]](#)
72. Scott, A.C.; Clancey, W.J.; Davis, R.; Shortliffe, E.H. *Explanation Capabilities of Production-Based Consultation Systems*; Stanford University CA Department of Computer Science: Stanford, CA, USA, 1977.
73. Swartout, W.R. Explaining and justifying expert consulting programs. In *Computers and Medicine*; Springer: Berlin/Heidelberg, Germany, 1985; pp. 254–271.
74. Samek, W.; Montavon, G.; Vedaldi, A.; Hansen, L.K.; Müller, K.-R. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*; Springer Nature: Berlin/Heidelberg, Germany, 2019; Volume 11700.
75. Hamidi, F.; Scheuerman, M.K.; Branham, S.M. Gender Recognition or Gender Reductionism? The Social Implications of Embedded Gender Recognition Systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal, QC, Canada, 21–26 April 2018; pp. 1–13.
76. Miller, T. Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artif. Intell.* **2019**, *267*, 1–38. [\[CrossRef\]](#)
77. Parasuraman, R.; Riley, V. Humans and Automation: Use, Misuse, Disuse, Abuse. *Hum. Factors J. Hum. Factors Ergon. Soc.* **1997**, *39*, 230–253. [\[CrossRef\]](#)
78. Dzindolet, M.T.; Peterson, S.A.; Pomranky, R.A.; Pierce, L.G.; Beck, H.P. The Role of Trust in Automation Reliance. *Int. J. Hum.-Comput. Stud.* **2003**, *58*, 697–718. [\[CrossRef\]](#)
79. Anjomshoa, S.; Najjar, A.; Calvaresi, D.; Främling, K. Explainable Agents and Robots: Results from a Systematic Literature Review. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019)*, International Foundation for Autonomous Agents and Multiagent Systems, Montreal, QC, Canada, 13–17 May 2019; pp. 1078–1088.
80. Samek, W.; Wiegand, T.; Müller, K.-R. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv* **2017**, arXiv:1708.08296.
81. Moulin, B.; Irandoust, H.; Bélanger, M.; Desbordes, G. Explanation and Argumentation Capabilities: Towards the Creation of More Persuasive Agents. *Artif. Intell. Rev.* **2002**, *17*, 169–222. [\[CrossRef\]](#)
82. Isern, D.; Moreno, A. A Systematic Literature Review of Agents Applied in Healthcare. *J. Med. Syst.* **2016**, *40*, 43. [\[CrossRef\]](#)
83. Belpaeme, T.; Kennedy, J.; Ramachandran, A.; Scassellati, B.; Tanaka, F. Social Robots for Education: A Review. *Sci. Robot.* **2018**, *3*. [\[CrossRef\]](#)
84. Kumar, V.; Dixit, A.; Javalgi, R.R.G.; Dass, M. Research Framework, Strategies, and Applications of Intelligent Agent Technologies (IATs) In Marketing. *J. Acad. Mark. Sci.* **2016**, *44*, 24–45. [\[CrossRef\]](#)
85. Langley, P.; Meadows, B.; Sridharan, M.; Choi, D. Explainable agency for intelligent autonomous systems. In *Proceedings of the Twenty-Ninth IAAI Conference*, San Francisco, CA, USA, 6–9 February 2017; pp. 4762–4763.
86. Dennett, D.C. Three Kinds of Intentional Psychology. In *Reduction, Time and Reality*; Cambridge University Press: Cambridge, UK, 1981; pp. 37–61.
87. Dennett, D.C. *The Intentional Stance*; MIT Press: Cambridge, MA, USA, 1989.
88. Nass, C.; Moon, Y.; Carney, P. Are People Polite to Computers? Responses to Computer-Based Interviewing Systems. *J. Appl. Soc. Psychol.* **1999**, *29*, 1093–1109. [\[CrossRef\]](#)



89. De Graaf, M.M.; Malle, B.F. How people explain action (and autonomous intelligent systems should too). In Proceedings of the 2017 AAAI Fall Symposium Series, Arlington, VA, USA, 9–11 November 2017.
90. Malle, B.F. How People Explain Behavior: A New Theoretical Framework. *Pers. Soc. Psychol. Rev.* **1999**, *3*, 23–48. [\[CrossRef\]](#)
91. Malle, B. Folk Theory of Mind: Conceptual Foundations of Social Cognition. *APA* **2012**. [\[CrossRef\]](#)
92. Harbers, M.; Van den Bosch, K.; Meyer, J.-J. Modeling Agents with a Theory of Mind: Theory–theory Versus Simulation Theory. *Web Intell. Agent Syst. Int. J.* **2012**, *10*, 331–343. [\[CrossRef\]](#)
93. Hilton, D.J.; McClure, J.; Sutton, R.M. Selecting Explanations from Causal Chains: Do Statistical Principles Explain Preferences for Voluntary Causes? *Eur. J. Soc. Psychol.* **2009**, *40*, 383–400. [\[CrossRef\]](#)
94. Harbers, M.; van den Bosch, K.; Meyer, J.-J. Design and Evaluation of Explainable BDI Agents. In Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Toronto, ON, Canada, 31 August–3 September 2010; Volume 2, pp. 125–132.
95. Kaptein, F.; Broekens, J.; Hindriks, K.; Neerincx, M. Personalised Self-Explanation by Robots: The Role of Goals Versus Beliefs in Robot-Action Explanation for Children and Adults. In Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Lisbon, Portugal, 28 August–1 September 2017; pp. 676–682.
96. Wheeler, S.C.; Petty, R.E.; Bizer, G.Y. Self-schema matching and attitude change: Situational and dispositional determinants of message elaboration. *J. Consum. Res.* **2005**, *31*, 787–797. [\[CrossRef\]](#)
97. Hagger, M.S.; Chatzisarantis, N.L.D.; Harris, J. From Psychological Need Satisfaction to Intentional Behavior: Testing a Motivational Sequence in Two Behavioral Contexts. *Pers. Soc. Psychol. Bull.* **2006**, *32*, 131–148. [\[CrossRef\]](#)
98. Bickmore, T.; Gruber, A. Relational Agents in Clinical Psychiatry. *Harv. Rev. Psychiatr.* **2010**, *18*, 119–130. [\[CrossRef\]](#) [\[PubMed\]](#)
99. Lisetti, C.L.; Wagner, E. Mental Health Promotion with Animated Characters: Exploring Issues and Potential. In Proceedings of the AAAI Spring Symposium: Emotion, Personality, and Social Behavior, Stanford, CA, USA, 26–28 March 2008; pp. 72–79.
100. Richards, D.; Caldwell, P. Building a Working Alliance with a Knowledge Based System Through an Embodied Conversational Agent. In *Knowledge Management and Acquisition for Intelligent Systems*; Springer International Publishing: Cham, Switzerland, 2016; pp. 213–227.
101. Bickmore, T.W.; Schulman, D.; Sidner, C.L. A Reusable Framework for Health Counseling Dialogue Systems Based on a Behavioral Medicine Ontology. *J. Biomed. Inform.* **2011**, *44*, 183–197. [\[CrossRef\]](#) [\[PubMed\]](#)
102. Lisetti, C.; Amini, R.; Yasavur, U.; Rische, N. I Can Help You Change! An Empathic Virtual Agent Delivers Behavior Change Health Interventions. *ACM Trans. Manag. Inf. Syst. (TMIS)* **2013**, *4*, 1–28. [\[CrossRef\]](#)
103. Olafsson, S.; Wallace, B.; Bickmore, T.W. Towards a Computational Framework for Automating Substance Use Counseling with Virtual Agents. In Proceedings of the AAMAS, Auckland, New Zealand, 9–13 May 2020; pp. 966–974.
104. DeVault, D.; Artstein, R.; Benn, G.; Dey, T.; Fast, E.; Gainer, A.; Georgila, K.; Gratch, J.; Hartholt, A.; Lhomme, M. SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems, Paris, France, 5–9 May 2014; pp. 1061–1068.
105. Lucas, G.M.; Gratch, J.; King, A.; Morency, L.-P. It's only a computer: Virtual humans increase willingness to disclose. *Comput. Hum. Behav.* **2014**, *37*, 94–100. [\[CrossRef\]](#)
106. Zhou, S.; Bickmore, T.; Paasche-Orlow, M.; Jack, B. Agent-user concordance and satisfaction with a virtual hospital discharge nurse. In Proceedings of the International Conference on Intelligent Virtual Agents, Boston, MA, USA, 27–29 August 2014; pp. 528–541.
107. Kang, S.-H.; Gratch, J. People like virtual counselors that highly-disclose about themselves. *Ann. Rev. Cyberther. Telemed.* **2011**, *167*, 143–148.
108. Richards, D.; Caldwell, P. Improving health outcomes sooner rather than later via an interactive website and virtual specialist. *IEEE J. Biomed. Health Inf.* **2017**, *22*, 1699–1706. [\[CrossRef\]](#) [\[PubMed\]](#)
109. Bickmore, T.; Picard, R. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **2005**, *12*, 293–327. [\[CrossRef\]](#)
110. Bickmore, T.; Pfeifer, L. Relational agents for antipsychotic medication adherence. In Proceedings of the CHI'08 Workshop on Technology in Mental Health, Florence, Italy, 6 April 2008.
111. Bickmore, T.W.; Pfeifer, L.M.; Jack, B.W. Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 4–9 April 2009; pp. 1265–1274.
112. Schulman, D.; Bickmore, T. Persuading users through counseling dialogue with a conversational agent. In Proceedings of the 4th International Conference on Persuasive Technology, Claremont, CA, USA, 26–29 April 2009; pp. 1–8.
113. Bickmore, T.; Gruber, A.; Picard, R. Establishing the Computer–patient Working Alliance in Automated Health Behavior Change Interventions. *Patient Educ. Couns.* **2005**, *59*, 21–30. [\[CrossRef\]](#) [\[PubMed\]](#)
114. Prochaska, J.O.; Velicer, W.F. The Transtheoretical Model of Health Behavior Change. *Am. J. Health Promot.* **1997**, *12*, 38–48. [\[CrossRef\]](#) [\[PubMed\]](#)
115. Bickmore, T.W.; Schulman, D.; Sidner, C. Automated interventions for multiple health behaviors using conversational agents. *Patient Educ. Couns.* **2013**, *92*, 142–148. [\[CrossRef\]](#)
116. De Rosis, F.; Nicole, N.; Valeria, C.; Addolorata, C.; Berardina, D.C. User Modeling and Adaptation in Health Promotion Dialogs with an Animated Character. *J. Biomed. Inf.* **2006**, *39*, 514–531. [\[CrossRef\]](#)



117. Lisetti, C.L. 10 advantages of using avatars in patient-centered computer-based interventions for behavior change. *SIGHIT Record* **2012**, *2*, 28. [\[CrossRef\]](#)
118. Lisetti, C.L.; Yasavur, U.; de Leon, C.; Amini, R.; Visser, U.; Rishe, N. Building an On-Demand Avatar-Based Health Intervention for Behavior Change. In Proceedings of the Twenty-Fifth International FLAIRS Conference, Marco Island, FL, USA, 23–25 May 2012.
119. Amini, R.; Lisetti, C.; Yasavur, U.; Rishe, N. On-Demand Virtual Health Counselor for Delivering Behavior-Change Health Interventions. In Proceedings of the 2013 IEEE International Conference on Healthcare Informatics, Philadelphia, PA, USA, 9–11 September 2013; pp. 46–55. [\[CrossRef\]](#)
120. Murali, P.; Shamekhi, A.; Parmar, D.; Bickmore, T. Argumentation is More Important than Appearance for Designing Culturally Tailored Virtual Agents. In Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, Auckland, New Zealand, 9–13 May 2020; pp. 1940–1942.
121. Roitman, H.; Messika, Y.; Tsimerman, Y.; Maman, Y. Increasing patient safety using explanation-driven personalized content recommendation. In Proceedings of the 1st ACM International Health Informatics Symposium, Arlington, VA, USA, 11–12 November 2010; pp. 430–434.
122. Aikins, J.S.; Kunz, J.C.; Shortliffe, E.H.; Fallat, R.J. Puff: An Expert System for Interpretation of Pulmonary Function Data. *Comput. Biomed. Res.* **1983**, *16*, 199–208. [\[CrossRef\]](#)
123. Clancey, W. The Epistemology of a Rule-Based Expert System—A Framework for Explanation. *Artif. Intell.* **1983**, *20*, 215–251. [\[CrossRef\]](#)
124. Darlington, K. Designing for Explanation in Health Care Applications of Expert Systems. *Sage Open* **2011**, *1*, 2158244011408618. [\[CrossRef\]](#)
125. Wendt, T.; Knaup-Gregori, P.; Winter, A. Decision support in medicine: A survey of problems of user acceptance. *Med. Infobahn Eur.* **2000**, *77*, 852–856.
126. Gage, B.F.; Waterman, A.D.; Shannon, W.; Boechler, M.; Rich, M.W.; Radford, M.J. Validation of Clinical Classification Schemes for Predicting Stroke: Results from the National Registry of Atrial Fibrillation. *JAMA* **2001**, *285*, 2864–2870. [\[CrossRef\]](#)
127. Lip, G.Y.; Nieuwlaat, R.; Pisters, R.; Lane, D.A.; Crijns, H.J. Refining Clinical Risk Stratification for Predicting Stroke and Thromboembolism in Atrial Fibrillation Using a Novel Risk Factor-Based Approach. *Chest* **2010**, *137*, 263–272. [\[CrossRef\]](#) [\[PubMed\]](#)
128. Letham, B.; Rudin, C.; McCormick, T.H.; Madigan, D. Interpretable Classifiers Using Rules and Bayesian Analysis: Building a Better Stroke Prediction Model. *Ann. Appl. Stat.* **2015**, *9*, 1350–1371. [\[CrossRef\]](#)
129. Keil, F.C. Explanation and understanding. *Annu. Rev. Psychol.* **2006**, *57*, 227–254. [\[CrossRef\]](#) [\[PubMed\]](#)
130. Abdulrahman, A.; Richards, D.; Bilgin, A.A. Reason Explanation for Encouraging Behaviour Change Intention. In Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems, London, UK, 3–7 May 2021; pp. 68–77.
131. Abdulrahman, A.; Richards, D.; Ranjbartabar, H.; Mascarenhas, S. Verbal empathy and explanation to encourage behaviour change intention. *J. Multimodal User Interfaces* **2021**, *15*, 189–199. [\[CrossRef\]](#)
132. Hillier, L. *Exploring the Nature of the Therapeutic Alliance in Technology-Based Interventions for Mental Health Problems*; Lancaster University: Lancaster, UK, 2018.
133. Tremain, H.; McEnery, C.; Fletcher, K.; Murray, G. The therapeutic alliance in digital mental health interventions for serious mental illnesses: Narrative review. *JMIR Ment. Health* **2020**, *7*, e17204. [\[CrossRef\]](#) [\[PubMed\]](#)
134. Holter, M.T.; Johansen, A.; Brendryen, H. How a fully automated eHealth program simulates three therapeutic processes: A case study. *J. Med. Internet Res.* **2016**, *18*, e176. [\[CrossRef\]](#)
135. Bachelor, A. Clients' and therapists' views of the therapeutic alliance: Similarities, differences and relationship to therapy outcome. *Clin. Psychol. Psychother.* **2013**, *20*, 118–135. [\[CrossRef\]](#)
136. Mackie, C.; Dunn, N.; MacLean, S.; Testa, V.; Heisel, M.; Hatcher, S. A qualitative study of a blended therapy using problem solving therapy with a customised smartphone app in men who present to hospital with intentional self-harm. *Evid.-Based Ment. Health* **2017**, *20*, 118–122. [\[CrossRef\]](#)
137. Mayer, R.C.; Davis, J.H.; Schoorman, F.D. An integrative model of organizational trust. *Acad. Manag. Rev.* **1995**, *20*, 709–734. [\[CrossRef\]](#)
138. Berry, D.C.; Butler, L.T.; De Rosis, F. Evaluating a realistic agent in an advice-giving task. *Int. J. Hum.-Comput. Stud.* **2005**, *63*, 304–327. [\[CrossRef\]](#)
139. Morency, L.-P.; Stratou, G.; DeVault, D.; Hartholt, A.; Lhommet, M.; Lucas, G.; Morbini, F.; Georgila, K.; Scherer, S.; Gratch, J. SimSensei Demonstration: A Perceptive Virtual Human Interviewer for Healthcare Applications. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015.
140. Bailenson, J.N.; Yee, N. Digital Chameleons: Automatic Assimilation of Nonverbal Gestures in Immersive Virtual Environments. *Psychol. Sci.* **2005**, *16*, 814–819. [\[CrossRef\]](#)
141. Herrera, F.; Bailenson, J.; Weisz, E.; Ogle, E.; Zaki, J. Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PLoS ONE* **2018**, *13*, e0204494.
142. Leite, I. Using adaptive empathic responses to improve long-term interaction with social robots. In *User Modeling, Adaption and Personalization*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 446–449.

- 
143. Leite, I.; Mascarenhas, S.; Pereira, A.; Martinho, C.; Prada, R.; Paiva, A. "Why Can't We Be Friends?" An Empathic Game Companion for Long-Term Interaction. In Proceedings of the Intelligent Virtual Agents, Philadelphia, PA, USA, 20–22 September 2010; pp. 315–321.
  144. Bickmore, T.; Schulman, D.; Yin, L. Maintaining engagement in long-term interventions with relational agents. *Appl. Artif. Intell.* **2010**, *24*, 648–666. [[CrossRef](#)]
  145. Cole-Lewis, H.; Ezeanochie, N.; Turgiss, J. Understanding health behavior technology engagement: Pathway to measuring digital behavior change interventions. *JMIR Form. Res.* **2019**, *3*, e14052. [[CrossRef](#)]
  146. Huber, M.; van Vliet, M.; Giezenberg, M.; Winkens, B.; Heerkens, Y.; Dagnelie, P.; Knottnerus, J. Towards a 'patient-centred' operationalisation of the new dynamic concept of health: A mixed methods study. *BMJ Open* **2016**, *6*, e010091. [[CrossRef](#)]
  147. Klaassen, R.; Bul, K.; Op den Akker, R.; Van der Burg, G.J.; Kato, P.M.; Di Bitonto, P. Design and evaluation of a pervasive coaching and gamification platform for young diabetes patients. *Sensors* **2018**, *18*, 402. [[CrossRef](#)]
  148. Op den Akker, H.; op den Akker, R.; Beinema, T.; Banos, O.; Heylen, D.; Bedsted, B.; Pease, A.; Pelachaud, C.; Salcedo, V.T.; Kyriazakos, S. Council of coaches a novel holistic behavior change coaching approach. In Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health, ICT4AWE, Funchal, Portugal, 22–23 March 2018; pp. 219–226.
  149. Hurmuz, M.Z.M.; Jansen-Kosterink, S.M.; Op den Akker, H.; Hermens, H.J. User experience and potential health effects of a conversational agent-based electronic health intervention: Protocol for an observational cohort study. *JMIR Res Protoc.* **2020**, *9*, e16641. [[CrossRef](#)]
  150. Harbers, M.; van den Bosch, K.; Meyer, J.-J.C. A Study into Preferred Explanations of Virtual Agent Behavior. In *International Workshop on Intelligent Virtual Agents*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 132–145.
  151. Porra, J.; Lacity, M.; Parks, M.S. Can Computer Based Human-Likeness Endanger Humanness?"—A Philosophical and Ethical Perspective on Digital Assistants Expressing Feelings They Can't Have. *Inf. Syst. Front.* **2019**, *22*, 533–547. [[CrossRef](#)]
  152. Hudlicka, E. Virtual affective agents and therapeutic games. In *Artificial Intelligence in Behavioral and Mental Health Care*; Elsevier: Hoboken, NJ, USA, 2016; pp. 81–115.
  153. Richards, D.; Dignum, V. Supporting and challenging learners through pedagogical agents: Addressing ethical issues through designing for values. *Br. J. Educ. Technol.* **2019**, *50*, 2885–2901. [[CrossRef](#)]
  154. Floridi, L.; Cows, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach.* **2018**, *28*, 689–707. [[CrossRef](#)]