How to Establish a Digital Therapeutic Alliance Between Chatbots and Users: The Role of Relational Cues

Authors: Mo Ran, Fang Zuozhi, Fang Jiandong, Fang Jiandong

Date: 2022-10-22T00:00:00+00:00

Abstract

In recent years, researchers have integrated the concept of "therapeutic alliance" (TA) with internet-based self-help interventions (ISIs) to address the issue of low user engagement in ISIs. This TA formed in digital environments is termed "digital therapeutic alliance" (DTA). With the rapid advancement of artificial intelligence, chatbots can simulate human guidance, making it easier to establish relationships with users compared to traditional ISIs. They can facilitate DTA development through relational cues such as friendliness, respect, listening, encouragement, sincerity, understanding, and trust, thereby offering a novel approach to addressing low user engagement. Future research could further investigate DTA from multiple perspectives, including influencing factors, technological iteration of ISIs, measurement standardization, and experimental manipulation.

Full Text

Establishing a Digital Therapeutic Alliance Between Chatbots and Users: The Role of Relational Cues

MO Ran[1], FANG Zuozhi[4], FANG Jiandong[1,2,3]

[1]Department of Psychology, Faculty of Education, Guangxi Normal University, Guilin 541006, China
[2]Guangxi University and College Key Laboratory of Cognitive Neuroscience and Applied Psychology, Guangxi Normal University, Guilin 541006, China
[3]Guangxi Ethnic Education Development Research Center, Key Research Base of Humanities and Social Sciences in Guangxi Universities, Guilin 541006, China
[4]School of Psychology, Shaanxi Normal University, Xi' an 710062, China

Abstract: To address the issue of poor user engagement in Internet-based Self-help Interventions (ISIs), researchers have recently integrated the concept of "therapeutic alliance" (TA) into digital environments, giving rise to the notion of "Digital Therapeutic Alliance" (DTA). With the rapid advancement of artificial intelligence, chatbots can simulate human guidance and are more adept at establishing relationships with users than traditional ISIs programs. DTA can be fostered through relational cues such as friendliness, respect, listening, encouragement, sincerity, understanding, and trust, offering a novel approach to solving the problem of low engagement. Future research should further explore DTA from perspectives including influencing factors, ISIs technological iteration, measurement standardization, and experimental manipulation.

Keywords: digital therapeutic alliance, chatbot, relational cues
Classification Number: B849

The feasibility and effectiveness of Internet-based Self-help Interventions (ISIs) have been widely validated [**?**, **?**, **?**, **?**, **?**, **?**], and they may serve as a powerful complement to face-to-face therapy [**?**, **?**]. However, high dropout rates and low engagement remain well-known challenges, particularly for unguided ISIs [**?**, **?**, **?**, **?**]. With the rapid development of Artificial Intelligence (AI), chatbots capable of simulating human conversation enable unguided ISIs to maintain both efficiency and cost-effectiveness after automation [**?**, **?**]. Specifically, incorporating relational cues—such as self-disclosure, sincerity, understanding, and humor—into chatbot design can meet users' cognitive and emotional needs [**?**, **?**, **?**, **?**] and establish a Digital Therapeutic Alliance (DTA) with users, thereby promoting engagement and treatment outcomes [**?**, **?**, **?**, **?**]. A review of existing research reveals that while foreign studies are abundant, they remain relatively fragmented, whereas domestic research on ISIs in China is still in its early stages [**?**, **?**, **?**, **?**, **?**]. In light of this, this article focuses on the potential influence of chatbots on DTA through relational cues in unguided ISIs, aiming to stimulate peer interest in this field and provide references for future research.

The Development of DTA

Therapeutic Alliance (TA), also known as Working Alliance (WA), refers to the quality and strength of the collaborative relationship between client and counselor to achieve therapeutic goals [**?**, **?**, **?**]. In the 1970s, Bordin [**?**, **?**] divided TA into three components—emotional bond, agreement on therapeutic tasks, and consensus on therapeutic goals—establishing the most popular definition of TA. Subsequently, Horvath and Greenberg [**?**, **?**] developed the first TA scale, the Working Alliance Inventory (WAI), based on Bordin' s definition. In recent years, as WAI has been gradually applied in digital mental health research [**?**, **?**], the term DTA has emerged. Therapeutic alliances established through intervention modalities such as email, online chat, video conferencing, and ISIs programs can all be referred to as DTA \cite{D' Alfonso et al., 2020; Henson et al., 2019; Lederman & D' Alfonso, 2021}.

Three factors may explain DTA' s development. First, the COVID-19 pandemic has accelerated societal digitalization, with concepts like Virtual Reality and the Metaverse also gaining traction [?, ?, ?]. From a philosophical creation perspective, although humans live in a predetermined universe excluded from the role of creator, they have always dreamed of transcending nature, and digital development provides humanity with opportunities for world-building [?, ?, ?]. Therefore, the digitalization of real life will become an objective trend, making human-computer relationships increasingly important. Second, ISIs are evolving toward maximizing benefits, but higher levels of automation are accompanied by problems of low user engagement and high dropout rates. Consequently, TA, which plays a crucial role in face-to-face counseling/psychotherapy, has naturally attracted researchers' attention. Third, based on Self-Determination Theory (SDT), satisfaction of the three basic needs—autonomy, competence, and relatedness—can promote the transformation of extrinsic motivation into intrinsic motivation, thereby ensuring healthy psychological development [?, ?]. Correspondingly, ISIs programs can assist users in problem-solving, enhancing their sense of autonomy and competence, which helps cultivate the emotional bond in TA. Meanwhile, "reaching consensus on goals and tasks" in TA relates to users' goal-setting in ISIs programs and the feedback they receive. Thus, DTA has theoretical support in terms of feasibility and effectiveness.

A series of studies have shown that DTA established in ISIs approaches the level of TA in face-to-face counseling/psychotherapy [?, ?, ?, ?, ?, ?]. Additionally, DTA is positively correlated with engagement [?, ?, ?, ?, ?, ?, ?], which is a key factor in improving ISIs treatment outcomes [?, ?, ?, ?, ?]. Another meta-analysis also indicated a correlation between DTA and treatment outcomes, with a moderate overall effect size, though actual research results have been mixed [?, ?], which is not unrelated to the development and selection of measurement tools. DTA measurement requires specific consideration for digital environments; simply adapting traditional WAI may fail to capture the complexity of TA in digital interventions. Researchers have gradually recognized this and begun adapting traditional WAI for digital contexts [?, ?]. For example, Berger et al. [?, ?] adapted WAI-SR for guided ISIs. Initially, Kiluk et al. [?, ?] proposed a digital adaptation of the original WAI (WAI-Tech) for measuring DTA in unguided ISIs, but it merely replaced "counselor" with "application" in the scale. With further exploration of DTA, Meyer et al. [?, ?] found in their research that TA between participants and ISIs programs was not equivalent to TA with human counselors. Therefore, they adapted the Helping Alliance Questionnaire (HAQ) to assess the extent to which participants perceived the program as helpful. In empirical studies, participants' HAQ scores at week 3 post-intervention successfully predicted treatment outcomes three months later. More recently, Berry et al. [?, ?] developed the Mobile Agnew Relationship Measure (mARM) based on the Agnew Relationship Measure (ARM), considering the characteristics of unguided ISIs. Subsequently, Henson et al. [?, ?] developed the D-WAI based on WAI-SR to specifically assess DTA in unguided ISIs. Gómez Penedo et al. [?, ?] also developed the WAI-I based on Berger [?, ?] to

better measure DTA in guided ISIs, validating the scale' s reliability in a large sample. To further adapt DTA measurement to digital intervention scenarios, D' Alfonso et al. \cite{D' Alfonso et al., 2020} combined Human-Computer Interaction (HCI) theory with TA theory based on mARM, attempting to develop a more reliable scale for assessing DTA in unguided ISIs. Currently, chatbot technology is continuously transforming the interactive experience of traditional unguided ISIs programs, providing guidance similar to humans while achieving full automation. Therefore, future DTA measurement development may need to adapt to the iterative trends of emerging AI technologies.

Relational Cues as Important Factors Influencing DTA

Currently, most ISIs are designed based on Cognitive Behavioral Therapy (CBT) and have demonstrated significant efficacy for issues including stress, depression, anxiety, smoking addiction, alcohol addiction, insomnia, and post-traumatic stress disorder [**?**, **?**]. According to Bielinski and Berger' s [**?**, **?**] classification, common types of ISIs include three categories: (1) Unguided Interventions, where no counselor is involved in the online intervention process and users only self-help through the program; (2) Guided Interventions, which combine user self-help with regular, brief online coaching (synchronous or asynchronous); and (3) Blended Interventions, which combine online interventions with face-to-face counseling/psychotherapy, using the former as an adjunct to the latter.

In ISIs contexts, TA is relatively easier to establish with counselor support. Research indicates that TA levels in guided ISIs show no significant difference from those in face-to-face therapy, and TA predicts both engagement and treatment outcomes [**?**, **?**, **?**]. Although guided ISIs generally demonstrate better overall effects than unguided ISIs [**?**, **?**], some studies have shown that in low-intensity guided ISIs, participants' emotional bond scores are low and grow slowly in the early intervention phase [**?**, **?**], with no significant differences in engagement and treatment outcomes compared to unguided ISIs [**?**, **?**], suggesting that lack of human involvement limits TA development. However, Berry et al. [**?**, **?**] noted that participants can actually develop virtual relationships with traditional unguided ISIs programs, which helps compensate for the lack of human guidance. Holter et al.' s [**?**, **?**] human-computer relationship model, built on Grounded Theory, also suggests that individuals can establish emotional bonds with unguided ISIs programs, provided that individuals' perception of the program alternates between social actors and inanimate programs. Based on this, researchers have attempted to incorporate traditionally programmed avatars into unguided ISIs programs to narrow the effectiveness gap with guided ISIs. For example, in Heim et al.' s [**?**, **?**] study, participants' emotional bond scores stabilized and developed with the addition of an avatar, correlating with improved insomnia. However, some participants expressed a desire to communicate with human counselors, which also predicted treatment efficacy. Similarly, Fenski et al. [**?**, **?**] pointed out that if avatars cannot accurately identify and appropriately respond to participants' negative emotions, they may have counterproductive

effects. Overall, avatars embedded in unguided ISIs programs show promise in establishing DTA with humans similar to that in guided ISIs, and DTA can positively influence engagement and treatment outcomes, but how to design avatars to ensure efficacy requires further discussion.

Human Cues are characteristics that computer programs possess by simulating human appearance, speech, behavior, and other conditions, enabling interacting individuals to experience feelings typically associated only with human interaction [?, ?]. The Computers as Social Actors (CASA) paradigm also suggests that humans often subconsciously respond to human cues presented by computer programs, regardless of how rudimentary these cues may be [?, ?]. To operationalize human cues, Gallen et al. [?, ?] categorized them into four types: Visual Cues (e.g., age, gender, appearance, expression, movement); Verbal Cues (e.g., text, voice, tone, speech rate); Quasi-Nonverbal Cues (e.g., emoticons); and Relational Cues (e.g., self-disclosure, understanding, humor). These cues may all impact the establishment and development of emotional bonds and overall DTA, serving as a starting point for guiding avatar design.

As mentioned above, participants often have higher emotional expectations for avatars, and whether these expectations are met also affects emotional bond development. If the emotional bond is not sufficiently strong, it may limit DTA development, leading to poor engagement and treatment outcomes. However, forming an emotional bond requires avatars to convey relational cues such as warmth, safety, and trust to participants [?, ?]. In early research, investigators found that introducing relational cues like small talk, humor, and empathy into avatar dialogue design had a greater impact on emotional bonds than on goal and task dimensions [?, ?]. In recent research, ter Stal et al. [?, ?] also noted that empathetic discourse is a key factor affecting human-computer relationships. Therefore, endowing avatars with appropriate relational cues may play an important role in developing DTA with users.

Designing Relational Cues to Promote DTA Development

If relational cues play an important role in DTA development, then how to design these cues and enable their efficient integration becomes crucial [?, ?]. At this point, chatbots based on AI Natural Language Processing (NLP) technology demonstrate advantages. They can not only present rich human cues but also continuously "learn"based on user behavioral data [?, ?] and provide personalized feedback [?, ?, ?], making them more efficient, flexible, and humanized than traditionally programmed avatars—truly more active social actors [?, ?, ?, ?, ?].

Since the birth of ELIZA, the world' s first true chatbot, in 1966 [?, ?], chatbot technology has continuously evolved and gradually integrated into digital mental health [?, ?, ?, ?]. Currently, chatbots typically function as standalone modules embedded in ISIs programs, providing assistance to users through Voice User Interface (VUI) and either replacing human counselor guidance to become a new type of unguided ISI (e.g., MYLO, Woebot) or serving as an auxiliary

function alongside human counselors (e.g., dll 心聆 “Xiaotian”). Additionally, based on response generation mechanisms, chatbots can be divided into two categories: (1) Retrieval-based, where chatbots retrieve predefined rules from a static knowledge base to generate responses; and (2) Generation-based, where chatbots dynamically generate responses through learning and reasoning mechanisms [**?**, **?**]. In terms of form, chatbots can be broadly classified into two types: (1) those with avatars, which combine chatbots with interactive embodied agents (computer-generated digital characters with human or cartoon appearances) that interact with humans through eye contact, expressions, movements, voice, and text (e.g., Replika); and (2) those that interact with humans only through voice and text (e.g., Siri, Microsoft “Xiaoice”).

In recent years, chatbot applications in ISIs have gradually increased. Studies have found that they not only promote participant engagement more effectively than traditional unguided ISI programs [**?**, **?**, **?**], but also establish TA levels with participants comparable to humans [**?**, **?**]. Nevertheless, researchers know little about why chatbots are effective. This article hypothesizes four possible reasons: First, Mind Perception Theory [**?**, **?**] suggests that individuals can perceive mental capacities in other objects and process them anthropomorphically. Therefore, the richer a chatbot’ s relational cues, the more likely it is to enhance Social Presence, creating a perception of interacting with real humans [**?**, **?**, **?**]. Meanwhile, anthropomorphic chatbots are typically more reliable and accessible than humans, and individuals interacting with them tend to gain more sense of security [**?**, **?**], thus being more inclined to cooperate [**?**, **?**]. Second, based on Reduced Social Cues (RSC) theory, during online text-based information exchange, thoughts and emotions must be converted into text to compensate for the lack of nonverbal information [**?**, **?**]. Therefore, individuals may experience Online Disinhibition Effect (ODE) during information processing, exhibiting behaviors different from face-to-face communication, including relaxation, less constraint, and more open emotional expression [**?**, **?**], which may make human-computer relationships tighter and stronger. Third, chatbots inherently possess human cues that anthropomorphically assist individuals in problem-solving, satisfying SDT principles and thereby promoting emotional bond development. Fourth, according to the Investment Model of Personal Relationships, the emotional support and valuable information provided by chatbots can continuously increase individuals’ perceived benefits and investments while reducing perceived costs and doubts, gradually building trust and enhancing engagement with ISIs programs [**?**, **?**].

In summary, chatbots equipped with relational cues and greater flexibility are more conducive to promoting rapid DTA development in unguided ISIs from cognitive and emotional perspectives, addressing the problem of low user engagement. However, although a few researchers have explored this issue, no one has yet summarized definitive and effective relational cues to guide chatbot design. For example, Rodrigues et al. [**?**, **?**] found that chatbots with only relational cues established better DTA and higher engagement than those with only visual cues. However, this study only examined differences between visual and

relational cues in DTA without testing the effects of different relational cues. To specify the role of chatbots, more critical relational cues should be explored in depth. Therefore, based on previous research, this article proposes several relational cues in ISIs that may positively influence DTA [**?**, **?**, **?**, **?**] to help chatbots develop Artificial Wisdom.

Friendliness and Respect

In face-to-face counseling, a warm, harmonious, relaxed, free, and safe conversation atmosphere and mutual respect between counselor and client are facilitative factors for TA [**?**, **?**]. Similarly, in previous ISIs studies, friendliness and respect have been considered essential foundational design elements, where the tone and manner of information delivered by ISIs programs affect intervention credibility, engagement, and effectiveness [**?**, **?**, **?**]. Although participant preferences vary, polite, respectful, friendly, humorous, and positive conversational tones are relatively more favored, while pressure, lecturing, and shaming are generally rejected [**?**, **?**, **?**]. When using chatbots, this influence may be intensified by real-time conversational interaction. For example, when a chatbot addresses participants by name and uses humor appropriately, it can enhance the friendly relationship [**?**, **?**]. The reason is that chatbots'"personality"characteristics affect the intensity of participants' emotional responses [**?**, **?**]. If participants attribute the chatbot' s interactive feedback as polite, friendly, and respectful, even knowing it is a virtual interaction, they will still treat this parasocial interaction as an intimate social interaction [**?**, **?**] and apply corresponding social norms to interactions with the chatbot. When chatbots have virtual embodiments, they present richer nonverbal information (e.g., expressions, postures, movements, lip synchronization) and can use Artificial Emotional Contagion mechanisms (e.g., imitation and emotional mirroring) to make participants more easily perceive their friendly and respectful characteristics [**?**, **?**]. For instance, in face-to-face counseling, clients can sensitively capture counselors' micro-expressions to assess their values and judgmental intentions [**?**, **?**]. In ISIs contexts, if virtual embodiments are given high polygon counts (Tris) and refined bone structures, they can not only simulate vivid macro-expressions but may also simulate positive micro-expressions to further promote an authentic and harmonious conversation atmosphere.

Listening and Encouragement

Counselors' investment in counseling work, typically including active listening and timely encouragement, is considered a predictor of TA [**?**, **?**, **?**]. In unguided ISI research, investigators have also found no significant difference in emotional catharsis effectiveness whether participants confide in real people or chatbots [**?**, **?**]. However, since ISIs programs often need to communicate with servers, response speed also affects social attractiveness when using text-only chatbots for listening and responding, with context-based dynamic responses being more effective than overly immediate or delayed ones [**?**, **?**, **?**]. Addition-

ally, if chatbots can simulate a "typing" status during dynamic responses, users can more clearly perceive appropriate pauses and hesitations, creating an impression that the chatbot is "thinking." Regarding encouragement, Chikersal et al. [?, ?] noted that in ISIs, supportive messages with more positive impact are not only shorter but also contain more positive, affirmative, and encouraging vocabulary. Based on this, chatbots can provide indirect encouragement through keyword repetition in multi-turn dialogues, enhancing users' confiding experience. When chatbots have virtual embodiments, they can supplement verbal encouragement with appropriate eye gaze, positive facial expressions, nodding, and open gestures in response to user expressions, effectively creating interpersonal interaction perception [?, ?], attributing relational cues to the chatbot [?, ?], and further enhancing listening and encouragement effects. However, encouragement may not be suitable for all groups [?, ?]. Therefore, chatbots should first identify which populations are more susceptible to encouragement' s positive effects before responding.

Sincerity and Understanding

Counseling outcomes often depend on the quality of the counselor-client relationship—when counselors are understanding, genuine, and provide unconditional positive regard, outcomes are better [?, ?]. In unguided ISI contexts, if chatbots frequently attend to participants' words and inquire with honest, humble attitudes, participants give more positive evaluations [?, ?], which not only increases participants' sense of involvement but also helps chatbots "learn" new concepts. Additionally, if chatbots proactively disclose their "personal information" to participants, it may also make participants perceive their "sincerity," leading to more self-disclosure [?, ?]. Regarding understanding, empathy techniques are commonly used in traditional counseling. In ISIs contexts, chatbots trained on real counseling corpora can also possess paraphrasing abilities to achieve a certain level of "empathy." However, paraphrasing is not mere parroting but requires extracting content using "one' s own words" plus important terms from the client' s speech. Currently, advanced Natural Language Generation models like GPT-3 can enable chatbots to make long stories short and capture essentials [?, ?]. But to infer clients'implied meanings for more advanced "empathy" while maintaining conversational coherence by considering dialogue history still requires further technological iteration. Moreover, chatbots can provide active, precise responses to users' unique needs, preferences, and emotions based on personalization technology [?, ?], enabling users to experience a unique "understanding" distinct from traditional counseling. For example, Liu-Thompkins et al. [?, ?] attempted to integrate a personalized system framework into real marketing scenarios, enabling chatbots to possess perspective-taking "empathy" capabilities through preference analysis, personality assessment, and goal inference. Studies have found that introducing personalized design helps participants reach consensus with ISIs programs on therapeutic tasks and goals [?, ?], thereby strengthening engagement motivation [?, ?] and DTA [?, ?, ?, ?].

Mutual Trust

In a counselor-client relationship, clients will only engage in more self-disclosure and promote TA development if they perceive the counselor as trustworthy [?, ?, ?, ?]. Similarly, credibility is highly correlated with DTA quality, and low credibility of ISIs programs may lead to low engagement or even dropout [?, ?]. Conversely, if participants find ISIs programs credible, their willingness to continue using them [?, ?] and their expectations for being cured will be higher [?, ?]. In unguided ISIs, the anthropomorphic first impression created by chatbots affects their credibility [?, ?, ?, ?], with external stimulus characteristics being the most critical predictive factor [?, ?, ?, ?]. However, unlike traditional counseling, ISIs contexts allow users to independently design and match or assign chatbots specific races, appearances, ages, genders, personalities, and voices based on user profiles [?, ?], with iterative adjustments based on user feedback, thus offering greater flexibility. Additionally, similar to paraphrasing, emotional reflection techniques containing emotional words are also worth exploring. In early intervention stages, through deep learning-based emotion prediction technology [?, ?], chatbots can quickly build trust with users through brief emotional reflections (e.g., "I sense you are very anxious now"). As DTA levels gradually increase, chatbots can further provide users with more critical emotional feedback and clarify their emotional experiences through inquiry, prompting deeper disclosure. However, since emotions are often expressed through metaphors, similes, and examples, chatbots can flexibly use proactive questioning for confirmation and "learning" when meanings are unclear, enriching their knowledge graphs [?, ?]. Finally, not all users welcome emotional reflection, so before providing emotional reflection, chatbots need to comprehensively evaluate personal knowledge graphs, DTA levels, and the frequency and intensity of contextual emotional words to determine response timing and content, learning user preferences based on subsequent feedback.

In summary, this article synthesizes existing research findings and outlines the antecedents and consequences of DTA. Based on this, the article proposes a model (see Figure 1 [Figure 1: see original paper]) and hypothesizes that DTA has a direct impact on treatment outcomes; DTA has a direct impact on engagement; and engagement has a direct impact on treatment outcomes. Meanwhile, relational cues such as friendliness and respect, listening and encouragement, sincerity and understanding, and mutual trust may influence DTA, thereby leading to better engagement and treatment outcomes.

Future Research Directions

Further Exploration of DTA Influencing Factors

Although the utility of human cues is significant [?, ?], current chatbot research has focused primarily on verbal and visual cues, with relatively few studies comparing relational cues with other cues [?, ?, ?, ?]. However, in traditional counseling fields, researchers compare relational cues with other variables to

determine their relative contributions to treatment outcomes, while this article only covers some relational cues that may influence DTA [**?**, **?**]. Therefore, which variables play key roles remains unknown [**?**, **?**]. In chatbot-based unguided ISI programs, influencing factors beyond human cues include user experience, AI conversation level, and user expectations. First, based on Hentati et al.' s [**?**, **?**] findings, the ease of use of program User Interface (UI) significantly affects participant engagement. Therefore, this additional variable may lead researchers to incorrectly assess chatbots' effects. Second, studies have found interaction effects among human cues: when chatbots present strong visual cues (human photos), AI conversation level is unrelated to participant attitudes, but when presenting weak visual cues (bubble icons), strong AI conversation level compensates for the low anthropomorphic effect of weak visual cues. Finally, this research also indicates that identity cues set participants' expectations for chatbot performance; when chatbots are identified as human, participants have higher expectations, while low AI conversation level leads to more negative evaluations [**?**, **?**]. Therefore, more human cues are not necessarily better in chatbots; different cues have different effects and complex internal relationships. Future research should evaluate more TA facilitative factors and compare or combine other human cues and variables to explore possible interaction effects among different variables.

Further Technological Iteration of ISIs

Currently, ISIs mostly digitize traditional psychological therapies, while many achievements in computer science can promote ISIs technological iteration. First, other mature structured techniques can be combined with ISIs programs to further systematize them. For example, using Persuasive System Design (PSD) to construct ISIs programs can increase engagement through more support, reminders, and scheduling [**?**, **?**]. Additionally, using Motivational Interviewing (MI), a structured conversational technique, can enhance users' motivation for change [**?**, **?**], thereby quickly and effectively improving their engagement [**?**, **?**]. Second, more advanced algorithmic models can be used to further improve ISIs program performance. For example, innovative BERT (Bidirectional Encoder Representation from Transformers) or GPT-3 models can replace the labor-intensive and less scalable LIWC (Linguistic Inquiry and Word Count) model [**?**, **?**]. This way, chatbots can not only dynamically assess users' emotions and DTA levels but also greatly enhance their emotion recognition/interaction capabilities [**?**, **?**]. However, when chatbot response generation becomes more flexible and creative, the uncertainty of generated content is a double-edged sword. Therefore, developing hybrid chatbots that combine retrieval-based and generation-based approaches in counseling contexts may be more beneficial for practical application [**?**, **?**]. Finally, future research can apply multimodal technology combining NLP and Computational Vision (CV) in ISIs. For example, through deep learning models, comprehensive analysis of voice, tone, speech rate, macro-expressions, micro-expressions, body movements, pupil dilation, and other factors can further improve chatbots'

accuracy in inferring user intentions and emotions [?, ?, ?, ?, ?, ?], providing interaction forms such as text, images, options, and voice to adapt to different groups' habits. Additionally, Virtual Reality (VR) technology can be combined to bridge the gap between virtual and real interaction, strengthening immersion and social presence [?, ?, ?].

Developing DTA Measurement Tools Adapted for ISIs and Reporting with Objective Data

Effective measurement tools are important conditions for advancing field research development, but currently, researchers have little consensus on how to measure DTA [?, ?]. For example, researchers either directly use the WAI scale or make minimal adjustments to it [?, ?], but simply replacing "counselor" with "application" may be biased. On one hand, positioning ISIs programs as human may generate higher expectations and raise evaluation standards [?, ?]. On the other hand, important factors in face-to-face therapy may not be equally important in ISIs, which have their own particularities and complexities [?, ?]. In chatbot-based ISIs, researchers must consider not only program interactive experience but also evaluate human cues within them. Therefore, future research should design specialized DTA scales for ISIs contexts and intervention forms based on traditional TA theory while also considering HCI theories such as the Computers as Social Actors (CASA) paradigm and the Uncanny Valley Effect (UVE) [?, ?, ?] \cite{D' Alfonso et al., 2020; Heim et al., 2018}. Additionally, an increasing number of TA studies emphasize the need to more accurately identify TA establishment and rupture occurrences [?, ?], but current DTA measurement relies almost exclusively on participant self-reports [?, ?] without more objective quantitative analysis combining behavioral and physiological data. Therefore, future research can combine more detailed objective data [?, ?], model participants' acoustic features, behavioral trajectories, and text and audio-visual data, and dynamically analyze current DTA quality to monitor when chatbot-participant DTA is established or ruptured, providing better guidance for chatbot behavioral decisions. Finally, researchers can also comprehensively evaluate quantitative data and subjective data from counselors and observers to strengthen research result rigor.

Attention to New Issues in Different Therapies and Populations in ISIs Regarding DTA

Currently, ISIs programs used in DTA research are mostly designed based on CBT. Although online CBT' s feasibility and effectiveness have been validated [?, ?, ?], some groups have not fully benefited from it [?, ?, ?]. Therefore, more alternative therapies should be developed and tested in ISIs. For example, Mindfulness-based Interventions (MBIs) are considered effective alternatives to CBT [?, ?]. Studies have shown that participants' TA scores in MBIs are higher than in CBT [?, ?], and state mindfulness is also highly correlated with TA [?, ?]. However, research on the relationship between MBIs and TA remains scarce,

and similar evidence in ISIs environments is even more lacking. Therefore, future DTA research can use MBIs such as Mindfulness-based Stress Reduction (MBSR), Mindfulness-based Cognitive Therapy (MBCT), and Acceptance and Commitment Therapy (ACT), and attempt to simulate mindfulness coaches with chatbots to optimize existing online MBIs experiences. Additionally, DTA levels established by ISIs based on a single therapy may differ across different psychological problems (e.g., depression, anxiety, phobia, addiction) and populations (e.g., adolescents, adults, elderly, or males, females), but existing research rarely discusses this [**?**, **?**, **?**, **?**]. Therefore, future research examining DTA relationships with specific symptoms can also divide participants into more subgroups to test result differences among different characteristic populations, thereby deepening understanding of DTA mechanisms.

Author Contributions:
MO Ran: Proposed the topic and design, conducted literature search, organization, and summary, drafted and revised the final manuscript;
FANG Zuozhi: Proposed important viewpoints, conducted literature search and organization, revised the final manuscript;
FANG Jiandong: Revised the final manuscript.

Note: Figure translations are in progress. See original paper for figures.