Let's chat about sex, baby!

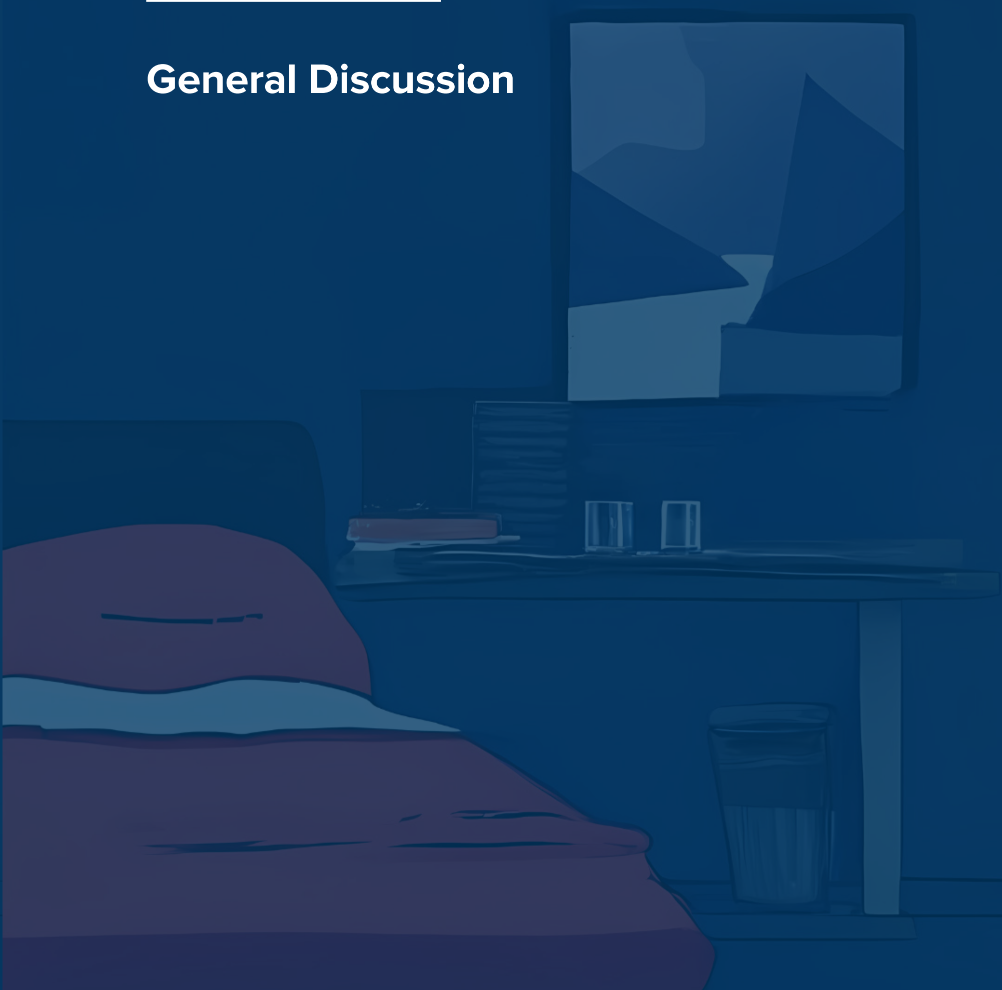
Leveraging chatbots for sexual health promotion

Balaji, D.

Publication date
2025

CHAPTER 6

General Discussion



Research Overview

The rates of sexually-transmitted infections (STIs) are increasing in young people, and young couples that have just entered a new, exclusive relationship are particularly susceptible as they often decide to quickly forgo condom use. Conversational agents (CAs), also referred to as chatbots, have the potential to be an effective solution to delivering digital sexual health interventions. This dissertation therefore set out to identify user acceptance factors relevant for sexual health CAs (Chapter 2), and to co-design and qualitatively evaluate two chatbots which address individuals (*Mish* (Chapter 3)) and couples (*MishSync* (Chapters 4 and 5)) as users. In this chapter, we summarize the findings related to the research questions proposed at the start of the dissertation, followed by theoretical and practical considerations, and end with the most important limitations and a general conclusion.

Summary of Findings

RQ1: What are the characteristics of conversational agents used for sexual health promotion, and what evidence exists regarding their effectiveness and acceptance?

To gauge their characteristics, effects and acceptance as a starting point for the design of our two chatbots, Chapter 2 consisted of a systematic review and meta-analysis regarding the effectiveness and user acceptability of current conversational agents that address one or more aspects of sexual health promotion.

Current CAs for sexual health promotion most commonly addressed treatment adherence, with a larger focus on HIV treatment (i.e. ART or antiretroviral therapy) rather than HIV prevention (i.e. PrEP or pre-exposure prophylaxis). In contrast, other preventive behaviors such as STI testing, HPV vaccine uptake, and safe sex practices such as condom use were less often addressed. Over half of the studies did not use an underlying theoretical framework; out of those that did, the most commonly used frameworks were Social Cognitive Theory (Bandura, 1999), Information-Motivation-Behavioral Skills model (Fisher & Fisher, 2009), and Motivational Interviewing (MI: Miller & Rollnick, 2012). Almost all CAs were delivered via short-messaging-service (SMS) in the case of mobile delivery, and through websites or desktop applications in the case of computer delivery; this is in contrast to the few that leveraged newer technologies such as smartphone apps and instant messaging platforms. Just over half of the CAs allowed multi-turn interactions (i.e. either the CA or the user was able to respond more than once within the same interaction), and the remainder used one-turn-only interactions largely as medication reminder systems.

CAs for sexual health promotion were found to be effective for improving ART adherence based on self-report measures but not objective measures, while a small number of studies showed that CAs to improve PrEP adherence were effective even on the basis of objective measures. Similarly positive outcomes were observed for condom use, HPV vaccination uptake, and STI testing.

Based on the factors present in the DeLone and McLean Information System Success Model (DeLone & McLean, 1992), the CAs were found to be acceptable in terms of ease of use, perceived usefulness, satisfaction, and intent to reuse. Their benefits included acting as reminder systems, and providing on-demand emotional and informational support anonymously. However, they fell short in terms of information quality, where participants wanted additional, personalized, and clearer content in an appropriate language style. The study also identified five additional constructs that may play an important role for user acceptance within the sexual health domain: trust, privacy, social presence, comfort, and emotional support. CAs tended to adequately address trust, privacy, comfort and emotional support, while the degree to which social presence is desirable was unclear. Based on these findings, we developed our own user acceptance model for conversational agents addressing sexual health, which subsequently guided the design and evaluation of Mish and MishSync in the remainder of the dissertation.

RQ2: How do young people perceive a sexual health counselling chatbot (Mish) that has been designed for individual use to promote condom use in new, exclusive relationships?

Mish was designed in a top-down manner, and delivered a single session of motivational interviewing (MI), a client-centered behavior change approach that emphasizes evocation i.e. encouraging the client to express themselves (Miller & Rollnick, 2012). In Chapter 3, we conducted a qualitative evaluation of Mish, focusing on participants' willingness to self-disclose relevant, personal information (e.g. demographics, sexual behaviors, attitudes and feelings about condom use), and how the use of basic MI strategies, through comfort and therapeutic alliance as facilitators, can promote participants' disclosures.

The qualitative interviews showed that efforts to translate motivational interviewing from a human-human to a human-chatbot interaction were successful in the sexual health context. Four out of the five MI strategies selected for the study (*affirmations, reflections, autonomy support*, and *information provision protocol*) were perceived both positively and as intended (based on the work of Miller and Rollnick (2012)). Consistent with S. Park et al. (2019), the fifth strategy i.e. *open questions* was only associated with less-positive perceptions due to the chatbot's technical inability to consistently and correctly interpret participants' free text input in response to those questions.

Participants did not experience a strong therapeutic alliance with the chatbot, and this can be explained by the need to improve the chatbot design, but also by the operation of machine schemas. Specifically, most participants only identified with certain aspects of the *bond (i.e. quality of interpersonal bond between user and chatbot)* dimension of the therapeutic alliance, and to a much lesser extent, the *goal* dimension; as such, many participants did not experience the *goal (i.e. agreement on goals of the interaction)* or *task (i.e. agreement that the tasks performed addressed their problems)* dimensions.

Consistent with Gambino et al. (2020), the majority of participants' comments reflected the presence of machine schemas, or specific beliefs and expectations about what machines can and should (not) do in different situations. In this study, participants applied machine schemas regarding the chatbot's capacity for social and affective behaviors. Specifically, participants felt that the chatbot embodied certain aspects of the *bond* dimension (friendly, helpful, respectful), but not others (caring, empathic) which were instead found to be uniquely human, and therefore ingenuine and not believable coming from a chatbot. They also perceived collaboration with a chatbot (*goal*) as being characterized by increased interactivity and personalization, rather than a social activity during which mutual agreements are iteratively negotiated. Importantly, these machine schemas appear to interfere with the formation of a strong therapeutic alliance with chatbots, which we know to be important for ensuring therapeutic outcomes in a health context (Farber, 2003), motivating the further study of machine-specific therapeutic alliance (DTA; D'Alfonso et al., 2020). On the other hand, and echoing the findings of Al-Natour et al. (2021), participants' comments suggest that the *goal* and *task* dimensions of therapeutic alliance may be adequately addressed through improved chatbot design, for example, by allowing more active participation and communicating the chatbot's goals and intents more clearly (*goal*), as well as support that is helpful and personalized to the user's context (*task*).

Chatbots have several attributes that act as facilitators of self-disclosure, but these may not outweigh the ultimate benefits of self-disclosing to a human entity in this context. Participants perceived chatbots to be non-judgmental, anonymous, accessible, objective, and the most comfortable setting to discuss sensitive topics. However, most participants adopted a gain-risk view of self-disclosure (Omarzu, 2000), preferring to disclose to a human entity, as they are seen as more capable of fluid and dynamic conversations that adapt to the individual's goals, providing "real" empathy and emotional support, and offering specific input on complex concerns – overall, self-disclosure with humans was seen as having more benefits despite the risks of feeling awkward or judged, and these benefits can be mapped onto the three dimensions of therapeutic alliance. This suggests that in the current context, the benefits of self-disclosure are more salient than the risks, but also that comfort and therapeutic alliance are important facilitators of self-disclosure to sexual health chatbots.

RQ3a: How can we design a couples-based chatbot to promote condom use in new, young couples?

Chapter 4 describes a two-stepped user-centered design approach taken alongside young people (target group) and sexual health counselors (experts) to develop MishSync, a couples-based chatbot. First, an additional question was pre-emptively added to the interview guide used in the study conducted in Chapter 3 to understand young people's preferences for a couples-based chatbot. Second, we conducted a co-design workshop with sexual health counselors. In addition to the design guidelines that emerged at each step, the workshop produced a paper-and-pen prototype that informed the final chatbot design of MishSync. Overall, both young people and experts overlapped significantly in their preferences for how MishSync should be designed – this is summarized below.

Four design guidelines emerged from the interviews with young people. (1) Young people want a chatbot that addresses sexual health holistically (i.e. addressing a range of topics, including relationship dynamics and pleasure), and (2) disliked the idea of the chatbot interfering directly with their sex lives towards explicit behavior change. (3) The main goal of the chatbot should be to improve intra-couple communication, but it is important that the chatbot does not speak on behalf of, or mediate between, the two individuals as participants strongly believed that they should ultimately be talking to each other and not to a chatbot. (4) They also expressed concerns about motivating their partner to use a couples chatbot with them given the existing discomfort of being in a new relationship.

Three design guidelines emerged from the workshop with experts. (1) The counselors believed the chatbot should take a holistic, sex-positive and educational approach to content, including topics such as communication, boundaries, and pleasure. (2) It was important that the chatbot focuses on providing basic support for healthy relationship functioning; it was not recommended that the chatbot interfere in the couple's lives, conduct couples' counselling nor provide complex support as this was considered outside of the chatbot's scope of acceptable behavior; that said, the use of basic motivational interviewing techniques was encouraged. (3) The experts agreed that it was important to present the content in a way that is engaging and interactive, for example, using stories, scenarios, and gamification.

RQ3b: How do new, young couples perceive a chatbot (MishSync) that has been designed for couple use to promote condom use?

MishSync was designed to engage new, young couples in a triadic, i.e., three-way group chat, and included three gamified activities: icebreakers (quiz), choices (interactive text adventure), and mythbusters (quiz). Given the novelty of the concept, Chapter 5 describes a holistic qualitative evaluation of MishSync with real couples through the wizard-of-oz method (i.e. a human pretends to be the chatbot without the participant being aware of this fact) (Dow et al., 2005).

Couples perceived several specific benefits of triadic chatbot interactions but gamification may be necessary to preserve engagement. Almost all participants were very positive about the three-way chat, and found the interaction fun, engaging and interesting. Importantly, it was considered very useful for learning more about the partner, communicating their own perspective, and aligning on values, opinions and knowledge regarding multiple relevant topics. The findings illustrated the potential of couples-based chatbot interactions in a context outside of couples' counselling and relationship health (Troitskaya & Batkhina, 2022; Yuksel & Kocaballi, 2022; Zheng et al., 2021). The text adventure and quiz were highly enjoyed by participants because these gamified elements added unpredictability and challenge, and helped participants in multiple ways i.e. encouraging perspective taking, increasing knowledge in different topics, and facilitating intra-couple communication.

Ultimately, participants may use MishSync, but in a different way than intended. Many participants said they would use the chatbot *if* it served a different purpose (e.g. conflict resolution, relationship troubleshooting, compiling information online), or could be used alone (e.g. for emotional support); that is, these participants were not inclined to use the chatbot again in its current form. Those that did not want to use the chatbot again cited several reasons, including preferring face-to-face communication, the challenge of discussing chatbot use with partner, the existence of well-established alternatives (e.g. ChatGPT and the Internet), and being unable to see how the chatbot could address interpersonal concerns, especially if they feel that they already communicate well with their partner.

Three key considerations emerged that increase the complexity of couples-based (i.e. triadic) interactions. **First**, the willingness to self-disclose openly was determined not only by chatbot characteristics, but also by the partner's presence, in that having the partner in the same conversation made some participants more willing to share, while making others more hesitant to do so. **Second**, disclosing the chatbot's identity led participants appears to have triggered *machine schemas,* or certain expectations about what machines can and should (not) do. In turn, these likely interfered with how participants experienced a therapeutic alliance with the chatbot – in this study, we identified two examples of when this occurred. Participants did not experience a *bond* with the chatbot in its capacity to be *caring, supportive* and *understanding,* attributing this to the fact that these are adjectives that apply to humans but cannot be used to describe machines like chatbots. Participants' *confidence* in the CA was high for its capacity to provide factual information, but low for its capacity to provide advice on interpersonal topics (i.e. relationships, sex) because it lacks the necessary understanding and context to do so. **Third**, we found that individuals were more likely to differ than align in the way they perceived and prefer to interact with the chatbot. Intra-couple differences were found on two levels; either they disagreed fully, or perceived it similarly but disagreed in their underlying reasoning. Given that it is important for couples to perceive the intervention in a similar manner to encourage joint use, this asymmetry may negatively impact the extent to which a couples-based chatbot may effectively be used.

Theoretical Implications

Drivers of User Acceptance

User acceptance of sexual health CAs is context-specific. In Chapter 2, we used the DeLone and McLean Information System Success Model (DeLone & McLean, 1992) to identify a set of user acceptance constructs that guided our evaluations of Mish and MishSync, which additionally included key communication constructs not present in the original model (privacy, trust, social presence, emotional support, and comfort). In both Chapters 3 and 5, participants talked about how their experiences surrounding these additional constructs played a role in their experience and their inclination to engage with the chatbot; importantly, they raised these points even when not explicitly prompted. Importantly, while prior studies support the inclusion of privacy and trust as additional user acceptance factors for healthcare chatbots (e.g. Laumer et al., 2019), this dissertation provides preliminary qualitative evidence for the importance of social presence (in terms of humanlikeness), emotional support (in terms of therapeutic alliance), and comfort for determining user acceptance for the sexual health domain. This evidence motivates future research that empirically tests the predictive and added value of these constructs for user acceptance with sexual health CAs.

Participants may perceive the chatbot interaction positively, and still prefer a non-chatbot entity due to innovation characteristics (or there lackof). The disconnect between perceptions and usage intentions may be explained by the innovation characteristics present in the Diffusion of Innovations Theory (Dearing & Cox, 2018; Rogers, 1983). For example, in Chapter 3 (Mish), we found that participants appreciated the use of motivational interviewing strategies, found the interaction comfortable, and were willing to self-disclose to the chatbot, citing several chatbot characteristics that facilitated this process. However, participants preferred to disclose to a human entity because of *relative advantage* (i.e. the benefits of disclosing to a human outweighed those of a chatbot). In Chapter 5 (MishSync), we found that participants perceived the interaction as fun and engaging, and found the triadic chat beneficial in several ways for their relationship. However, many participants said they would only use the chatbot again on the condition that it could be used alone or served a different purpose, while a small number said they would not use it again. The underlying reasons can be mapped onto *compatibility* (i.e. the added challenge of communicating about chatbot use with partner, interfering with F2F), *observability* (i.e. immediate distrust in chatbot's capability to support interpersonal matters), and *relative advantage* (i.e. existing alternatives such as ChatGPT, Google, and the internet as a whole). In an age of rapid technological advancements and the multitude of choice, these findings are consistent with Dearing's (2021) commentary on the role that innovation characteristics can play in the diffusion of AI agents such as chatbots.

Humanlikeness, Machine Schemas, and Chatbot Perceptions

Chatbots are often designed to be humanlike, and this is usually done through the inclusion of social and/or anthropomorphic cues (Lombard & Xu, 2021). Encountering such cues during an interaction can activate social scripts that lead human users to respond in social ways (CASA; Reeves & Nass, 1996). However, our findings suggest that in addition to social scripts, humans also apply machine-specific schemas and scripts that have been developed through experience over time (Edwards et al., 2016; Gambino et al., 2020), and can influence users' perceptions and behavior towards the chatbot.

In Chapters 3 and 5, the findings suggest that disclosing the chatbot's identity led to the activation of *machine schemas*, or certain expectations of what chatbots can and should do in a specific context (Gambino et al., 2020), which appears to have influenced how users perceived specific behaviors of Mish and MishSync. Across both studies, the presence of machine schemas appears to have (a) interfered in the formation of a strong therapeutic alliance (primarily through the *bond* dimension), (b) hindered self-disclosure processes, (c) reduced confidence in the chatbot's capacity to provide advice on emotional or interpersonal matters, and (d) made it less acceptable for the chatbot to engage in explicit behavior change and couples counselling. Overall, while surface-level humanlike characteristics (e.g. friendly tone, jokes, polite behavior) were found to be highly acceptable and even desired, complex humanlike behaviors (e.g. showing empathy, perceived understanding, providing relationship advice) were more likely to be perceived as unique to humans, making it inauthentic and less desirable in chatbots (Seitz, 2024; Stein & Ohler, 2017).

Additionally, simply disclosing the chatbot's identity may be sufficient to trigger these machine schemas, altering how the same behavior is perceived when performed by a chatbot versus a human. For example, Basar and colleagues (2024) further demonstrate that if participants are unaware of the source, they tend to find language model-generated counselling responses as much more appropriate, specific, and relevant than human expert-written responses. These findings add to the body of literature that illustrates the downstream negative consequences of chatbot identity disclosure e.g. cooperation (Ishowo-Oloko et al., 2019), persuasive capacity (Castelo et al., 2019; Shi et al., 2020), trust (Mozafari et al., 2020), perceived supportiveness (Medeiros et al., 2022), and overall user acceptance (Markovitch et al., 2024; Murgia et al., 2016; R. Yu et al., 2024). Given that chatbot identity non-disclosure poses both ethical and legal considerations (European Commission, 2021), it becomes important for future research to investigate how we can mitigate the negative consequences of chatbot identity disclosures (e.g. Mozafari et al., 2021).

Ethical Consideration: Do Our Chatbots Violate User Autonomy?

This dissertation explored two chatbots as potential digital solutions to a growing public health concern. However, it could be argued that despite our good intentions, interventions designed to result in behavior change inherently run the risk of violating user autonomy. Park and colleagues (2024) found that current research on sexual health CAs tend not to address ethical concerns of working with vulnerable groups such as adolescents and young adults, an example of which is empowerment and autonomy (Cascalheira et al., 2023). To reflect on this notion, the PhD candidates of the Look Who's Talking project (that this dissertation is part of), collaborated on a theoretical paper on how chatbots for behavior change may (not) support user autonomy (Balaji et al., 2025).

Drawing from Frankfurt's theory of volition (Frankfurt, 2001), we differentiate between first- and second-order volitions, where an autonomous entity (such as an adult human) is seen as having one or more desires (or first-order volitions), while also being able to reflect and actively choose to act on a specific desire (or second-order volition). In an attempt to strike a balance between promoting positive health behaviors and supporting user autonomy, we propose that a chatbot can be equipped with the skills to probe users about the *reasons-responsiveness* (i.e. do they have good reasons?) of their second-order volitions. Importantly, this procedure is conducted even if the user has adopted, or is considering adopting an “undesirable” second-order volition e.g. not using condoms. If users are unable to provide strong reasons for adopting or continuing to do a specific behavior, this introduces an *acceptable* opportunity for the chatbot to support the user’s process of reflection.

Applying the above line of reasoning to the work conducted in this dissertation, Mish and MishSync *do* support the probing of reasoned responsivenness. This was possible because both chatbots made use of motivational interviewing (to varying degrees), and therefore included simple and short *evocation* exercises in which users are asked to express their reasons in support of a specific behavior (Miller & Rollnick, 2012). The compatibility of motivational interviewing with autonomy support is unsurprising given its roots in self-determination theory (Deci & Ryan, 1985). To ensure truly volitional, and therefore ethical, behavior change, it can be useful to further explore how chatbots can more effectively support evocation processes as a way to help users adopt their desired second-order volitions.

Practical Takeaways

The Marriage of Top-Down and Bottom-Up Design

The design phase of intervention development can focus on either theory (top-down), or working with stakeholders (bottom-up) (O’Cathain et al., 2019). In this research, we utilised both approaches for Mish (Chapter 3) and MishSync (Chapter 5) respectively, and found that the two chatbots ended up bearing little resemblance to each other, despite addressing the same target group and behavior. This was because the principles that guided chatbot design differed between the two approaches.

Consistent with the tenets of traditional approaches behavior change interventions such as intervention mapping (Bartholomew et al., 1998), Mish (Chapter 3) was designed to focus solely on one sexual health behavior i.e. condom use, and adhered to a structured approach that addressed the key determinants of the behavior. However, in Chapter 4, we found that both our target group (young people) and professionals (sexual health counselors) agreed that MishSync (Chapter 5) should address behaviors beyond condom use and take a holistic approach to sexual health. Similarly, while Mish utilised motivational interviewing to facilitate positive behavior change towards condom use, both young people and sexual health counselors agreed that chatbots should avoid directly interfering with behavior but take a more educational and engaging approach to increase awareness and empowerment.

These differences hint at the research-practice gap, and suggest the need to combine the best of both approaches. Focusing on a specific health behavior may lead to a more effective intervention but can also reduce its perceived usefulness and overall user acceptance, which can negatively impact intervention effectiveness. On the contrary, holistic approaches may be more attractive, leading to increased acceptance and usage, but may not result in the desired intervention outcomes. It may therefore be useful to use both user-centered design and theory-based approaches in conjunction (e.g., Groos et al. (2024)) for the development of digital health interventions that are more likely to meet users' needs and be effective in the real world.

Towards Sexual Health Counselling Chatbots for Couples

Earlier, we presented chatbots as being uniquely suited for the delivery of language-based strategies and approaches, such as motivational interviewing (MI), for health promotion. Through the series of studies conducted in this dissertation, we found evidence to support the potential of conducting sexual health counselling in couples through chatbots. In Chapter 3, we looked at the potential for a chatbot to carry out motivational interviewing, a counselling approach whose evocative nature relies heavily on natural language. Mish was able to conduct a 20-minute counselling session with individual users, addressing the four stages of MI (engage, focus, evoke, and plan) while effectively making use of basic MI techniques. In Chapter 5, we showed that young couples positively evaluated a triadic chatbot interaction to promote sexual health, and associated it with several benefits that can be linked to improved dyadic functioning and communication. Given the potential of chatbots to conduct MI, and the preliminary acceptability of triadic chatbot interactions in this context, future research can then pursue a couples-based motivational interviewing chatbot for sexual health promotion. Starks et al. (2018, 2020) have developed a motivational interviewing framework for couples specifically for sexual health promotion, which can serve as a strong foundation to guide future couples-based sexual health chatbots that actively promote positive behavior change.

It would be unwise to ignore the input obtained through the user-centered design activities conducted in Chapter 4, where both young people and experts expressed that the chatbot should not engage in active behavior change or counselling with the couples as this was considered out of scope for acceptable chatbot behavior. However, an autonomy-supportive approach taken by the chatbot may better support health-related dyadic functioning in a way that addresses these concerns. In line with the input received from young people and experts, the work by Starks and colleagues (2018, 2020), and the findings from couples-based relationship counselling chatbots (Troitskaya & Batkhina, 2022; Yuksel & Kocaballi, 2022; Zheng et al., 2021), couples-based chatbots are in the unqiue position of being able to facilitate dyadic functioning that can indirectly benefit couples' sexual health without directly influencing sexual behaviors. Regarding the appropriate approach, motivational interviewing seems to remain a good candidate as it was utilised in Mish (Chapter 3), and to a much lesser degree, in MishSync (Chapter 5), and participants did not express that this was unacceptable behavior; it may be the case that in contrast to other behavior change or counselling approaches, motivational interviewing may be particularly acceptable in light of its client-centered, autonomy-supportive focus. Furthermore, our ethical exploration of autonomy support in behavior change chatbots suggests that it is possible for the chatbot design to include additional mechanisms that encourage reflection without directly interfering with behavior; these are complementary to the *evocation* aspect of motivational interviewing, and serve only to identify the reason-responsiveness of both individuals' volitions, and facilitate exploration in case of ambivalent or non-existing volitions.

Design Recommendations

Technical Capabilities

Chapter 3 found that if the user is given the opportunity to use free text to express themselves (which is appreciated), it is important that the chatbot is able to accurately and consistently understand user intent.

Intervention Approach

Based on the findings from Chapter 5, we found that it is more acceptable for chatbots to provide information, question-and-answer support, and pointing to resources, while it is less acceptable for the chatbot to engage in active counselling, or providing advice on interpersonal matters.

Intervention Content

In Chapter 4, both young people and experts recommended that the chatbot take a holistic, sex-positive approach to sexual health rather than addressing any single behavior. In Chapter 5, participants confirm that they appreciated that the chatbot addressed multiple topics.

Building Rapport

Findings from Chapters 3 and 5 indicate that it is desirable for the chatbot to exhibit surface-level humanlike behaviors or relational strategies, such as being respectful and polite, and using friendly language, emotive language and punctuation, referring to user(s) by name, and making jokes. Even in the absence of an explicit behavior change goal, basic motivational interviewing techniques (e.g. autonomy support, asking for permission, and reflections) are appreciated. However, we suggest careful consideration of the chatbot exhibiting behaviors that imply the capacity for emotional intelligence, such as care, understanding, and empathy.

Supporting Therapeutic Alliance Formation

In Chapter 3, we found that the *goal* and *task* aspects of the therapeutic alliance can be better addressed through chatbot design. To address *goal,* the chatbot should allow the user to participate more actively, and communicate its own goals and intents clearly throughout the interaction. To address *task*, the chatbot should provide more personalized responses, or responses that are or seem more specific and appropriate, and therefore helpful to the user.

Limitations and Future Directions

In this section, we discuss two main limitations of this research, followed by a final reflection on avenues for future research.

Participant Recruitment

The specificity of the target group (i.e. new, young couples) made it difficult to reach desired sample sizes, which motivated us to reconsider the inclusion criteria. In Chapter 3, the inclusion criteria used to recruit the sample was intentionally made broader. That is, in addition to young individuals in a new relationship, we also included individuals that were currently single, or casually dating; while these individuals were not in a relationship at the time of recruitment, they could enter a new relationship in the near future and therefore their perspective was still considered useful. As a result, the findings are not specific to our target group, but to a slightly broader but still relevant group. In Chapter 5, the recruitment of new, young couples proved to be a greater challenge, leading us to extend the maximum relationship duration from three to six months. During the interviews, participants often expressed how they had "*already* been together for x months", despite this number always being under six. This confirmed the idea that relationships are not perceived as *new* for very long, and that the findings are once again applicable to a slightly broader but still relevant group.

In Chapters 3 and 5, people took part of their own volition, and many indicated that they were interested in conversational agents and/or sexual health. There may have been a participant bias as we likely did not evaluate the CAs sufficiently with people who may not have similarly positive attitudes. Future studies can ensure adequate sample size and representativeness by developing a proactive recruitment plan in conjunction with institutions such as high schools, medical institutions, and public health organizations.

Wizard-of-Oz Method

We used the wizard-of-oz (WoZ) method to assess user acceptance of chatbot MishSync in Chapter 5. Using the WoZ method means that the participant interacts with what appears to be a fully-functional chatbot prototype when in reality, a human operator is responding to the participant's input from behind the scenes (Dow et al., 2005). WoZ allows researchers and designers to evaluate the *design* of a technology (such as a chatbot), and understand users' needs, preferences, and attitudes without devoting extensive resources to fully implement a functional prototype of the system (E. Mitchell & Mamykina, 2021). As such, WoZ has been used in many studies to capture users' perceptions of chatbots (e.g. Bosco et al., 2024; Cai et al., 2025; Jentzsch et al., 2019; Medhi Thies et al., 2017; Teng et al., 2024).

However, the WoZ approach may have introduced response inconsistencies, errors, and delays due to human error (Schlögl et al., 2015). Additionally, the success of the WoZ is contingent on the participant being successfully deceived about the source of the interaction, as this determines the extent to which users' perceptions would reflect those of a fully-functional chatbot. We took several measures to address these concerns, including (1) developing structured, rule-based dialogue flows with explicit instructions for the wizard to follow, (2) developing a dedicated web interface to allow quick, accurate, and consistent responses across participants, and (3) addressing potential response delays by informing participants that "the language models being used by the chatbot are still in development and may respond slowly".

Based on the chat transcripts as well as the debriefing sessions conducted with participants, not a single participant guessed that they had actually been interacting with a human, and many were surprised at this fact because they truly believed they had been talking to a chatbot. Furthermore, the chatbot dialogue flows had also been discussed with the computational linguist that developed the chatbot in Chapter 3, and were developed in a way that is technically possible given adequate development time and resources; this increases the likelihood that a fully-functional prototype of MishSync would behave identically to the one participants interacted with in Chapter 5. It then becomes highly likely that the same findings regarding participants' perceptions would be observed in the case of a real chatbot. In this case, using WoZ was an appropriate method to answer RQ3b ("How do new, young couples perceive a chatbot (MishSync) that has been designed for couple use to promote condom use?"). However, the lack of a fully-functional prototype means we did not truly test the technical ability of chatbots to conduct a triadic chat with couples. The logical next step to address this concern is to incorporate participants' feedback obtained through the WoZ method, and replicate the study using a functioning chatbot prototype.

Conclusion

This dissertation delved into the potential of conversational agents (CAs) for sexual health promotion, particularly in encouraging condom use among young couples. By systematically designing and evaluating two chatbots—Mish for individual use and MishSync for couples—this research shed light on key factors influencing user acceptance of sexual health chatbots, and highlighted the role of machine-specific schemas in shaping user perceptions of chatbot interactions. Findings revealed that motivational interviewing can be successfully adapted for a chatbot-based sexual health intervention, fostering comfort and self-disclosure. Moreover, the user-centered design approach demonstrated that couples-based CAs are well-received but require distinct considerations, particularly regarding trust-building and partner dynamics. While CAs show promise for engaging young people in sexual health discussions, future research is needed to refine human-chatbot interactions for greater real-world adoption.

References

Al-Natour, S., Benbasat, I., & Cenfetelli, R. (2021). Designing Online Virtual Advisors to Encourage Customer Self-disclosure: A Theoretical Model and an Empirical Test. *Journal of Management Information Systems*, *38*(3), 798–827. https://doi.org/10.1080/07421222.2021.1962595

Balaji, D., Basar, E., He, L., & Haselager, P. (2025). Autonomy-Supportive Chatbots: Endorsing Volitional Behavior Change. *Manuscript under Review for Publication*.

Bandura, A. (1999). Social Cognitive Theory: An Agentic Perspective. *Asian Journal of Social Psychology*, 2(1), 21–41. https://doi.org/10.1111/1467-839X.00024

Bartholomew, L. K., Parcel, G. S., & Kok, G. (1998). Intervention Mapping: A Process for Developing Theory and Evidence-Based Health Education Programs. *Health Education & Behavior*, *25*(5), 545–563. https://doi.org/10.1177/109019819802500502

Basar, E., Hendrickx, I., Krahmer, E., Bruijn, G.-J., & Bosse, T. (2024). To What Extent Are Large Language Models Capable of Generating Substantial Reflections for Motivational Interviewing Counseling Chatbots? A Human Evaluation. In N. Soni, L. Flek, A. Sharma, D. Yang, S. Hooker, & H. A. Schwartz (Eds.), *Proceedings of the 1st Human-Centered Large Language Modeling Workshop* (pp. 41–52). ACL. https://doi.org/10.18653/v1/2024.hucllm-1.4

Bosco, C., Shojaei, F., Theisz, A. A., Torres, J. O., Cureton, B., Himes, A. K., Jessup, N. M., Barnes, P. A., Lu, Y., Hendrie, H. C., Hill, C. V., & Shih, P. C. (2024). Testing 3 Modalities (Voice Assistant, Chatbot, and Mobile App) to Assist Older African American and Black Adults in Seeking Information on Alzheimer Disease and Related Dementias: Wizard of Oz Usability Study. *JMIR Formative Research*, *8*(1), e60650. https://doi.org/10.2196/60650

Cai, N., Heo, J., & Yan, J. (2025). Understanding Consumer Reactions to Chatbot Service Failures: Evidence From A Wizard-Of-Oz Experiment. *Acta Psychologica*, *253*, 104707. https://doi.org/10.1016/j.actpsy.2025.104707

Cascalheira, C. J., Pugh, T. H., Hong, C., Birkett, M., Macapagal, K., & Holloway, I. W. (2023). Developing Technology-Based Interventions for Infectious Diseases: Ethical Considerations For Young Sexual And Gender Minority People. *Frontiers in Reproductive Health*, *5*, 1303218. https://doi.org/10.3389/frph.2023.1303218

Castelo, N. 1, Bos, M. W. 2, & Lehmann, D. 3 1 P. of M. (2019). Let the Machine Decide: When Consumers Trust or Distrust Algorithms. *NIM Marketing Intelligence Review,* 24–29. https://doi.org/10.2478/nimmir-2019-0012

D'Alfonso, S., Lederman, R., Bucci, S., & Berry, K. (2020). The Digital Therapeutic Alliance and Human-Computer Interaction. *JMIR Mental Health*, *7*(12), e21895. https://doi.org/10.2196/21895

Dearing, J. W., & Cox, J. G. (2018). Diffusion Of Innovations Theory, Principles, And Practice. *Health Affairs*, *37*(2), 183–190. https://doi.org/10.1377/hlthaff.2017.1104

Deci, E. L., & Ryan, R. M. (1985). Conceptualizations of Intrinsic Motivation and Self-Determination. In E. L. Deci & R. M. Ryan (Eds.), *Intrinsic Motivation and Self-Determination in Human Behavior* (pp. 11–40). Springer US. https://doi.org/10.1007/978-1-4899-2271-7_2

DeLone, W. H., & McLean, E. R. (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, *3*(1), 60–95. https://doi.org/10.1287/isre.3.1.60

Dow, S., MacIntyre, B., Lee, J., Oezbek, C., Bolter, J. D., & Gandy, M. (2005). Wizard Of Oz Support Throughout an Iterative Design Process. *IEEE Pervasive Computing*, *4*(4), 18–26. IEEE Pervasive Computing. https://doi.org/10.1109/MPRV.2005.93

Edwards, C., Edwards, A., Spence, P. R., & Westerman, D. (2016). Initial Interaction Expectations with Robots: Testing the Human-To-Human Interaction Script. *Communication Studies*, *67*(2), 227–238. https://doi.org/10.1080/10510974.2015.1121899

European Commission. (2021, April 21). Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts. https://artificialintelligenceact.eu/wp-content/uploads/2021/08/The-AI-Act.pdf

Farber, B. A. (2003). Patient Self-Disclosure: A Review of The Research. *Journal of Clinical Psychology*, 59(5), 589–600. https://doi.org/10.1002/jclp.10161

Fisher, J. D., Fisher, W. A., & Shuper, P. A. (2009). The Information-Motivation-Behavioral Skills Model of HIV Preventive Behavior. In *Emerging theories in health promotion practice and research*, 2nd ed (pp. 21–63). Jossey-Bass/Wiley.

Frankfurt, H. (2001). Freedom of the Will and the Concept of a Person. In *Agency and Responsibility*. Routledge.

Gambino, A., Fox, J., & Ratan, R. (2020). Building a Stronger CASA: Extending the Computers Are Social Actors Paradigm. *Human-Machine Communication*, *1*, 71–86. https://doi.org/10.30658/hmc.1.5

Groos, S. S., Linn, A. J., Kuiper, J. I., van Schoor, N. M., van der Velde, N., & van Weert, J. C. M. (2024). Combining User-Centered Design and Behavioral Theory To Enhance Health Technologies: A Personas-Based Approach For A Primary-Care Based Multifactorial Falls Risk Assessment Tool. *International Journal of Medical Informatics*, *186*, 105420. https://doi.org/10.1016/j.ijmedinf.2024.105420

Ishowo-Oloko, F., Bonnefon, J.-F., Soroye, Z., Crandall, J., Rahwan, I., & Rahwan, T. (2019). Behavioral Evidence for A Transparency–Efficiency Tradeoff In Human–Machine Cooperation. *Nature Machine Intelligence*, *1*(11), 517–521. https://doi.org/10.1038/s42256-019-0113-5

Jentzsch, S. F., Höhn, S., & Hochgeschwender, N. (2019). Conversational Interfaces for Explainable AI: A Human-Centered Approach. *Explainable, Transparent Autonomous Agents and Multi-Agent Systems: First International Workshop, EXTRAAMAS 2019, Montreal, QC, Canada, May 13–14, 2019, Revised Selected Papers*, 77–92. https://doi.org/10.1007/978-3-030-30391-4_5

Laumer, S., Maier, C., & Gubler, F. (2019). Chatbot Acceptance in Healthcare: Explaining User Adoption Of Conversational Agents For Disease Diagnosis. *Research Papers*. https://aisel.aisnet.org/ecis2019_rp/88

Lombard, M., & Xu, K. (2021). Social Responses to Media Technologies in The 21st Century: The Media Are Social Actors Paradigm. *Human-Machine Communication*, *2*, 29–55. https://doi.org/10.3316/INFORMIT.100016122150335

Markovitch, D. G., Stough, R. A., & Huang, D. (2024). Consumer Reactions to Chatbot Versus Human Service: An Investigation In The Role Of Outcome Valence And Perceived Empathy. *Journal of Retailing and Consumer Services*, *79*, 103847. https://doi.org/10.1016/j.jretconser.2024.103847

Medeiros, L., Bosse, T., & Gerritsen, C. (2022). Can a Chatbot Comfort Humans? Studying the Impact of a Supportive Chatbot on Users' Self-Perceived Stress. *IEEE Transactions on Human-Machine Systems*, *52*(3), 343–353. https://doi.org/10.1109/THMS.2021.3113643

Medhi Thies, I., Menon, N., Magapu, S., Subramony, M., & O'Neill, J. (2017). How Do You Want Your Chatbot? An Exploratory Wizard-of-Oz Study with Young, Urban Indians. *Human-Computer Interaction - INTERACT 2017*, 441–459. https://doi.org/10.1007/978-3-319-67744-6_28

Miller, W. R., & Rollnick, S. (2012). *Motivational Interviewing: Helping People Change*. Guilford Press.

Mitchell, E., & Mamykina, L. (2021). From the Curtain to Kansas: Conducting Wizard-of-Oz Studies in the Wild. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3411763.3443446

Mozafari, N., Weiger, W. H., & Hammerschmidt, M. (2021). Resolving the Chatbot Disclosure Dilemma: Leveraging Selective Self-Presentation to Mitigate the Negative Effect of Chatbot Disclosure. *Hawaii International Conference on System Sciences.* https://doi.org/10.24251/HICSS.2021.355

Mozafari, N., Weiger, W., & Hammerschmidt, M. (2020). The Chatbot Disclosure Dilemma: Desirable and Undesirable Effects of Disclosing the Non-Human Identity of Chatbots. *International Conference of Information Systems.* https://aisel.aisnet.org/icis2020/hci_artintel/hci_artintel/6

Murgia, A., Janssens, D., Demeyer, S., & Vasilescu, B. (2016). Among the Machines: Human-Bot Interaction on Social Q&A Websites. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 1272–1279. https://doi.org/10.1145/2851581.2892311

O'Cathain, A., Croot, L., Sworn, K., Duncan, E., Rousseau, N., Turner, K., Yardley, L., & Hoddinott, P. (2019). Taxonomy Of Approaches to Developing Interventions To Improve Health: A Systematic Methods Overview. *Pilot and Feasibility Studies*, *5*(1), Article 1. https://doi.org/10.1186/s40814-019-0425-6

Omarzu, J. (2000). A Disclosure Decision Model: Determining How and When Individuals Will Self-Disclose. *Personality and Social Psychology Review*, *4*(2), 174–185. https://doi.org/10.1207/S15327957PSPR0402_05

Park, J. K., Singh, V., & Wisniewski, P. (2024). Current Trends and Future Directions for Sexual Health Conversational Agents (CAs) for Youth: A Scoping Review. *arXiv*. http://arxiv.org/abs/2409.14226

Park, S., Choi, J., Lee, S., Oh, C., Kim, C., La, S., Lee, J., & Suh, B. (2019). Designing a Chatbot for a Brief Motivational Interview on Stress Management: Qualitative Case Study. *Journal of Medical Internet Research*, *21*(4), e12231. https://doi.org/10.2196/12231

Reeves, B., & Nass, C. I. (1996). *The Media Equation: How People Treat Computers, Television, And New Media Like Real People and Places* (pp. xiv, 305). Cambridge University Press.

Rogers, E. M. (1983). *Diffusion of Innovations* (3rd ed.). The Free Press, A Division of MacMillan Publishing Co., Inc.

Schlögl, S., Doherty, G., & Luz, S. (2015). Wizard of Oz Experimentation for Language Technology Applications: Challenges and Tools. *Interacting with Computers*, *27*(6), 592–615. https://doi.org/10.1093/iwc/iwu016

Seitz, L. (2024). Artificial Empathy in Healthcare Chatbots: Does It Feel Authentic? *Computers in Human Behavior: Artificial Humans*, *2*(1), 100067. https://doi.org/10.1016/j.chbah.2024.100067

Shi, W., Wang, X., Oh, Y. J., Zhang, J., Sahay, S., & Yu, Z. (2020). Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376843

Starks, T. J., Millar, B. M., Doyle, K. M., Bertone, P., Ohadi, J., & Parsons, J. T. (2018). Motivational Interviewing with Couples: A Theoretical Framework for Clinical Practice Illustrated In Substance Use And HIV Prevention Intervention With Gay Male Couples. *Psychology of Sexual Orientation and Gender Diversity*, *5*(4), 490–502. https://doi.org/10.1037/sgd0000297

Starks, T. J., Robles, G., Doyle, K. M., Pawson, M., Bertone, P., Millar, B. M., & Ingersoll, K. S. (2020). Motivational Interviewing with Male Couples to Reduce Substance Use And HIV Risk: Manifestations Of Partner Discord And Strategies For Facilitating Dyadic Functioning. *Psychotherapy*, *57*(1), 58–67. https://doi.org/10.1037/pst0000278

Stein, J.-P., & Ohler, P. (2017). Venturing Into the Uncanny Valley Of Mind—The Influence Of Mind Attribution On The Acceptance Of Human-Like Characters In A Virtual Reality Setting. *Cognition*, *160*, 43–50. https://doi.org/10.1016/j.cognition.2016.12.010

Teng, M., Yu, M., Jin, W., Hwang, D. W., Stitt, D., Conati, C., Carenini, G., Jarus, T., & Field, T. (2024). Human-computer interactions and compassionate healthcare: A Wizard of Oz study using a self-administered AI-assisted cognitive assessment (p. 2024.12.17.24317999). *medRxiv*. https://doi.org/10.1101/2024.12.17.24317999

Troitskaya, O., & Batkhina, A. (2022). Mobile Application for Couple Relationships: Results Of A Pilot Effectiveness Study. *Family Process*, *61*(2), 625–642. https://doi.org/10.1111/famp.12733

Yu, R., Feng, J., Wang, K., Yang, L., & Feng, J. (2024). Who Are You Talking To? How Chatbot Identity Disclosure Affects Service Satisfaction. *Journal of Travel & Tourism Marketing*, *41*(8), 1090–1106. https://doi.org/10.1080/10548408.2024.2369755

Yuksel, B., & Kocaballi, A. B. (2022). Conversational Agents to Support Couple Therapy. *Proceedings of the 34th Australian Conference on Human-Computer Interaction*, 291–297. https://doi.org/10.1145/3572921.3572922

Zheng, Q., Markazi, D. M., Tang, Y., & Huang, Y. (2021). “PocketBot Is Like a Knock-On-the-Door!”: Designing a Chatbot to Support Long-Distance Relationships. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–28. https://doi.org/10.1145/3479589