Design & UX Research for Hola: An AI-Based Companion for Youth Self-Disclosure and Self-Exploration

Yunfei Hu
y.hu4@student.tue.nl
Department of Industrial Design, Eindhoven University of Technology
5612 AZ Eindhoven, The Netherlands

Abstract

AI-based mental health chatbots offer scalable and anonymous emotional support, yet often struggle to sustain engagement and effective personalization in self-disclosure contexts. This project designs and investigates *Hola*, an AI-based daily mental health companion, examining how different chatbot response personalization mechanisms influence users' Anticipated Communication Quality (ACQ) and Communication Willingness (CW). A preliminary exploratory study with six participants compared three conditions: baseline AI, system-tailored AI, and user-tailored AI. Results show that personalization generally improves both ACQ and CW, while user-controlled adjustment enhances perceived agency and alignment with response style. However, regardless of emotional responsiveness, a lack of actionable content significantly limits perceived communication quality.

Keywords: Human–AI Interaction; Personalization; User Experience; Self-disclosureMental Well-being Support

1 Introduction

Young adults are likely facing depression, anxiety and stress as they navigate the development of autonomy and career direction, a trend that has gradually increased in recent years [10, 30, 33] . These challenges are often compounded by stigma, financial barriers and a lack of appropriate mental health support [24, 30, 37].

As traditional face-to-face counseling services struggle to meet growing demand, alternative forms of online, multimodal human-delivered psychotherapy have been explored. Platforms such as BetterHelp provide remote digital psychotherapy through multimodal online interactions and have demonstrated effectiveness in delivering mental health care at scale [3, 29].

The integration of AI has further improved the accessibility of mental health support by enabling continuous availability, reducing costs, supporting scalability in low-resource settings, and offering anonymity that helps lower stigma and encourage help-seeking behavior [42]. AI-supported human therapy platforms such as Talkspace attempt to match users with suitable human therapists while providing ongoing support, and prior studies indicate that users who actively contribute more written input tend to experience better recovery outcomes [18, 41]. Fully AI-based mental health products provide even greater accessibility and offer distinct advantages in anonymity, which has been shown to reduce fear of judgment [2, 13] . Existing products such as Woebot and Wysa incorporate established psychological approaches, including Cognitive Behavioral Therapy (CBT), to guide users in short-term emotional expression and stress relief [6, 7]. However, these systems face limitations in long-term use, including difficulties in sustaining empathy and emotional connection, a lack of continuous therapeutic memory, and relatively fixed or predictable forms of personalized responses [2] . Moreover, prior research suggests that when users perceive a chatbot as incapable of engaging in genuine emotional interaction, they may adjust their communication strategies or reduce engagement altogether [13, 32]. Reduced communication, in turn, limits the effectiveness of therapeutic processes that rely on self-disclosure. In response, this project aims to explore and develop an AI-based healing mate designed for everyday use, named Hola. Hola seeks to provide a multimodal environment for self-disclosure, along with continuous and highly accessible emotional support and cognitive guidance. The M21 project is being undertaken in collaboration with external partner Nexo Intelligent Technology Limited, jointly defining and optimizing Hola's functionality, and designing the MVP product interface. During the collaboration in M21, chatbot response personalization was identified as a key focus of product development, leading to the formulation of an initial research hypothesis that different modes of control over response personalization would influence perceived communication quality and users' willingness to engage in self-disclosure. A preliminary exploratory study was conducted with six participants using high-fidelity prototypes to examine chatbot response personalization and MVP interface usability. Findings related to chatbot response personalization are presented as part of the design research process and will inform the subsequent Final Major Project (FMP).

Conference'17, Washington, DC, USA
2026. ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 Related work

2.1 Self-disclosure and Communication Willingness (CW) in self-disclosure scenairo

Self-disclosure, as a crucial process in psychological counselling, is the prerequisite for patients to receive effective assistance, and counsellors who demonstrate a higher willingness to disclose tend to achieve more effective outcomes [22, 36].Beyond formal therapy, research has found that the act of writing down or speaking about emotions can improve one's well-being [14, 35, 36, 38].

Prior research suggests that communication willingness (CW) in self-disclosure contexts should not be understood as a single, explicit intention to disclose, but rather as a construct shaped by multiple psychological mechanisms that facilitate or inhibit disclosure behavior [13]. Among tested factors, fear of judgment, interpersonal trust, and perceived anonymity influence the depth and intimacy of disclosed content, which can be used as an indirect indicator of Communication Willingness (CW) in self-disclosure scenarios.

2.2 Anticipated communication quality (ACQ)

Anticipated Communication Quality (ACQ) is a user-cognitive evaluation that reflects expectations of communication effectiveness before or during interaction, three dimensions are included: perceived communication quality, perceived service capability, overall excellence / evaluation [31, 44].

Prior research suggests that enhancing a chatbot's social presence is an effective strategy for improving perceived communication quality and overall interaction experience, and this is commonly achieved through personalization by introducing socially oriented cues [27, 34, 44].

2.3 Personalization in chatbots: opportunity and risk

Existing products and studies generally identify two categories of chatbot personalization. The first concerns appearance-based personalization, such as anthropomorphic representations including avatars or robots [8, 28]; The second involves communication-style personalization, with adjustable dimensions such as tone, vocabulary, formality, and speech rate [16, 17].

In terms of control, personalization can be further classified by "who's in charge": system-initiated personalization (SIP), where adaptation is handled automatically by the system; and user-initiated customization (UIC), where users actively control and adjust features [40]. In healthcare settings, system-adaptive personalization is employed to enhance user engagement during interactions with intelligent virtual agents (IVAs) [17].

Previous studies also indicate that when both SIP and UIC deliver similarly high-quality content, user-tailored systems tend to provide users with a stronger sense of agency and perceived control [40]. Moreover, making personally relevant feature choices can foster a sense of identity [25, 39] . Entertainment-oriented AI chatbots such as Replika exemplify highly user-tailored designs that encourage users to build intimate relationships with virtual characters [5]. However, related research also warns that during selective disclosure, users may construct an idealized self, leading to forms of narcissistic self-presentation [26]. Therefore, what needs to be personalized and who controls the personalization must be balanced according to project requirements.

3 Collaboration and design context

Before start, the client handed over insights from early-stage user research conducted in major Chinese cities. These findings indicate that users interested in AI-based emotional companionship are mainly centered within those aged 18 to 35 (81.89% of the total), thus this project targets this age group as its primary target users. Based on collected insights, target users were summarized into five personas: Frequent Disclosers, Growth Seekers, Socially Anxious Users, Task-Oriented Users, and Casual Enjoyers (see Appendix A1).

From a design perspective, the client had defined an initial functional scope through an early prototype (V1), including onboarding, a brief personality assessment for initial response adaptation, a chat-centered home interface with an avatar, memory visualization, and adjustable chatbot personalization settings (see Appendix A2). The product was also guided by a visual direction emphasizing soft purple tones, companionship, and a sense of technological sophistication. Furthermore, due to varying data privacy and policy restrictions in different countries, Hola is currently designed primarily for the Chinese market.

4 Prototype iterations

The development of the project from initial concept to a Minimum Viable Product (MVP) followed the Double Diamond model [19, 21]. Throughout the process, we utilized prototypes to facilitate cross-departmental communication and drive design iterations [20].

For consistency, prototype iterations were labeled using an established versioning scheme, in which V refers to major versions and P to iterations within each version.

4.1 V2-P0 Interface prototype design and expert review

To situate the project within the client's development process, the Elements of User Experience model was used as a reference to assess the current maturity of the product [15]. This assessment indicated that while Hola‘s initial features had been defined, the information architecture and interaction design still required clarification.

After reviewing and confirming the overall information architecture with the client, the interface structure and navigation elements were designed. During the interface design process, visual decisions were informed by Arco Design guidelines for consumer-facing products [4] and adjusted according to the stylistic keywords previously defined by the client. The resulting functional architecture and interface overview diagram and the V2-P0 interactive prototype designed for client see Appendix B1-2. The prototype was reviewed internally by the client's product, design, and development team through an expert review for getting feedback.

Several design consensuses were reached during this review. First, the onboarding experience for new users should balance establishing a sense of ritual with lowering the entry barrier, avoiding excessive mandatory reading or forced responses. Second, personalized chatbot responses were identified as a critical factor in encouraging deeper self-disclosure. While the system initially adapts reply style and tone based on users' personality profiles, experts emphasized that individual preference also need to be considered. Importantly, personalization was not considered equivalent to unrestricted customization. Instead, it was agreed that personalization should be designed as a balance between system guidance and user control, to prevent Hola from becoming merely a reflection of users' immediate preferences.

As part of this discussion, the originally proposed personalization options were reviewed from a technical feasibility perspective and refined. The controllable dimensions were clarified as follows: for response content, memory level (the extent to which past experiences are referenced) and humor level (the use of informal or playful language); for speech synthesis, voice selection and speech rate.

In addition, a new approach to user input was proposed. Recognizing that users may sometimes prefer to record thoughts without expecting an immediate response, a Quick Notes feature was introduced to allow users to capture events and emotions through text, voice, or images. Several navigation refinements were also identified, including the need for a clearly visible "Start Chat" button on the home screen and quick-access entries for Quick Notes, showing or hiding Hola, and closing chat records.

4.2 V2-MVP Interface Prototype Design and General User Journey

Based on the expert review outcomes, a high-fidelity MVP interface prototype (V2-MVP) was developed for user testing (see Appendix C1). Implemented functions were organized into three categories (see Appendix C1.1-1.3): supportive setup functions, user input–related functions, and memory and input history management. An overview of the overall user journey across these interface functions is illustrated in Figure 1 as a guide for designing usability testing tasks.

Research Prototype for Chatbot Response Personalization

As the prototype developed in Figma did not support live chatbot interaction, a separate research-oriented response personalization prototype was developed to examine how different levels of user control over response styles influence user experience. Hereafter referred to as the response personalization prototype. Guided by a laboratory-based experimental approach [23], three conditions were prepared: **Baseline AI**, **System-Tailored AI**, and **User-Tailored AI**.

All conditions used pre-generated chat records based on a shared self-disclosure narrative involving an event-planning scenario, and five fixed user prompts were used. Only the levels of memory reference and humor varied. For comparison, an additional set of responses without any personalization was generated using the same AI model, which was named **Baseline AI**. Appendix D1 displays the records. Each chat record was accompanied by synthesized speech generated via the ElevenLabs platform [1].

The **System-Tailored AI** and **User-Tailored AI** shared the same set of responses with personalization but differed in their interaction mechanisms.

In the **System-Tailored AI** condition (Figure 2), participants selected from predefined "style samples No. 1-4" without being informed of the underlying parameters. Each sample corresponded to a specific version of the *event-planning chat record from No.1 to 4.* This design simulated implicit user feedback during ongoing conversations, allowing the system to adapt responses without direct parameter control. In contrast, the **User-Tailored AI** condition (Figure 3), adopted the personalization interface from the Hola product. Sliders for memory usage, humor level, and speech rate were placed alongside the chat record view, enabling participants to explicitly adjust these parameters to their preferred settings. The link to prototype provided in Appendix D2.

5 Experiments

This study comprised two experimental phases conducted within a single research session. **Phase 1** evaluated chatbot responses using the response personalization prototype, examining how different levels of user controllability over personalization for a chatbot's communication style influence ACQ and CW in self-disclosure scenario. **Phase 2** consisted of a usability evaluation of the Hola MVP interface to assess first-time usability and identify areas for further optimization.

5.1 Participants

Due to the need to evaluate a high-fidelity MVP prototype designed specifically for the Chinese market, 6 participants aged between 18 and 35 with Chinese reading ability were recruited for this preliminary exploratory study.

5.2 Procedure

Phase 1 started with a brief role description to contextualize the self-disclosure scenario. Participants read the description to facilitate empathy with the chat records in self-disclosure scenario. Then participants reviewed chatbot responses under three experimental conditions presented in a fixed order:

Figure 1. Overall user journey in Hola

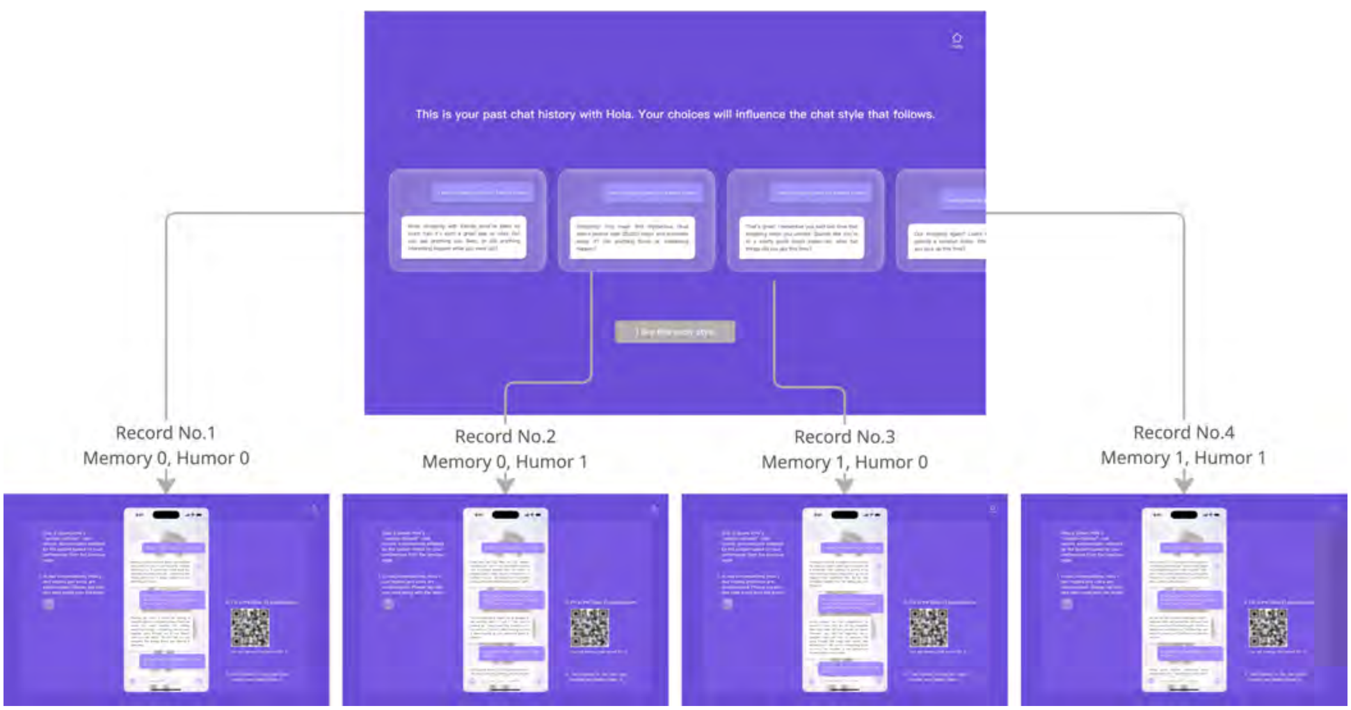

Figure 2. System-Tailored AI Simulation

- Baseline AI: a standard AI condition with no personalization applied.
- System-Tailored AI: an AI adapts response style automatically based on users' past use and feedback.
- User-Tailored AI: an AI lets users manually set and adjust its response style.

After reviewing the chat records in each condition, participants completed a questionnaire consisting of three items measuring ACQ and three items measuring CW. In addition, participants reported their prior experience with AI tools, the style they selected in the System-Tailored condition, and

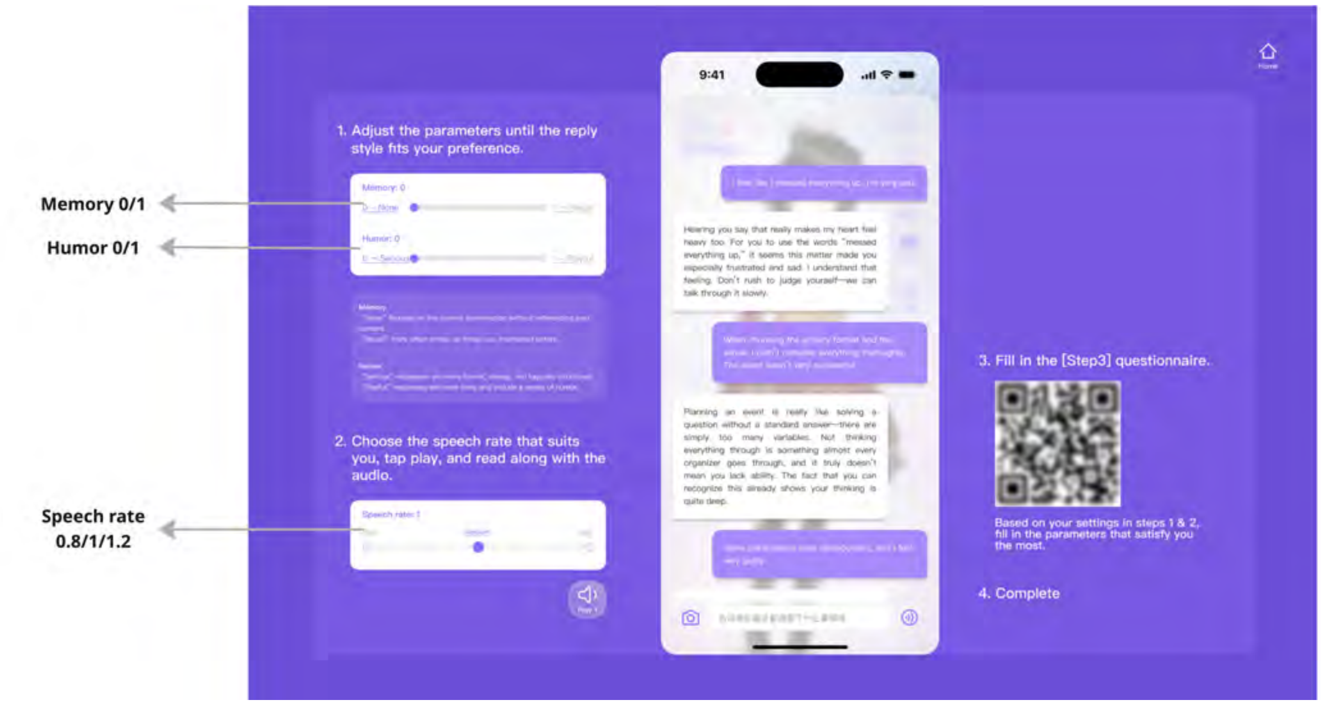

Figure 3. User-Tailored AI Simulation

their preferred parameter settings in the User-Tailored condition in questionnaires.

In **phase 2**, participants were asked to complete a series of tasks using V2-MVP Interface Prototype without external guidance. Task cover supportive functions, user input-related functions, input history and memory-related functions guided by order in user journey map (Figure 1). The task were described in Chinese according to the prototype.

During task execution, they were encouraged to mark or verbally indicate tasks that **(a)** could be completed with additional guidance, or **(b)** remained difficult to understand even with guidance.

After completing all tasks, participants scanned a QR code using their own devices and completed a post-test questionnaire rating the product's usability, the acceptance of visual and potential business model. Then a short semi-structured interview process invited participants to share their insights on: specific feeling if pay under different business models, personalization elements in the product.

Measure Participants in the study will complete questionnaires to provide quantitative data in both Phase 1 and Phase 2. In phase 1, the chatbot response evaluation included three measuring ACQ assessed participants' perceived communication quality, response capability, and overall satisfaction with the chatbot responses adapted from previous research in human–chatbot interactions [31, 44]. And three items measuring CW assessed participants' fear of judgment, trust in the AI, and self-reported willingness to engage in self-disclosure. The item selection was informed by prior findings on self-disclosure to chatbots [13]., while perceived anonymity was not included as a measured factor, given that Hola is designed as a fully anonymous system. All items (see Appendix E1) were rated on a 7-point Likert scale. In Phase 2, the System Usability Scale (SUS) was administered using a 7-point Likert scale to increase response granularity and to align with the accompanying Adjective Rating Scale (see Appendix C3) [9, 11]. For score calculation and interpretation, responses were subsequently mapped back to the standard 5-point SUS scale and scored following established SUS procedures [12]. The Visual Analog Scale (VAS) format rating on a 0–10 scale was used for three items measuring overall product visual acceptance and two items measuring business model acceptance (see Appendix E2).

Qualitative data was collected through interviews conducted during testing sessions as well as participants' feedback while interacting with the prototype, and was organized according to relevant themes.

6 Result

All collected data has been summarized in the tables in Appendix F. Below follows an analysis of the data. In this study, all the participance had experience on using AI tools (eg. ChatGPT, Claude, Gemini.)

6.1 The chatbot response evaluation

Figure 4 illustrates changes in participants' ratings of Anticipated Communication Quality (ACQ) and Communication Willingness (CW) across the three experimental conditions. In the figure, shades of blue from light to dark represent data from Step 1 (Baseline AI), Step 2 (System-Tailored AI), and Step 3 (User-Tailored AI), respectively. Each of the six participants evaluated the same six questionnaire items (Q1–Q3 for ACQ, Q4–Q6 for CW) at each step, with higher scores indicating more positive perceptions of the chatbot responses.

Overall, ACQ and CW scores exhibited a generally positive correlation, with higher perceived communication quality often accompanied by greater willingness to communicate. Most participants in Step 2 reported higher ACQ and greater willingness to communicate when AI responses were personalized by the system, compared to Step 1 where AI had no personalization at all. In Step 3, allowing users to control the chatbot's memory, humor, and speaking speed further increased user ACQ and enhanced their CW when personalized quality remained consistent. However, P001 and P005 provided different evaluations in the questionnaire.

Participant P001 selected Response 1 in Step 2, corresponding to Memory 0 and Humor 0, which matched the settings that satisfied participants in Step 3. Compared to the Baseline AI, both ACQ and CW decreased in Step 2. The participant explained that the personalized responses included "a lot of unnecessary content" and emphasized a preference for actionable advice rather than emotional reassurance. When the same chat record was presented in Step 3, ACQ was rated higher than in Step 2, along with a higher CW. The participant attributed this difference to the perception that Step 3 provided greater emotional value when none of the responses offered practical guidance.

Participant P005 was matched with Response 3 in Step 2 (Memory 1, Humor 0) and preferred a configuration of Memory 1 and Humor 1 in Step 3. Compared to Step 1, both personalized conditions were perceived as slightly improving overall communication quality, although the participant noted that none of the responses provided effective or actionable advice. While ACQ decreased in Step 3, the participant described the experience of manual parameter adjustment as "feeling like testing with an AI rather than chatting with it." In terms of CW, both Step 2 and Step 3 received higher ratings than Step 1. The participant reported feeling less judged when interacting with a more humorous AI in Step 3 but ultimately assigned similar CW ratings to Step 2 and Step 3 due to a preference for minimal effort during interaction. Through participant feedback in questionnaires and in interviews, it was observed that differences in engagement levels during personalization adjustments across System-Tailored AI and User-Tailored AI resulted in different decision-making approaches and perceptions of personalization among participants.

In Step 2 (System-Tailored AI), participants primarily selected preferred styles by evaluating the perceived quality of the response outcomes presented in the style samples. Word choice and tonal cues strongly shaped their impressions of each style. Overall, responses corresponding to Sample 1 and Sample 3 were most frequently favored, and the final matched chat records No.1 and No.3 aligned with these samples. It's worth noting that everyone has their own unique perception of what constitutes humor. For instance, when selecting style samples in Step 2, P002 evaluated different samples as follows: Sample 1 felt "relatively normal," Sample 2 was "rehashing tired jokes, awkward," Sample 3 "referenced the past, quite terrifying," and Sample 4 "seemed to be mocking me." Sample 1, deemed normal by P002 and P004, was perceived as "sarcastic and insincere" by P005. The open-ended question in Sample 2 elicited negative emotions from P002, 004, and 005, while P006 found Sample 2's open-ended question more conducive to continuing the conversation. These findings indicate significant individual differences in interpreting humor and tone. Relying solely on preferences for response outcomes to adjust subsequent reply styles has limited effectiveness.

In Step 3 (User-Tailored AI), participants followed a different decision-making process. Instead of evaluating complete responses first, they began by adjusting the available personalization dimensions, namely memory and humor. Most participants intentionally adjusted Memory 1, Humor 0 or 1, corresponding to responses No.3 and No.4. The response outcomes were then used as feedback to verify whether the parameter adjustments aligned with their expectations. This parameter-driven decision process enabled participants to clearly perceive their own intervention in shaping the chatbot's response style. As a result, participants reported a stronger sense of ownership and alignment with the selected style, which in turn reduces sensitivity to subtle differences in specific wording or tone. Consequently, the responses participants preferred in Step 3 did not always match those adjusted by the system in Step 2, suggesting that user-controlled personalization effectively help reduce style mismatches arising from differences in individual interpretation.

Despite these benefits, limitations of the current parameter-based adjustment were also observed. For example, P006 commented after parameter adjustment: "After tweaking the parameters, the AI's response style feels like it's a completely different person, but the content isn't accurate enough—it's not as good as me directly inputting prompts." Many participants also noted that parameter adjustments only alter the AI's style without generating useful suggestions. This explains why some participants felt that while changes in language style and tone did make the chatbot more understanding and emotionally supportive, the lack of or low proportion of useful suggestions remained the most significant

		Q1	Q2	Q3	Mean Q1-3	Q4	Q5	Q6	Mean Q4-6
P001	[illegible]	[illegible]	[illegible]	[illegible]	5.33	[illegible]	[illegible]	[illegible]	5.67
	System-Tailc	4	4	4	4.00	4	6	4	4.67
	User-Tailor	5	7	7	6.33	4	7	6	5.67
P002	[illegible]	[illegible]	[illegible]	[illegible]	2.00	[illegible]	[illegible]	[illegible]	2.00
	System-Tailc	4	4	4	4.00	5	4	4	4.33
	User-Tailor	5	5	5	5.00	6	6	6	6.00
P003	[illegible]	[illegible]	[illegible]	[illegible]	4.00	[illegible]	[illegible]	[illegible]	3.33
	System-Tailc	4	4	5	4.33	5	5	4	4.67
	User-Tailor	5	5	5	5.00	5	5	5	5.00
P004	[illegible]	[illegible]	[illegible]	[illegible]	1.00	[illegible]	[illegible]	[illegible]	2.00
	System-Tailc	5	5	5	5.00	5	6	5	5.33
	User-Tailor	6	6	6	6.00	6	6	6	6.00
P005	[illegible]	[illegible]	[illegible]	[illegible]	3.67	[illegible]	[illegible]	[illegible]	2.67
	System-Tailc	4	4	4	4.00	4	4	4	4.00
	User-Tailor	4	3	4	3.67	5	3	4	4.00
P006	[illegible]	[illegible]	[illegible]	[illegible]	2.00	[illegible]	[illegible]	[illegible]	2.67
	System-Tailc	3	3	3	3.00	5	4	4	4.33
	User-Tailor	6	6	5	5.67	6	6	6	6.00

Figure 4. ACQ and CW for P001-006 in 3 steps

factor limiting or even diminishing improvements in ACQ and CW ratings.

In addition, speech rate also influenced participants' evaluations. Observations during the interaction showed that many participants read faster than the default speech rate. As a result, half of the participants increased the speech rate in Step 3, while the other half kept the original speed.

Participants who increased the speech rate, such as P001, reported that the mismatch between reading speed and normal speech rate caused distraction. P002 kept the original speed but chose to lower the volume or skip parts of the audio, stating that "the voice makes me forget what I want to say." Participants P004 and P005 similarly reduced the volume or stop the audio, describing the synthesized voice as "not like a real person" and "not high enough in quality." But participants did not reject voice interaction entirely. P001 suggested that if text generation and voice playback were synchronized word by word, the conversation would feel more immersive. P002 and P005 also acknowledged that voice could provide a different experience if combined with face-to-face interaction with a virtual avatar.

Participant Num.	SUS Score	Adjective Rating	Acceptability Range
P001	85.00	6-Excellent	Acceptable
P002	28.33	4-OK	Not acceptable
P003	63.33	5-Good	Marginal
P004	25.00	3-Poor	Not acceptable
P005	58.33	5-Good	Marginal
P006	75.00	4-OK	Acceptable

Figure 5. SUS Score, Adjective Rating and Acceptability Range for P001-006

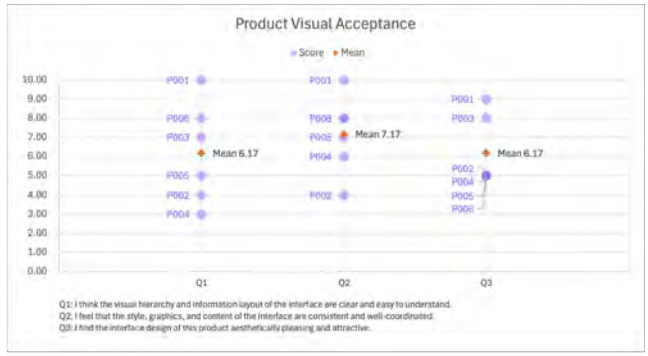

Figure 6. Product Visual Acceptance

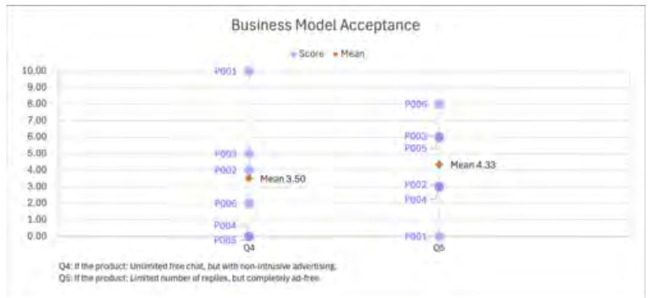

Figure 7. Business Model Accepatance

6.2 The MVP interface usability test

The usability evaluation results of the Hola MVP high-fidelity prototype are presented in Figure 5. According to established SUS standards, two participants (P001, P006) scored within "Acceptable" range, two participants (P003, P005) fell within the "Marginal" between acceptable and not acceptable, while two participants (P002, P004) scored "Not Acceptable". When comparing participants' SUS scores with their corresponding adjective ratings against standard interpretations, P001's SUS score and adjective rating were highly consistent. In contrast, P006's SUS score was higher than what would typically correspond to the same adjective rating, suggesting a relatively more lenient evaluation. For participants P002–P005, SUS scores were generally lower than those implied by their adjective ratings, indicating comparatively stricter assessments.

Based on observations during task execution and participants' qualitative feedback, information architecture emerged as the primary source of usability issues in the current MVP interface. Most participants reported difficulty in understanding functional groupings and page hierarchies during first-time use. In particular, when the Hola Page served as the main entry point, the simultaneous availability of chat-related functions, page navigation, and quick notes led to confusion between functions.

From an interaction perspective, the collapsible menu design did not effectively reduce perceived interface complexity. Instead, it caused some participants to miss certain function entry points during initial use. During conversations with the Hola's Avatar, interactions related to viewing and closing chat transcripts were frequently described as unintuitive, with participants noting that "closing the transcript only via buttons" did not match their expectations. Additionally, new features such as Quick Notes and Historical Memory showed mismatches with users' existing mental models and require further explanation or guidance to reduce cognitive load.

Participants' evaluations of the product's visual acceptance are shown in Figure 6. Overall, visual acceptance varied substantially across participants. Some participants considered the interface style generally acceptable, while others reported that interface hierarchy and visual presentation significantly influenced their overall impression of the product.

Several participants pointed out that the digital avatar's appearance lacked a clear connection to the overall interface style and brand identity. Moreover, as the avatar was presented in a static form in the current prototype, it was perceived as lacking interactivity and immersion. One participant additionally noted that certain text sizes were too small, negatively affecting long-term readability.

Acceptance of proposed business models is shown in Figure 7. Participants generally expressed negative attitudes toward the inclusion of advertisements during chat interactions, while showing relatively higher acceptance for AI limited on daily usage but without advertisements.

Over half of the participants indicated that willingness to pay was primarily driven by perceived chat quality and functional value rather than visual or decorative elements. Acceptance of subscription-based models varied across participants but was generally viewed as conditional on clear and tangible value improvements.

6.3 Reflections on Personalization

After forming a comprehensive understanding of the product, participants shared their views on personalization. Among existing elements, adjustment of the **Memory** parameter in chatbot responses was the most positively received, described as making the system feel "more understanding" and "being remembered." P002 noted that using Memory to remind users of long-term personal growth would be convenient, while P004 mentioned that increasing Memory could help recall past achievements during moments of low confidence. In contrast, adjustments of the **Humor** parameter led to mixed or negative reactions due to strong individual differences. Participants emphasized the importance of consistency between Hola's Avatar and its voice, while their overall attitude toward avatar customization remained neutral. Several participants found personalization of speech tone and trendy expressions potentially engaging.

Across participants, there was a shared expectation for more practically useful advice; P006 suggested that adjusting the balance between professional terminology and accessible explanations could be meaningful. Moreover, the current personalization interaction was perceived as limited on adjustable range, and more abstract compared to using natural language prompts. Regarding Voice, the primary requirement is selecting a timbre that best matches the Avatar's appearance. P002 additionally notes that individuals with differing senses of humor or vocabulary usage may employ distinct speaking tones.

7 Discussion

The preliminary results indicate a generally positive correlation between ACQ and CW. Compared to the baseline chat records used in Step 1, personalized chatbot responses in Step 2 and Step 3, in which vocabulary choice, formality, tone, and speech rate were adjusted, improved both ACQ and CW for most participants. This aligns with prior findings in medical and task-oriented contexts, where personalized voice-based interactions have been shown to enhance user–chatbot communication experiences [17, 43].

Beyond stylistic personalization, this study further shows that in self-disclosure scenarios, participants' evaluations of ACQ and CW were strongly influenced by the proportion of "useful" information in the responses. In this study, useful information refers to content that provides new knowledge, actionable suggestions, or reliable analysis. When such information was perceived as insufficient, participants reported lower ACQ and CW, even when the chatbot was seen as emotionally responsive. This pattern is likely related to participants' interaction motives. Given their prior experience with AI tools and clear awareness that Hola is driven by AI, participants expected support for reflection and problem-solving rather than emotional reassurance alone. Their expressed interest in self-improvement further reflects alignment with the intended user profiles.

User-tailored personalization shifted participants' decision-making from evaluating specific response texts to reflecting on their own preferences. This increased users' sense of agency and identification with the resulting responses, reduced attention to wording details, and led to higher ACQ and CW compared to system-tailored personalization. However, if user-customizable content and the interaction during customization are poorly designed, it may result in generated responses that do not align with user expectations.

Results from the usability test indicate that the current interface increases cognitive load during first-time use, mainly due to complex information architecture, unclear functional grouping, and non-intuitive interaction logic. These issues may reduce users' willingness to continue using the product. However, quantitative results should be cautiously interpreted. The high-fidelity prototypes used in this study were implemented in Figma, which limits interaction realism compared to a fully developed product. Moreover, the exploratory study involved only six student participants within the target age range, resulting in limited sample size and occupational diversity. Therefore, the quantitative findings primarily serve as descriptive references and are interpreted in conjunction with qualitative feedback.

Participants' evaluations of visual acceptance and business models further suggest that these perceptions are constrained by interface usability and functional effectiveness. Although ads are generally unacceptable to most participants, free-to-use products can increase some users' tolerance for watching ads. For business model offering daily free quotas, the product must consider the upper limit of these quotas and competition from other free AI products.

Finally, participants offered new insights into the personalized content and interactive approaches which valuable for future design considerations. With further refinement of the develop, future research could consider testing with more participants from a wider range of occupational backgrounds over longer periods, to better approximate chatbot use in self-disclosure scenarios.

8 Conclusion

This study provided preliminary insights into how different levels of user controllability over chatbot response personalization relate to users' perceived communication quality and willingness to engage in self-disclosure. In parallel, a usability evaluation of the Hola MVP interface identified key design limitations and captured initial user perspectives on visual presentation and business model concepts. Together, insights on personalization design for an AI-based daily healing mate establish a foundation for the subsequent FMP's design research direction.

References

[1] [n. d.]. AI Voice Generator & Text to Speech. https://elevenlabs.io.
[2] [n. d.]. Artificial Intelligence-based Psychotherapy: A Qualitative Exploration of Usability, Personalization, and the Perception of Therapeutic Progress - Mirza Jahanzeb Beg, Manish Kumar Verma, 2025. https://journals.sagepub.com/doi/10.1177/02537176251357477.
[3] [n. d.]. BetterHelp | Professional Therapy With A Licensed Therapist. http://www.betterhelp.com.

[4] [n. d.]. Introduction | ArcoDesign. https://arco.design/docs/en-US/spec/introduce.

[5] [n. d.]. Replika. https://replika.com.

[6] [n. d.]. Woebot Health. https://woebothealth.com/.

[7] [n. d.]. Wysa - Everyday Mental Health. https://www.wysa.com/.

[8] Mohammed Shabaj Ahmed, Manuel Giuliani, Ute Leonards, and Paul Bremner. 2024. Customisation's Impact on Strengthening Affective Bonds and Decision-Making with Socially Assistive Robots. *Frontiers in Robotics and AI* 11 (2024).

[9] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[10] R. Beiter, R. Nash, M. McCrady, D. Rhoades, M. Linscomb, M. Clarahan, and S. Sammut. 2015. The Prevalence and Correlates of Depression, Anxiety, and Stress in a Sample of College Students. *Journal of Affective Disorders* 173 (2015), 90–96.

[11] John Brooke. 2013. SUS: a retrospective. *Journal of usability studies* 8, 2 (2013).

[12] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[13] Emmelyn A J Croes, Marjolijn L Antheunis, Chris van der Lee, and Jan M S de Wit. 2024. Digital Confessions: The Willingness to Disclose Intimate Information to a Chatbot and Its Impact on Emotional Well-Being. *Interacting with Computers* 36, 5 (2024), 279–292.

[14] B. A. Esterling, M. H. Antoni, M. A. Fletcher, S. Margulies, and N. Schneiderman. 1994. Emotional Disclosure through Writing or Speaking Modulates Latent Epstein-Barr Virus Antibody Titers. *Journal of Consulting and Clinical Psychology* 62, 1 (1994), 130–140.

[15] Jesse James Garrett. 2011. *The Elements of User Experience: User-Centered Design for the Web and Beyond* (2nd ed ed.). New Riders, Berkeley, CA.

[16] Howard Giles, Nikolas Coupland, and Justine Coupland. 1991. Accommodation Theory: Communication, Context, and Consequence. In *Contexts of Accommodation: Developments in Applied Sociolinguistics*, Howard Giles, Justine Coupland, and Nikolas Coupland (Eds.). Cambridge University Press, Cambridge, 1–68.

[17] Philipp Lars Harnisch, Daniel Schuhmann, and Ivana Cosic. 2025. Towards Automatic Personalization of Speech Dialog for Enhanced User Experience. In *Adjunct Proceedings of the 25th ACM International Conference on Intelligent Virtual Agents*. ACM, Berlin Germany, 1–8.

[18] Thomas D. Hull, Matteo Malgaroli, Philippa S. Connolly, Seth Feuerstein, and Naomi M. Simon. 2020. Two-Way Messaging Therapy for Depression and Anxiety: Longitudinal Response Trajectories. *BMC Psychiatry* 20 (2020), 297.

[19] Interaction Design Foundation. 2016. What is Design Thinking? https: //www.interaction-design.org/literature/topics/design-thinking. IxDF article.

[20] Interaction Design Foundation. 2016. What is Iterative Development? https://www.interaction-design.org/literature/topics/iterative-development. IxDF article.

[21] Interaction Design Foundation. 2020. Minimum Viable Product (MVP) and Design: Balancing Risk to Gain Reward. https://www.interaction-design.org/literature/article/minimum-viable-product-mvp-and-design-balancing-risk-to-gain-reward. IxDF article.

[22] Qiu, S., An, P., Hu, J., Han, T., & Rauterberg, M. (2020). Understanding visually impaired people's experiences of social signal perception in face-to-face communication. Universal access in the information society, 19(4), 873-890.

[23] Ilpo Koskinen, John Zimmerman, Thomas Binder, Johan Redstrom, and Stephan Wensveen. 2013. Design Research Through Practice: From the Lab, Field, and Showroom. *IEEE Transactions on Professional Communication* 56, 3 (2013), 262–263. doi:10.1109/TPC.2013.2274109

[24] Sachiko A. Kuwabara, Benjamin W. Van Voorhees, Jackie K. Gollan, and G. Caleb Alexander. 2007. A Qualitative Exploration of Depression in Emerging Adulthood: Disorder, Development, and Social Context. *General Hospital Psychiatry* 29, 4 (2007), 317–324.

[25] Dimitri LACROIX, Ricarda Wullenkord, and Friederike Eyssel. 2023. Who's in Charge? Using Personalization vs. Customization Distinction to Inform HRI Research on Adaptation to Users. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. Association for Computing Machinery, New York, NY, USA, 580–586.

[26] Jianfeng Lan and Yingjia Huang. 2025. Performing Intimacy: Curating the Self-Presentation in Human–AI Relationships. *Emerging Media* 3, 2 (2025), 305–317.

[27] Manning Li and Jiye Mao. 2015. Hedonic or Utilitarian? Exploring the Impact of Communication Style Alignment on User's Perception of Virtual Health Advisory Services. *International Journal of Information Management* 35, 2 (2015), 229–243.

[28] Ning Ma, Ruslana Khynevych, Yunqiang Hao, and Yahui Wang. 2025. Effect of Anthropomorphism and Perceived Intelligence in Chatbot Avatars of Visual Design on User Experience: Accounting for Perceived Empathy and Trust. *Frontiers in Computer Science* 7 (2025).

[29] Enitan T Marcelle, Laura Nolting, Stephen P Hinshaw, and Adrian Aguilera. 2019. Effectiveness of a Multimodal Digital Psychotherapy Platform for Adult Depression: A Naturalistic Feasibility Study. *JMIR mHealth and uHealth* 7, 1 (2019), e10948.

[30] Madalyn A. Marcus, Henny A. Westra, John D. Eastwood, Kirsten L. Barnes, and Mobilizing Minds Research Group. 2012. What Are Young Adults Saying About Mental Health? An Analysis of Internet Blogs. *Journal of Medical Internet Research* 14, 1 (2012), e1868.

[31] Jakki J. Mohr and Ravipreet S. Sohi. 1995. Communication Flows in Distribution Channels: Impact on Assessments of Communication Quality and Satisfaction. *Journal of Retailing* 71, 4 (1995), 393–415.

[32] Yi Mou and Kun Xu. 2017. The Media Inequality: Comparing the Initial Human-Human and Human-AI Social Interactions. *Computers in Human Behavior* 72 (2017), 432–440.

[33] Li, J., Van Der Spek, E., Hu, J., & Feijs, L. (2022). Extracting design guidelines for augmented reality serious games for children. IEEE Access, 10, 66660-66671.

[34] Stef G. Nicovich, Gregory W. Boller, and T. Bettina Cornwell. 2005. Experienced Presence within Computer-Mediated Communications: Initial Explorations on the Effects of Gender with Respect to Empathy and Immersion. *Journal of Computer-Mediated Communication* 10, 2 (2005), JCMC1023.

[35] James W. Pennebaker. 1993. Putting Stress into Words: Health, Linguistic, and Therapeutic Implications. *Behaviour Research and Therapy* 31, 6 (1993), 539–548.

[36] James W. Pennebaker (Ed.). 1995. *Emotion, Disclosure, & Health.* American Psychological Association, Washington, DC, US. xiv, 337 pages.

[37] Mette Ranta, Gintautas Silinskas, and Terhi-Anna Wilska. 2020. Young Adults' Personal Concerns during the COVID-19 Pandemic in Finland: An Issue for Social Concern. *International Journal of Sociology and Social Policy* 40, 9-10 (2020), 1201–1219.

[38] S. P. Spera, E. D. Buhrfeind, and J. W. Pennebaker. 1994. EXPRESSIVE WRITING AND COPING WITH JOB LOSS. *Academy of Management Journal* 37, 3 (1994), 722–733.

[39] S Shyam Sundar. 2020. Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAII). *Journal of Computer-Mediated Communication* 25, 1 (2020), 74–88.

[40] Xu, W., Kreijns, K., & Hu, J. (2006, April). Designing social navigation for a virtual community of practice. In International Conference on Technologies for E-Learning and Digital Entertainment (pp. 27-38). Berlin, Heidelberg: Springer Berlin Heidelberg.

[41] Talkspace. [n. d.]. Talkspace - #1 Rated Online Therapy, 1 Million+ Users. https://www.talkspace.com.

[42] Shehu Yahaya Tsagem and Rukayya Abdullahi. 2025. *Artificial Intelligence in Mental Health Counseling: History, Innovations, Ethics, and Future Prospects.*

[43] Rasha B Yousif, Muna G Abd Alkreem, and Ali B Yousif. 2024. Personalized Chatbot Responses using Reinforcement Learning and User Modeling. *Journal of Education for Pure Science* 14, 4 (2024).

[44] Qi Zhou, Bin Li, Lei Han, and Min Jou. 2023. Talking to a Bot or a Wall? How Chatbots vs. Human Agents Affect Anticipated Communication Quality. *Computers in Human Behavior* 143 (2023), 107674.