Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Patient Acceptability of Artificial Intelligence in a DHN Workflow for Hybrid Mental Health Care

**Katherine Lim**
Beth Israel Deaconess Medical Center
**Haruka Notsu**
Beth Israel Deaconess Medical Center
**Matthew Flathers**
Beth Israel Deaconess Medical Center
**John Torous**
jtorous@bidmc.harvard.edu

Beth Israel Deaconess Medical Center

**Research Article**

**Keywords:** Digital mental health, Digital health navigator, Digital phenotyping, Artificial intelligence, Patient acceptability

**Posted Date:** November 24th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11718201/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

**Patient Acceptability of Artificial Intelligence in a DHN Workflow for Hybrid Mental Health Care**


## Abstract

**Background:** Digital mental health tools generate growing volumes of active and passive behavioral data, but integrating these data into clinical care and returning them to patients as usable feedback remains a major challenge. Artificial intelligence (AI), including large language models, are able to interpret such data and draft personalized feedback, offering a potentially scalable way to support this work. Whether AI can be adopted for this purpose, however, may depend less on its technical performance than on whether patients trust and accept it. This study examined how former and current patients in mental health service perceive AI involvement in delivering data-report feedback, comparing fully human (no AI), AI-assisted (hybrid), and fully AI-generated formats.

**Methods:** Twenty-five adults who had received outpatient mental health treatment viewed six standardized videos representing two clinical scenarios and the three delivery formats. Participants rated each video using a seven-item acceptability scale, compared the formats, and answered open-ended questions. They were not told which formats used AI until after completing the initial ratings.

**Results:** Acceptability differed across formats, $F(1.20, 28.7) = 21.22$, $p < .001$. Fully human feedback (mean 6.00) and AI-assisted feedback delivered by a human (mean 5.93) received similar ratings, and both were rated more favorably than fully AI-delivered feedback (mean 4.65). Participants were also more comfortable with AI analyzing their data when a human delivered the feedback than with AI both analyzing and delivering it. Overall, 23 of 25 participants preferred a format that retained human involvement after AI use was disclosed. Participants valued AI for producing clear and thorough summaries but wanted a person to explain and discuss the results.

**Conclusions:** Patients accepted AI-generated mental health feedback when a human remained involved in reviewing or delivering it. These findings support using AI to assist with data analysis and report preparation rather than replacing the human interaction involved in mental health care. Larger studies should test this approach in routine clinical settings.



## Keywords

Digital mental health; Digital health navigator; Digital phenotyping; Artificial intelligence; Patient acceptability


## Introduction

Smartphone-derived signals are increasingly recognized as clinically meaningful indicators of real-world functioning that may support measurement-based and personalized care (Bond et al., 2023; Fortney et al., 2017;

Löchner et al., 2025). Digital tools can capture patients' experiences in near real time, whereas treatment response is still often assessed through periodic, retrospective rating scales that may not support timely clinical decision-making. Open-source platforms such as mindLAMP have made it easier for clinical teams to collect active data, including brief symptom surveys, and passive data, including sleep, physical activity, and phone-use patterns (Vaidyam et al., 2022). However, easier data collection has created a new implementation challenge as continuous digital data streams are often too large and complex for routine clinical workflows to organize, interpret, and translate into actionable insight (Smith et al., 2025). As a result, the clinical value of digital phenotyping depends on both data collection and the ability to transform these data into clear, interpretable information at the point of care.

The Digital Clinic model at Beth Israel Deaconess Medical Center (BIDMC) was developed to address this implementation gap (Calvert et al., 2025; Macrynikola et al., 2023). In this model, patients receiving outpatient psychotherapy use mindLAMP to complete brief daily surveys and passively collect behavioral data (Rodriguez-Villa et al., 2020). A Digital Health Navigator (DHN), a trained non-clinician team member, supports the clinical use of these data by reviewing smartphone-based passive sensing data and digital phenotyping outputs, identifying relevant patterns, preparing individualized reports, and helping patients and clinicians engage with the findings (Chen et al., 2024). By converting complex data streams into patient-facing summaries, the DHN makes measurement-based care more feasible within a hybrid care workflow. However, the model faces substantial scalability constraints, including limited workforce capacity, lack of reimbursement pathways, and time-intensive data-analysis and reporting processes (Flathers et al., 2025; Jin et al., 2025).

Recent advances in artificial intelligence (AI), particularly large language models (LLMs), may help address these challenges. Early studies suggest that AI can interpret digital-phenotyping data and generate individualized summaries (Hua et al., 2025; Jin et al., 2025). These capabilities may be especially relevant to the DHN role, where a central task is translating complex behavioral data into accessible, clinically meaningful feedback. However, AI applications in behavioral health monitoring, particularly within DHN workflows, remain limited. The present work examines how AI can be integrated into the DHN role by supporting and partially automating digital data interpretation and reporting while preserving the human connection and clinical judgment central to the role.

However, technical feasibility alone does not ensure clinical implementation. Whether AI-supported tools can be successfully integrated into mental health care may depend less on technical capability than on whether patients trust and accept them. Trust and acceptance are well-recognized determinants of whether health technologies are adopted

(Bracic et al., 2026), and they are especially consequential in mental health, where the therapeutic relationship is itself a core mechanism of benefit (Baier et al., 2020). If patients perceive AI involvement as impersonal, untrustworthy, or inappropriate, distrust may become a decisive implementation barrier that technical performance alone cannot overcome (Benda et al., 2024). Understanding the sources of this distrust, as well as the conditions under which patients find AI acceptable, is therefore a prerequisite for responsible deployment.

Despite this, patient perspectives remain under-represented in literature, which has focused largely on technical performance or clinician-facing applications (Hussein et al., 2026; Lee et al., 2025; Zidaru et al., 2021). Despite rapid advances in AI for mental health, relatively little research has examined how patients in mental health service perceive AI's role in their care or where it might best fit within existing DHN workflows. In particular, little is known about how patients respond to different configurations of AI involvement in generation and delivery. Centering patient perspectives is both an ethical imperative and a practical necessity, as successful implementation depends on designing AI-supported interventions that align with patient preferences.

The present study addresses this gap using a within-subjects, mixed-methods vignette experiment examining patient preferences for different modes of delivering digital mental health data reports. Participants viewed standardized clinical vignettes presented in three formats: fully human, in which a human-written script was delivered by a person (no AI involvement); AI-assisted, in which AI-drafted content was edited and delivered by a person (human–AI hybrid); and fully AI-generated, in which an AI-written script was delivered using a synthetic voice. Former and current mental health patients rated and compared each format and explained their preferences. By combining quantitative ratings with qualitative explanations, this study evaluates patient acceptance of AI within the DHN workflow and identifies the configurations in which AI involvement is most acceptable.

## Methods

### *Design*

This mixed-methods, within-subjects experimental study used fictional vignettes to examine participants' perceptions of three standardized data-report delivery formats. Participants provided both quantitative ratings and qualitative feedback. To minimize bias, the study incorporated planned mild deception (incomplete disclosure), with participants not being informed during the initial evaluation that some feedback scripts were AI-generated. Following the rating task, participants were fully debriefed, informed which formats involved AI, and asked how

this information affected their perceptions. The protocol was approved by the BIDMC Institutional Review Board (IRB #2026P000062), and all participants provided informed consent.

***Participants***

Eligible participants were adults (≥18 years) who had received mental health treatment with the Digital Clinic, were able to provide informed consent in English, and had access to a device and Internet connection sufficient to join a video session and complete online questionnaires. Potential participants were identified through electronic health record eligibility screening and consented online. No medical exclusions were applied. Because participants had previously participated in the Digital Clinic, they were familiar with the format of the digital report and the DHN role.

***Stimuli and Conditions***

Study materials were based on two fictional clinical vignettes. One featured "Sarah," whose primary presenting issue was work-related stress, and "David" whose primary presenting issue was family-related stress. Sarah's digital report highlighted problems in physical activity, and David's highlighted problems in sleep quality. For each vignette, the same underlying digital phenotyping data (e.g., step count and sleep) were presented in three delivery formats (Fig. 1): fully human (script written and delivered by a human DHN), AI-assisted hybrid (script generated by AI, then delivered by a human DHN), and fully AI-generated (script generated by AI and delivered using AI text-to-speech). Each participant viewed all six videos (2 vignettes × 3 delivery formats) in a randomized order. Visual stimuli were standardized across conditions, and all videos were approximately 2–3 minutes in length.

Scripts for the AI-assisted (hybrid) and fully AI-generated conditions were produced using Google Gemini 3.5 Flash, from a standardized prompt that supplied the vignette's digital phenotyping data and the required report structure. The delivery script is included in Appendix 1. For the hybrid condition, the AI-drafted script was reviewed by a DHN before delivery. For the fully AI-generated condition, the same class of AI-drafted script was rendered as speech using NaturalReader, with no human editing. No script was prepared for the fully human condition. This was deliberate: to simulate a real session, DHNs delivered the report as they would in routine practice, guided by the standard review guidelines used in the Digital Clinic (Appendix 1), rather than reading prepared text.

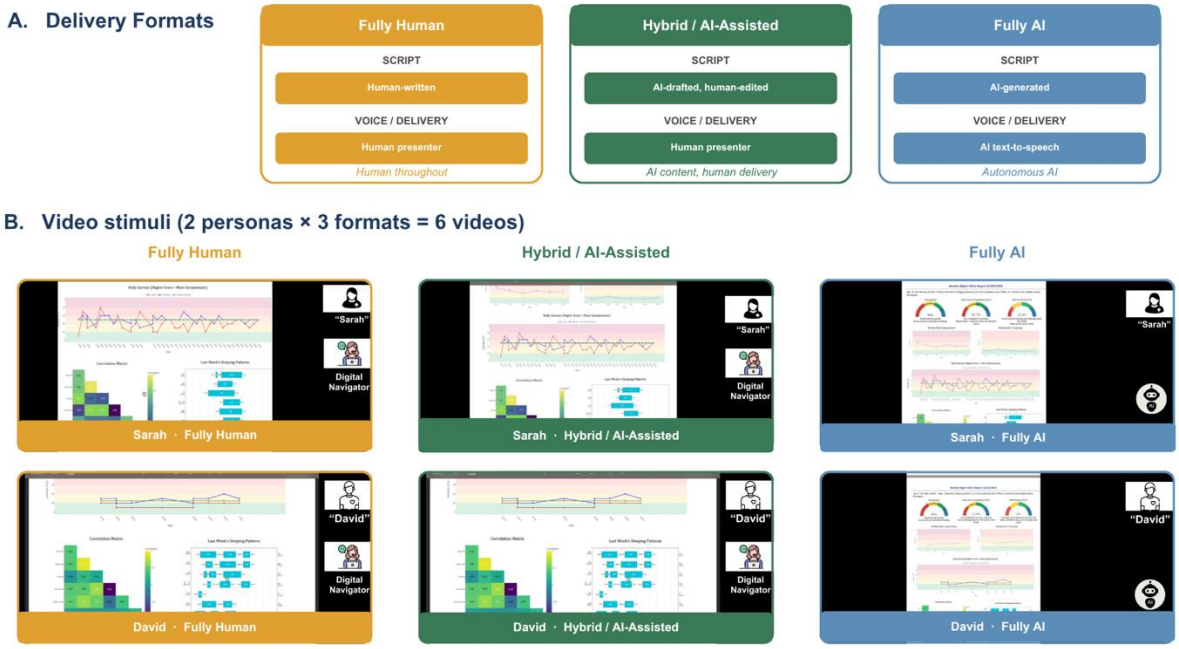


**Fig. 1** Study design and video stimuli. The formats differed in script source and voice/delivery (Panel A). Two clinical-vignette personas (rows), Sarah (physical activity and work-related stress) and David (sleep hygiene and family-related stress), were each presented in three data-report delivery formats (columns), for a total of six standardized videos. During the rating task, participants were not informed which videos used AI-generated content; accordingly, no persona or visual element was used to represent the AI format in the study materials or figure. The identity of the AI-generated videos was disclosed only during debriefing after participants completed the questionnaire

***Procedure***

Participants first completed a pre-study questionnaire. They then viewed the six videos in randomized order and completed the acceptability scale after each video. After viewing all six videos, participants completed the forced-choice format comparisons and indicated their preferred format for a clinical session. Throughout the viewing, rating, and comparison tasks, participants were blinded to the identity of the formats: the videos were labeled only with neutral symbols, and participants were not told which videos involved AI. Following debriefing, in which the AI involvement in each format was disclosed, participants completed a final questionnaire (Fig. 2).

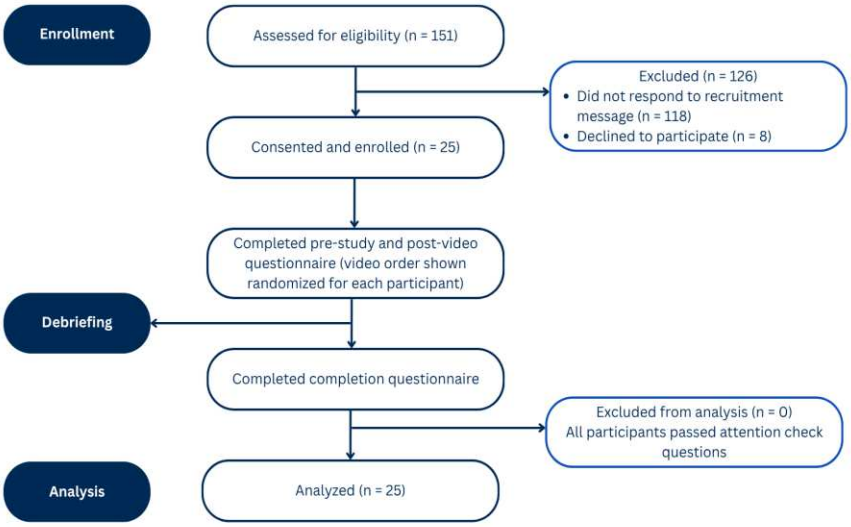


**Fig. 2** CONSORT-style participant flow ($n$ = 25). Video order shown was randomized for each participant; the design is within-subjects, so every participant viewed all three delivery formats

***Measures***

The full set of questions is provided in Appendix 2.

Participant characteristics and baseline attitudes were assessed with a brief questionnaire comprising age range, the perceived helpfulness of the DHN during past treatment (rated from 1 [not at all helpful] to 7 [among the most helpful aspects of care]), and overall opinion of AI (rated from 1 [very negative] to 7 [very positive]). The questionnaire was administered once, before viewing the videos. Responses were summarized descriptively and used as individual-difference variables in exploratory analyses.

Acceptability was assessed after each video using a seven-item self-report scale adapted from established technology-acceptance and user-experience frameworks (Glikson & Woolley, 2020; Henson et al., 2019). Items assessed perceived accuracy and relevance, clarity and ease of understanding, empathy and emotional comfort, and motivation and engagement, with domains adapted from the Technology Acceptance Model, System Usability Scale, and User Experience Questionnaire (Lewis, 2018; Rahimi et al., 2018; Schrepp et al., 2017). Responses were rated from 1 (strongly disagree) to 7 (strongly agree) and averaged to create a combined acceptability score ranging from 1 to 7. The scale showed excellent internal consistency (Cronbach's $\alpha$ = .93). Each post-video questionnaire also included one randomly selected attention-check item about a simple visual detail to verify engagement. The

primary outcome was the combined score for each delivery format, averaged across the two vignettes and compared within subjects.

After viewing all six videos, and while still blinded to format identity, participants selected the single best-performing format for each of eight attributes spanning explanation content, tone, and engagement: comprehensiveness of the summary, personalization, activity–mood insight, professional supportiveness, achievability, tracking value, therapist discussion, and health-habit discussion. "No preference" was an available option. Responses were summarized by the number of times each format was chosen.

Format preference was assessed with a single item: "Which format would you prefer to receive in a clinical setting?" This item was administered twice: once after all videos while participants remained blinded to each format's AI involvement, and again after debriefing once AI involvement had been disclosed. Responses were summarized by format and examined for changes after disclosure.

Following debriefing, participants rated their comfort with two AI use cases: AI analyzing their data and generating feedback for them to read, and a human navigator using AI to analyze the data but delivering the feedback personally. Each was rated from 1 (not at all comfortable) to 7 (extremely comfortable). Participants also rated agreement with the statement, "Using AI like this feels appropriate in a healthcare setting," from 1 (strongly disagree) to 7 (strongly agree), adapted from Unified Theory of Acceptance and Use of Technology (UTAUT) and implementation science frameworks (Sekhon et al., 2017; Venkatesh et al., 2016). The primary analysis compared comfort ratings between the two AI configurations. These post-debrief items are a separate instrument from the per-video acceptability scale and are not directly comparable to it.

Finally, five free-response questions administered after debriefing explored perceived benefits, concerns, and preferences regarding AI- and human-generated feedback. Responses were analyzed thematically.

***Analysis***

Analyses were conducted in Python 3 using SciPy 1.17. Composite scores, formed as the mean of seven items, were treated as interval-scaled and analyzed parametrically, whereas single-item ordinal ratings were analyzed with rank-based tests. Normality of composite difference scores was assessed using Shapiro–Wilk tests and sphericity using Mauchly's test. Difference scores for all three format contrasts were consistent with normality ($W = .93–.96$, all $p > .05$). Mauchly's test indicated a violation of sphericity, $\chi^2(2) = 25.57$, $p < .001$, consistent with the substantially

greater variance in the fully AI condition. Accordingly, Greenhouse–Geisser-corrected degrees of freedom ($\varepsilon = .60$) are reported for the primary repeated-measures ANOVA. All tests were two-tailed.

The primary outcome, composite acceptability, was compared across delivery formats using a one-way repeated-measures ANOVA, with format as the within-subjects factor and partial $\eta^2$ as the effect size. Pairwise comparisons used paired-samples t tests with Bonferroni correction for three comparisons and are reported with mean differences, 95% confidence intervals, t statistics, and Cohen's dz. The seven individual acceptability items were compared between the fully human and fully AI formats using Wilcoxon signed-rank tests with Holm–Bonferroni correction. Comfort with the two AI configurations was compared using a Wilcoxon signed-rank test.

Because each participant contributed eight non-independent forced-choice selections, analyses were conducted at the participant rather than pooled-selection level. For each participant, the number of attributes for which each format was selected (0–8) was calculated and compared across formats using a repeated-measures ANOVA with Bonferroni-corrected paired t tests, with a Friedman test as a distribution-free confirmation.

Binary outcomes are summarized as proportions with Wilson 95% confidence intervals and tested against chance using exact binomial tests. Changes in preference from before debriefing to after debriefing were evaluated using McNemar's exact test. Exploratory individual-difference analyses used Spearman rank-order correlations and subgroup, and order comparisons used Mann–Whitney U tests. Open-ended responses were analyzed using thematic analysis to identify recurring themes, which were integrated with the quantitative findings. Given the pilot nature of the study, statistical findings should be interpreted as exploratory.

**Results**

Of 151 individuals assessed for eligibility, 118 did not respond to the recruitment message and 8 declined, yielding 25 participants who consented, completed all study procedures, and were included in the analysis (Fig. 2). Participants were predominantly younger adults (roughly four in five were under 45 years). All participants rated their previous experience with DHNs as at least slightly helpful. Baseline attitudes toward AI were mixed (approximately half positive, one third negative, the remainder neutral). All 25 participants passed the embedded attention checks (100%). Participant characteristics and baseline measures are summarized in Table 1.

**Table 1** Participant characteristics and baseline measures ($N = 25$)

| Characteristic | Category | n | % |
|---|---|---|---|
| Age range | 25–34 years | 12 | 48% |
| | 35–44 years | 9 | 36% |

| Characteristic | Category | n | % |
|---|---|---|---|
| | 45–54 years | 2 | 8% |
| | 55–64 years | 2 | 8% |
| DHN helpfulness | Extremely / Very helpful | 16 | 64% |
| | Moderately / Somewhat / Slightly | 9 | 36% |
| Baseline opinion of AI | Positive (slightly–very) | 13 | 52% |
| | Neutral | 4 | 16% |
| | Negative (slightly–very) | 8 | 32% |

Composite acceptability differed across formats, $F(1.20, 28.7) = 21.22$, $p < .001$, partial $\eta^2 = .47$. Figure 3 shows that fully human ($M = 6.00$, $SD = 0.57$, 95% CI [5.76, 6.23]) and AI-assisted hybrid ($M = 5.93$, $SD = 0.67$, 95% CI [5.65, 6.21]) explanations were rated almost identically, and both were rated above fully AI-generated explanations ($M = 4.65$, $SD = 1.22$, 95% CI [4.14, 5.15]).

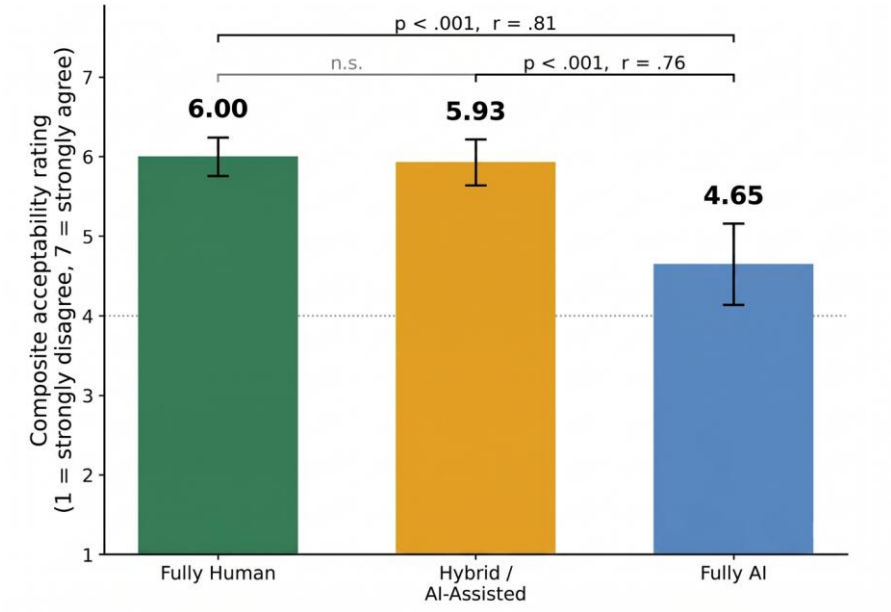


**Fig. 3** Composite acceptability ratings by explanation format. Bars are group means; whiskers are 95% confidence intervals; points are individual participants (N = 25). The dotted line marks the scale midpoint (4 = neutral). Fully human and hybrid formats were statistically indistinguishable, with both significantly exceeding the fully AI format. Bonferroni-corrected paired-samples t tests confirmed large advantages of both human and hybrid formats over fully AI delivery (human vs AI: mean difference = 1.35, 95% CI [0.79, 1.92], $t(24) = 4.93$, $p < .001$, $dz = 0.99$; hybrid vs AI: mean difference = 1.29, 95% CI [0.71, 1.87], $t(24) = 4.57$, $p < .001$, $dz = 0.91$). The human and hybrid formats did not differ (mean difference = 0.07, 95% CI [−0.14, 0.27], $t(24) = 0.66$, $p = 1.00$, $dz = 0.13$). Distribution-free

tests gave the same pattern (Friedman $\chi^2(2) = 19.68$, $p < .001$, Kendall's W = 0.39; Wilcoxon signed-rank tests, human vs AI r = 0.81 and hybrid vs AI r = 0.76, both $p < .001$, human vs hybrid $p = .69$).

At the item level acceptability scale (Fig. 4), the fully AI format performed less favorably on relational and motivational attributes, including perceived personalization, the balance between professionalism and support, and motivation to discuss the feedback with their therapist. In contrast, fully AI was statistically indistinguishable from the human format in explaining the relationship between activities and mood.

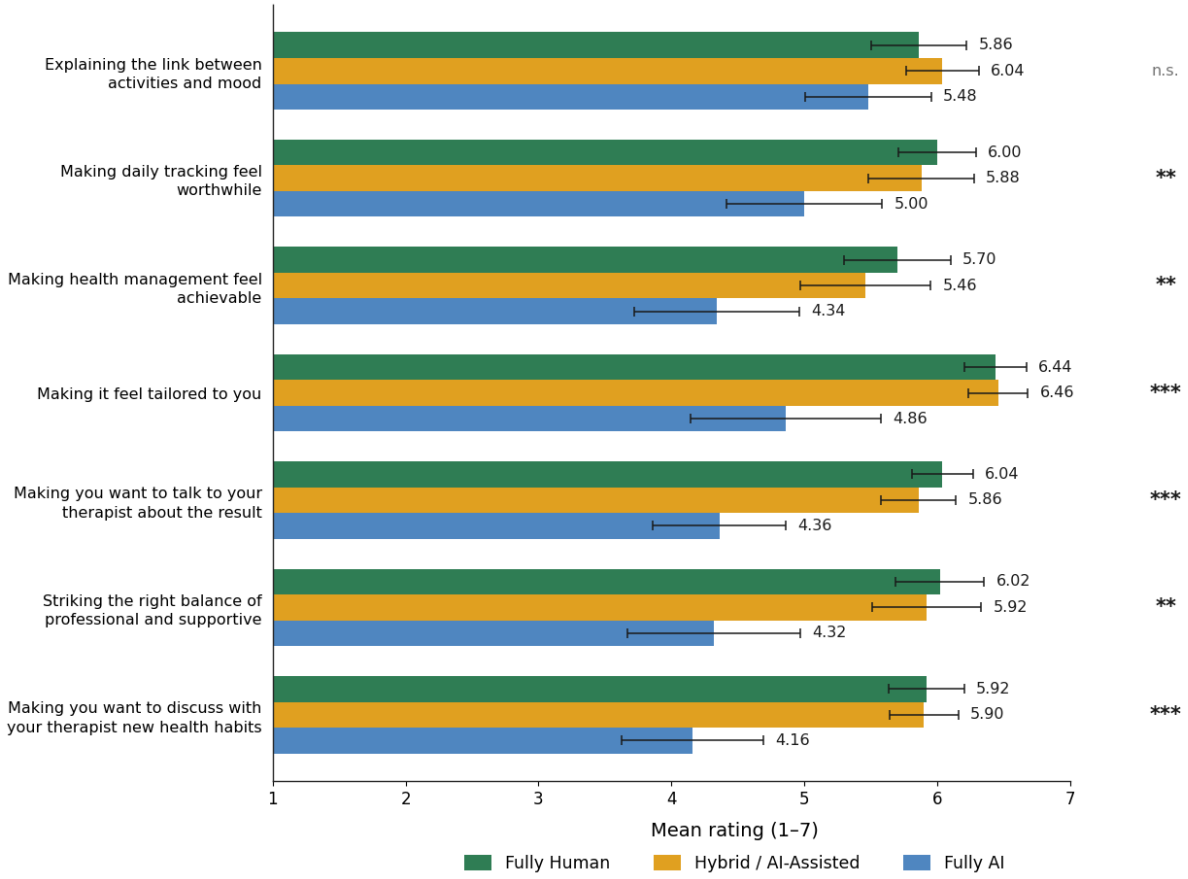


**Fig. 4** Mean ratings for each of the seven acceptability items by format, ordered from smallest to largest fully-human-minus-fully-AI difference. Bars are group means; whiskers are 95% confidence intervals. Stars denote Holm–Bonferroni-corrected Wilcoxon comparisons of the human versus fully AI format (*p < .05, **p < .01, ***p < .001; n.s. = not significant).

In the forced-choice comparison task, where participants, still blinded to format identity, selected the single best-performing format across eight aspects of explanation quality (N = 200 selections, 25 participants × 8 attributes), the hybrid format was selected most often (86 selections, 43.0%), followed by fully human (58 selections, 29.0%) and

fully AI (41 selections, 20.5%); the remaining 15 selections (7.5%) were "No preference" (Fig. 5). At the participant level, each participant selected the hybrid format for a mean of 3.44 attributes (SD = 2.72), compared with 2.32 (SD = 2.50) for fully human and 1.64 (SD = 2.04) for fully AI; this difference did not reach significance, $F(1.93, 46.25) = 2.43$, $p = .10$, partial $\eta^2 = .09$ (Friedman $\chi^2(2) = 2.88$, $p = .24$), and no pairwise contrast survived Bonferroni correction. The fully AI format was most competitive for providing a comprehensive summary of the data and making daily tracking feel worthwhile. When asked which format they would prefer to be used in a clinical setting, 12 participants selected the hybrid format, 9 selected the fully human format, and 3 selected fully AI; one participant reported no preference.

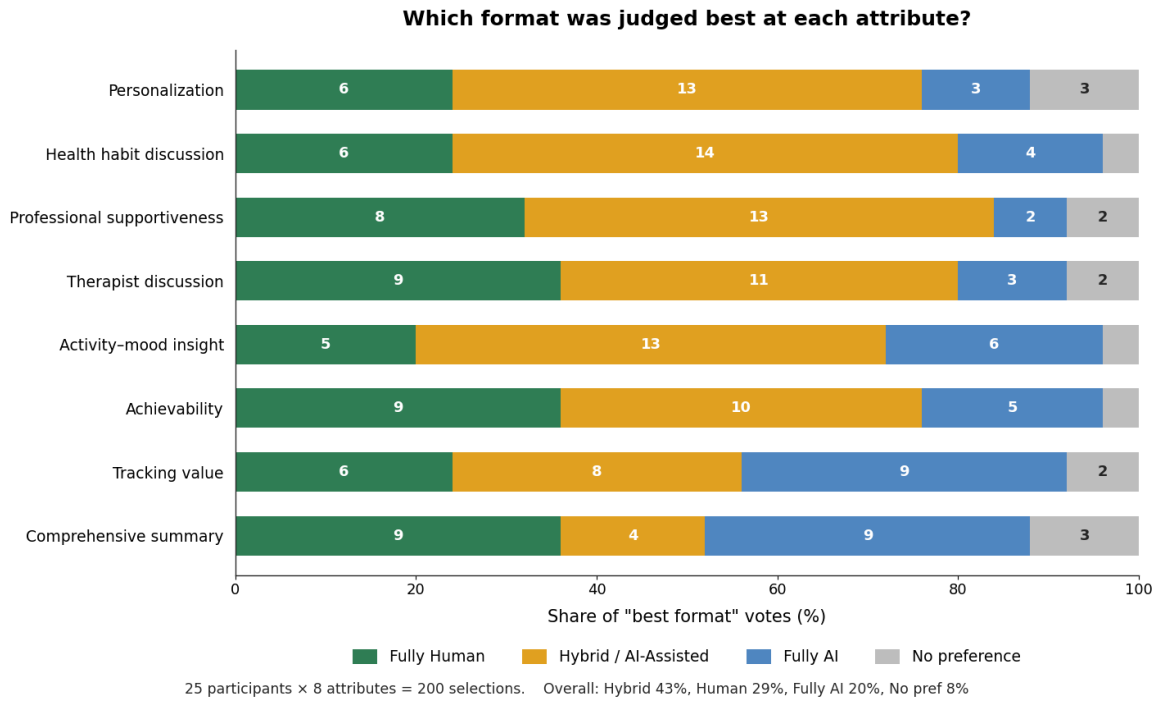


**Fig. 5** Forced-choice "best format" selections for each of eight attributes (percentage of selections; counts within bars; 25 participants × 8 attributes = 200 selections)

When the same clinical-session preference question was repeated after participants were told about the extent of AI involvement in each of the videos, 23 of 25 participants continued to select a human-involved format (12 fully human, 11 hybrid, 2 fully AI; Fig. 6). Comfort with AI, rated on a 7-point scale, depended on human involvement (Fig. 7), with 20 of 25 participants reporting feeling at least slightly comfortable with AI analyzing their data to inform a human feedback session ($M = 5.28$, $SD = 1.79$), compared with 11 of 25 for AI both analyzing and delivering feedback ($M = 4.00$, $SD = 2.10$; Wilcoxon $p = .012$, $r = .56$). Agreement that AI is appropriate for

reviewing patient data and helping generate data report feedback was, on average, near the scale midpoint (M = 4.44, SD = 1.76), with substantial variability across participants.

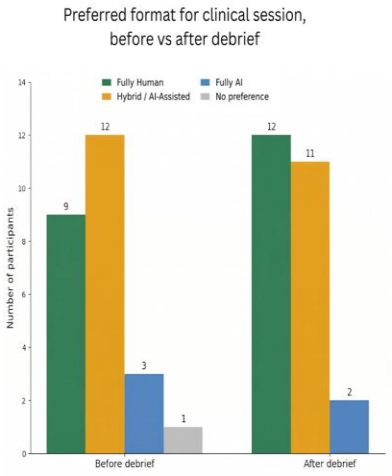


**Fig. 6** Preferred format for a clinical session, in response to the question "If you were receiving this service in a clinical setting, which type of session would you prefer?", asked before debriefing (while participants were blinded to format identity) and again after debriefing, once AI involvement in each format had been disclosed

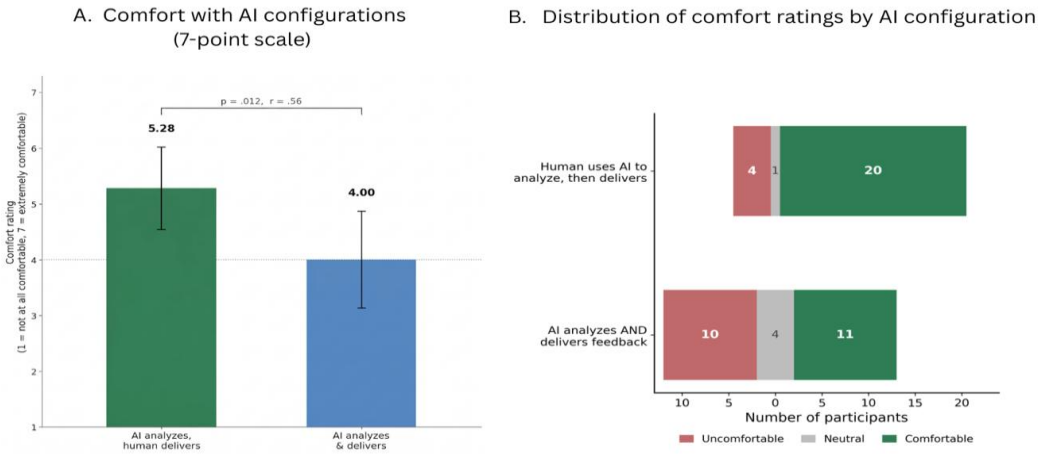


**Fig. 7** A) Comfort with two uses of AI (a human navigator using AI to analyze the data then delivering the feedback vs. AI analyzing the data and delivering the feedback directly), rated on a 7-point comfort scale (1 = not all comfortable, 7 = extremely comfortable). Bars are group means; whiskers are 95% confidence intervals. The dotted

line marks the scale midpoint (4 = neutral). B) Distribution of participant responses categorized as uncomfortable (ratings 1–3), neutral (rating 4), and comfortable (ratings 5–7). Participants reported greater comfort when a human remained the point of delivery than when AI both analyzed and delivered feedback (Wilcoxon signed-rank test, p = .012, r* = .56)

Individual-difference analyses indicated that receptivity to AI was associated with general attitude toward AI. Higher ratings of the fully AI format were associated with greater comfort with AI ($\rho$ = .55, $p$ = .004) and stronger endorsement of its appropriateness in healthcare ($\rho$ = .46, $p$ = .021). More favorable baseline attitudes toward AI were associated with higher ratings of the fully AI format ($\rho$ = .36, $p$ = .080), but not the human or hybrid formats. Older participants rated the fully AI format more favorably ($\rho$ = .41, $p$ = .040) and exhibited a smaller preference gap between the human and fully AI formats ($\rho$ = −.42, $p$ = .039). Qualitative findings complemented these results: participants valued AI for generating clear, concise, and comprehensive data summaries but perceived it as impersonal and one-directional, favoring a workflow in which AI analyzes data while humans deliver and contextualize the findings. Although a minority preferred AI for its perceived objectivity, several indicated a preference for a written AI-generated report over spoken, anthropomorphic delivery (Table 2).

**Table 2** Qualitative themes from open-ended responses

| Theme | Description |
|---|---|
| Human connection and reciprocal conversation | Human and hybrid delivery were valued as a responsive, back-and-forth exchange; whereas fully AI delivery was likened to a monologue. |
| AI's analytic strengths | AI was valued for clear, concise, and comprehensive data summaries. |
| AI as impersonal / "talking at" the patient | Fully AI delivery was criticized as emotionally flat and one-directional, leaving no room for the patient. |
| Trust, privacy, and appropriateness | Concerns centered on data privacy and a belief that mental health care intrinsically requires human connection. |
| Preference for hybrid format | Many endorsed AI analyzing the data with a human delivering and interpreting it. |

**Discussion**

These findings suggest a clear and consistent pattern: the acceptability of digital mental health feedback depends less on who or what generates it than on whether a human remains involved in its delivery. AI-assisted feedback that was drafted by AI but edited and delivered by a person was rated comparably to fully human feedback, whereas fully AI-generated feedback delivered via synthetic voice was rated substantially less favorably. The consistency of this pattern across acceptability ratings, forced-choice preferences, session preferences, and post-debrief responses

strengthens confidence in the findings despite the small sample size, although the participant-level forced-choice comparison did not reach significance.

The most relevant finding was the near-equivalence between hybrid and fully human care. The two formats differed by 0.07 scale points, with a confidence interval spanning less than half a point in either direction. This suggests that AI can be integrated into existing care workflows without reducing patients' perceived quality of care, provided it complements rather than replaces human interaction. Participants did not view AI as inherently acceptable or unacceptable. Instead, their acceptance depended on how AI was used. They welcomed AI for analytic, behind-the-scenes tasks but were less comfortable with its use in the relational and interpretive aspects of care that they viewed as central to the therapeutic relationship. Consistent with this pattern, participants were substantially more comfortable when a human used AI to generate content and then delivered the feedback. This preference persisted even after explicit disclosure of AI involvement, suggesting that human oversight is a prerequisite for trust, consistent with prior research (Lee et al., 2025).

These findings also reframe the design challenge. The least preferred condition, the fully AI-generated format, was also the most ambitious, attempting to replicate human interaction through a spoken, direct monologue. Participants often described this experience as being "talked at" rather than engaged in conversation, and several explicitly stated they would prefer AI-generated feedback as a written report rather than a simulated spoken interaction. Combined with the strong preference for the hybrid format, these findings suggest that greater anthropomorphism does not necessarily increase acceptability. Rather than mimicking a human, AI may be better suited to analyzing data and generating a clear written summary that a DHN can interpret and deliver.

These findings also suggest a practical approach to using AI in digital mental health care. Highly conversational, anthropomorphic AI systems are relatively expensive to develop, validate, and maintain, and they may create relational expectations that patients typically reserve for humans (Hussein et al., 2026). In contrast, simpler AI systems could analyze patient data and generate draft summaries while leaving interpretation and communication to the DHN. This approach may be less costly, easier to govern, and more scalable without reducing patient acceptability. For hybrid clinics seeking to expand DHN capacity while preserving the human relationship that supports engagement, this human-in-the-loop model may offer an effective balance of scalability, cost, and patient acceptance.

Participants' responses to AI were also internally consistent. Those who rated the fully AI format more favorably reported greater comfort with AI and stronger beliefs that AI is appropriate in healthcare. More favorable baseline attitudes toward AI were also positively, though not significantly, associated with ratings of the fully AI format, with no similar trend for the human or hybrid formats. Age showed a similar pattern: it was associated with ratings of the fully AI format, but not with ratings of the human or hybrid formats, and this association remained when baseline AI attitude was controlled for. This finding was unexpected and contrasts with the common assumption that younger individuals are more receptive to automation (Fazi et al., 2025). Because the differences between these correlations were not statistically significant, the apparent specificity to fully AI delivery should be treated as a hypothesis rather than a finding. If replicated, this pattern may have a practical implication: individual differences may matter most when AI operates autonomously, whereas human-delivered formats may be acceptable across patients regardless of their prior attitudes toward AI. This supports preserving a human role in the delivery of care while using AI to assist with underlying clinical tasks.

These findings suggest that AI should be used primarily for data analysis and summarization while preserving DHNs as the primary point of relational contact. Explicit disclosure of AI involvement may also support, rather than undermine, patient trust as the proportion of participants preferring a format that kept human involvement remained essentially unchanged after AI use was disclosed. Within the Digital Clinic, a practical first step may be to integrate AI-generated data summaries that are reviewed and delivered by DHNs, providing an incremental, low-risk approach that appears acceptable to patients. Future research may evaluate these models using patients' own data in real-world clinical workflows, examining the effects of AI modality, and assessing engagement, therapeutic alliance, workforce efficiency, and implementation outcomes in addition to acceptability.

These findings should be interpreted cautiously. The analyses were exploratory and based on a small pilot sample. Participants evaluated hypothetical vignettes rather than their own data, and the sample was drawn from a single site and consisted primarily of highly engaged patients. Outcomes were limited to cross-sectional self-reported acceptability and did not include behavioral or longitudinal measures. The study was powered to detect the large differences observed between human-involved and fully AI delivery. A larger study, with a more robust sample and formal qualitative analysis, will provide an opportunity to confirm and extend these preliminary findings.

## Conclusion

In this mixed-methods study, former and current mental health patients found AI-generated data-report feedback acceptable when delivered under human oversight. AI-assisted delivery achieved acceptability comparable to fully human delivery and was consistently preferred over fully AI delivery, with both comfort and preference depending on continued human involvement. Qualitative findings further supported a model in which AI performs data analysis and summarization while humans deliver and contextualize the findings, favoring simpler, scalable AI integrated into the existing DHN workflow over more complex, anthropomorphic systems. If replicated in a larger sample, these findings support designing AI to augment, rather than replace, the human relationships that sustain engagement in mental health care.

**References**


Baier, A. L., Kline, A. C., & Feeny, N. C. (2020). Therapeutic alliance as a mediator of change: A systematic review and evaluation of research. *Clinical Psychology Review, 82*, 101921. https://doi.org/10.1016/j.cpr.2020.101921

Benda, N., Desai, P., Reza, Z., et al. (2024). Patient perspectives on AI for mental health care: Cross-sectional survey study. *JMIR Mental Health, 11*, e58462. https://doi.org/10.2196/58462

Bond, R. R., Mulvenna, M. D., Potts, C., et al. (2023). Digital transformation of mental health services. *npj Mental Health Research, 2*, 13. https://doi.org/10.1038/s44184-023-00033-y

Bracic, A., Spector-Bagdady, K., Towle, S., Zhang, R., James, C. A., & Price, W. N., II. (2026). Factors for patient trust and acceptance of medical artificial intelligence. *JAMA Network Open, 9*(3), e260815. https://doi.org/10.1001/jamanetworkopen.2026.0815

Calvert, E., Cipriani, M., Chen, K., Dhima, A., Burns, J., & Torous, J. (2025). Evaluating clinical outcomes for anxiety and depression: A real-world comparison of the digital clinic and primary care. *Journal of Affective Disorders, 377*, 275–283. https://doi.org/10.1016/j.jad.2025.02.051

Chen, K., Lane, E., Burns, J., Macrynikola, N., Chang, S., & Torous, J. (2024). The digital navigator: Standardizing human technology support in app-integrated clinical care. *Telemedicine and e-Health, 30*(7), e1963–e1970. https://doi.org/10.1089/tmj.2024.0023

Fazi, L., Zaniboni, S., & Wang, M. (2025). Age differences in the adoption of technology at work: A review and recommendations for managerial practice. *Journal of Organizational Change Management, 38*(8), 138–175. https://doi.org/10.1108/JOCM-12-2024-0767

Flathers, M., Xia, W., Hau, C., et al. (2025). Interpreting psychiatric digital phenotyping data with large language models: A preliminary analysis. *BMJ Mental Health, 28*(1), e301817. https://doi.org/10.1136/bmjment-2025-301817

Fortney, J. C., Unützer, J., Wrenn, G., et al. (2017). A tipping point for measurement-based care. *Psychiatric Services, 68*(2), 179–188. https://doi.org/10.1176/appi.ps.201500439

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Henson, P., Wisniewski, H., Hollis, C., Keshavan, M., & Torous, J. (2019). Digital mental health apps and the therapeutic alliance: Initial review. *BJPsych Open, 5*(1), e15. https://doi.org/10.1192/bjo.2018.86

Hua, Y., Na, H., Li, Z., et al. (2025). A scoping review of large language models for generative tasks in mental health care. *npj Digital Medicine, 8*, 230. https://doi.org/10.1038/s41746-025-01611-4

Hussein, R., Zink, A., Ramadan, B., et al. (2026). Advancing healthcare AI governance through a comprehensive maturity model based on systematic review. *npj Digital Medicine, 9*(1), 236. https://doi.org/10.1038/s41746-026-02418-7

Jin, Y., Liu, J., Li, P., et al. (2025). The applications of large language models in mental health: Scoping review. *Journal of Medical Internet Research, 27*, e69284. https://doi.org/10.2196/69284

Lee, H. S., Wright, C., Ferranto, J., et al. (2025). Artificial intelligence conversational agents in mental health: Patients see potential, but prefer humans in the loop. *Frontiers in Psychiatry, 15*, 1505024. https://doi.org/10.3389/fpsyt.2024.1505024

Lewis, J. R. (2018). The System Usability Scale: Past, present, and future. *International Journal of Human–Computer Interaction, 34*(7), 577–590. https://doi.org/10.1080/10447318.2018.1455307

Löchner, J., Carlbring, P., Schuller, B., Torous, J., & Sander, L. B. (2025). Digital interventions in mental health: An overview and future perspectives. *Internet Interventions, 40*, 100824. https://doi.org/10.1016/j.invent.2025.100824

Macrynikola, N., Nguyen, N., Lane, E., Yen, S., & Torous, J. (2023). The digital clinic: An innovative mental health care delivery model utilizing hybrid synchronous and asynchronous treatment. *NEJM Catalyst Innovations in Care Delivery, 4*(9). https://doi.org/10.1056/CAT.23.0100

Rahimi, B., Nadri, H., Lotfnezhad Afshar, H., & Timpka, T. (2018). A systematic review of the technology acceptance model in health informatics. *Applied Clinical Informatics, 9*(3), 604–634. https://doi.org/10.1055/s-0038-1668091

Rodriguez-Villa, E., Rauseo-Ricupero, N., Camacho, E., Wisniewski, H., Keshavan, M., & Torous, J. (2020). The digital clinic: Implementing technology and augmenting care for mental health. *General Hospital Psychiatry, 66*, 59–66. https://doi.org/10.1016/j.genhosppsych.2020.06.009

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Design and evaluation of a short version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence, 4*(6), 103–108. https://doi.org/10.9781/ijimai.2017.09.001

Sekhon, M., Cartwright, M., & Francis, J. J. (2017). Acceptability of healthcare interventions: An overview of reviews and development of a theoretical framework. *BMC Health Services Research, 17*, 88. https://doi.org/10.1186/s12913-017-2031-8

Smith, K. A., Ward, T., Lambe, S., et al. (2025). Engagement and attrition in digital mental health: Current challenges and potential solutions. *npj Digital Medicine, 8*, 398. https://doi.org/10.1038/s41746-025-01778-w

Vaidyam, A., Halamka, J., & Torous, J. (2022). Enabling research and clinical use of patient-generated health data (the mindLAMP platform): Digital phenotyping study. *JMIR mHealth and uHealth, 10*(1), e30557. https://doi.org/10.2196/30557

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2016). Unified theory of acceptance and use of technology: A synthesis and the road ahead. *Journal of the Association for Information Systems, 17*(5), 328–376. https://doi.org/10.17705/1jais.00428

Zidaru, T., Morrow, E. M., & Stockley, R. (2021). Ensuring patient and public involvement in the transition to AI-assisted mental health care: A systematic scoping review and agenda for design justice. *Health Expectations, 24*(4), 1072–1124. https://doi.org/10.1111/hex.13299