Patient-centered communication preferences in AI-powered mental health chatbots: evidence from two preregistered studies

Katharina Angermayr, Nathalie Laura Neuendorf, Sebastian Scherr

Angaben zur Veröffentlichung / Publication details:

Angermayr, Katharina, Nathalie Laura Neuendorf, and Sebastian Scherr. 2026. "Patient-centered communication preferences in AI-powered mental health chatbots: evidence from two preregistered studies." *Health Communication*. https://doi.org/10.1080/10410236.2026.2666885.

Health Communication

ISSN: 1041-0236 (Print) 1532-7027 (Online) Journal homepage: www.tandfonline.com/journals/hhth20

Patient-Centered Communication Preferences in AI-Powered Mental Health Chatbots: Evidence from Two Preregistered Studies

Katharina Angermayr, Nathalie Laura Neuendorf & Sebastian Scherr

To cite this article: Katharina Angermayr, Nathalie Laura Neuendorf & Sebastian Scherr (15 Jun 2026): Patient-Centered Communication Preferences in AI-Powered Mental Health Chatbots: Evidence from Two Preregistered Studies, Health Communication, DOI: 10.1080/10410236.2026.2666885

To link to this article: https://doi.org/10.1080/10410236.2026.2666885

HEALTH COMMUNICATION
https://doi.org/10.1080/10410236.2026.2666885

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Patient-Centered Communication Preferences in AI-Powered Mental Health Chatbots: Evidence from Two Preregistered Studies

Katharina Angermayr [a,b], Nathalie Laura Neuendorf [a,b], and Sebastian Scherr [a,b]

[a]Center for Interdisciplinary Health Research, University of Augsburg; [b]Department of Media, Knowledge, and Communication, University of Augsburg

ABSTRACT

Access to mental health information is shifting from static search to conversational AI. Guided by patient-centered communication (PCC), two preregistered U.S. studies identified preferred communication features for interactions with AI chatbots about mental health and how individuals trade them off within feature bundles. Study 1 ($N = 414$, US quota sample) used a Best–Worst Scaling (BWS) to identify the six most relevant PCC-aligned features for healthcare providers. Study 2 analyzed an AI chatbot subsample ($n = 268$) drawn from a U.S. quota-representative sample in a Discrete Choice Experiment (DCE) to quantify trade-offs between combinations of these preferred features. Across both studies, users strongly wanted two communication features simultaneously in AI mental-health chatbots: *reflective listening* and *multi-symptom assessment*. Importantly, relational and clinical PCC-aligned features are most highly valued in interactions with AI mental-health chatbots. These preferences remained largely consistent across users and their preferences for communication accommodation.

Digital access to mental health information is transforming as users shift from navigating search engine hyperlinks to receiving immediate, synthesized AI-generated responses, exemplified by tools such as Google's "AI Overviews" and Bing's "Copilot Answer." Although these tools deliver mainly static, one-off responses, this evolution points toward conversational AI systems (e.g., AI chatbots) enabling ongoing interaction and new possibilities for personalization, empathy, and tailored communication (Esmaeilzadeh et al., 2025). This shift highlights not only access to facts but also the quality of patient-centered communication (PCC), the gold standard for in-person medical encounters (Niu et al., 2025). PCC involves validating patients' perspectives, interpreting them in psychological and social contexts, fostering trust and shared understanding, and supporting evidence-based decisions aligned with patient preferences (Epstein & Street, 2007).

PCC principles are rarely operationalized as communicative design features in AI systems. With few exceptions (e.g., Angermayr et al., 2025), most studies examine isolated features such as empathetic tone (Kang & Ki, 2025), nonjudgmental style (Jang et al., 2021), or factual accuracy (Cornelison et al., 2024) rather than combined configurations. Only a small subset of experimental health AI chatbot studies implement communicative design features, which typically reflect partial rather than comprehensive PCC models (Qin et al., 2025). As a result, mental health AI research has largely overlooked how users evaluate co-occurring communicative features, leaving open how trade-offs between feature combinations shape user-centered preferences and inform AI chatbot communication design.

To address this research gap, we draw on PCC and Communication Accommodation Theory (CAT; Giles, 1973) as guiding frameworks and combined Best–Worst Scaling (BWS; Study 1; Louviere et al., 2015; Schuster et al., 2024) to identify preferred PCC-aligned features of healthcare providers and a Discrete Choice Experiment (DCE; Study 2; Szinay et al., 2021) to explore trade-offs between those features. Both studies were preregistered using a repository preregistration [weblink]. In Study 1, we used a quota sample of the U.S. population ($N = 414$) to assess the importance of communication attributes when seeking mental health advice, and in Study 2, we used an AI chatbot subsample ($n = 268$) of an overall U.S. sample ($N = 1{,}011$) to examine trade-offs between combinations of those preferred attributes when seeking mental health advice from an AI chatbot.

The promises and pitfalls of AI chatbots in digital mental health care

People increasingly seek mental health information online via search engines for quick, anonymous answers (Esmaeilzadeh et al., 2025), social media for peer dialogue (Calvin et al., 2024), and online communities with forums and clinician consultations (Liu & Wang, 2021). These channels increase flexibility and reduce stigma (Kong et al., 2025) but lack tailored conversation (Esmaeilzadeh et al., 2025) and just-in-time guidance (von Lützow et al., 2025). AI chatbots extend care with adaptive real-time, anonymous, low-cost support, personalize pathways (Zhong et al., 2024), and generate consistent, context-aware, multimodal dialogue (Scherr et al., 2025).

CONTACT Katharina Angermayr katharina.angermayr@uni-a.de Center for Interdisciplinary Health Research & Department of Media, Knowledge, and Communication, University of Augsburg, Universitaetsstr. 10, Augsburg 86159, Germany

Supplemental data for this article can be accessed online at https://osf.io/t69rx/

AI mental health chatbots such as Woebot (Fitzpatrick et al., 2017) or Wysa (Inkster et al., 2018) show positive impact by measurably reducing depression, anxiety, and stress. They improve psychoeducation and health literacy (Guo et al., 2024; Sallam, 2023), reduce stigma via empathetic, anonymous interaction (Bhatt, 2024; Lawrence et al., 2024), and support symptom identification, (early) diagnosis, and monitoring (Guo et al., 2024; Olawade et al., 2024). Evidence on conversational quality aligns with these gains: Clinicians preferred ChatGPT's responses to human ones in 78.6% of cases for quality and empathy (Ayers et al., 2023). Patients judged AI more empathetic (Chen et al., 2025), with advice often matching human standards (Kuhail et al., 2025). AI motivates therapy (Sallam, 2023), personalizes care, improves decision-making (Guo et al., 2024; Lawrence et al., 2024), increases efficiency, and saves costs (Sallam, 2023). While less personal (Saha et al., 2025), AI replies are more readable, professional, and practical (Lopes et al., 2024). Effectiveness rises with human-like embodiment, multimodality, and empathetic tone (Jang et al., 2021; Lim et al., 2022).

However, AI chatbots face limitations, including repetitive dialogs, restricted functionalities, lack of novelty (Laymouna et al., 2024), plus generic suggestions requiring clinicians (Scholich et al., 2025). Safety is constrained by inadequate crisis responses, limited personal understanding and reliability issues (Guo et al., 2024; Sallam, 2023). Opacity and missing sourcing raise ethical and legal concerns regarding privacy, security, and accountability (Guo et al., 2024; Olawade et al., 2024). Uncorrected algorithmic bias and confirmation loops can yield unfair outcomes (Hasanzadeh et al., 2025; Lopez-Lopez et al., 2025), while hallucinations undermine trust and agency (Asgari et al., 2025). Additional risks include oversimplification, outdated data, and overreliance on LLMs (Guo et al., 2024). Overall, the value of AI mental health chatbots depends less on automation than on communication quality, motivating a shift from a technology-centered to a *communication*-centered perspective.

Patient-centered communication as design principle for AI mental-health chatbots

For mental health communication, a core challenge is moving beyond generic, one-size-fits-all messaging toward tailored, individualized, human-like interactions. This maps onto *patient-centered communication* (PCC; Epstein & Street, 2007), a collaborative process in which clinicians and patients share information, understand perspectives, and coordinate care around specific concerns, preferences, and needs. PCC in psychotherapy strengthens therapeutic alliance (Flückiger et al., 2018), improves outcomes, reduces dropout (Swift et al., 2018), and increases satisfaction (Elwyn et al., 2017). It enhances interpersonal functioning (Muir et al., 2021), advances skills via experiential learning (Liao et al., 2022), and extends to teletherapy and online communities (Farber & Ort, 2024; Niu et al., 2025; van der Eijk et al., 2013).

PCC elements are inherently relational and context-dependent (Epstein & Street, 2007). From this perspective, AI systems cannot fully achieve PCC because they lack human intentionality, moral agency, and experiential understanding. However, communicative effectiveness depends less on the system's objective qualities than on *users' perceptions of interaction quality* (Reeves & Nass, 1996). Consistent with the *Computers Are Social Actors* (CASA) framework, people apply social rules to machines (Nass & Moon, 2000; Reeves & Nass, 1996). Therefore, and most importantly, PCC in human–AI interaction can be functionally approximated through language-based and structural design features (Qin et al., 2025), with varying success across dimensions. They excel at technical competence (e.g., structured assessments; Goodman et al., 2023), simplified explanations (Stephan et al., 2025) and information quality (Laranjo et al., 2018; Walker et al., 2023). Relational goals require adaptation via empathetic language (Seitz, 2024), reflective listening (Xue et al., 2023), and discourse cues (Klein, 2025), yet sustained engagement and accountability remain difficult. Consequently, this study is guided by PCC to examine user-expected communication attributes rather than assuming human equivalence. At the same time, we recognize perceived attribute value may vary across users based on accommodative preferences and communicative expectations.

What patients value in clinician-patient communication: evidence from preference studies in the patient–clinician encounter

PCC sets quality criteria, but its practical meaning is anchored in patients value. Decision-makers seek to understand how patients prioritize what matters (Hollin et al., 2022). *Priorities* are topics accorded greater relative *importance* (i.e., insights that become essential when time and resources are limited), which are central to shared decision-making (Tinetti et al., 2019) and can inform patient-centered outcomes (Edgman-Levitan & Schoenbaum, 2021). PCC guides systematic identification and organization of communication attributes. Consequently, we synthesize evidence on the relative importance of clinician–patient communication attributes to derive a concrete set of contextualized PCC attributes.

Evidence from preference studies shows a consistent pattern: relational and clinical quality dominate patient priorities (Burton et al., 2017; Cheraghi-Sohi et al., 2008; Gerard et al., 2012; Hole, 2008; Tinelli et al., 2015; Whitaker et al., 2017). *Relational continuity* (i.e., seeing a known clinician who understands one's history) builds trust, problem-specific knowledge, and perceived safety while reducing diagnostic and communicative friction (Chudner et al., 2025; Hole, 2008; Lagarde et al., 2015; Tinelli et al., 2015; von Weinrich et al., 2024; Whitaker et al., 2017). *Communication and information quality* (i.e., being actively listened to, treated with respect, receiving clear, tailored explanations, and involvement in shared decisions; Burton et al., 2017; Cheraghi-Sohi et al., 2008; Gerard et al., 2012; Mengoni et al., 2013; Tinelli et al., 2015) drive reassurance, understanding, and adherence. Patients often prioritize these relational elements over time or consultation mode (Buchanan et al., 2021; von Weinrich et al., 2024; Whitaker et al., 2017). *Technical quality and clinical competence* remain crucial, with diagnostic certainty superseding convenience (Tinelli et al., 2015). *Logistical factors* (e.g., appointment time) are subordinate. Consultation mode often proxies perceived quality or safety unless strong continuity and timely access exist (Buchanan et al., 2021; Hole, 2008; Lagarde et al., 2015; von Weinrich et al., 2024). Reputation is a distant signal, while direct experience through continuity tends to dominate (Buchanan et al., 2021; Lagarde et al., 2015). Price matters less in low or zero marginal cost systems, shifting focus to quality and continuity (Lagarde et al., 2015). In-visit waiting time and consultation length are valued for enabling unhurried, high-quality interaction, not as ends in themselves (Gerard et al., 2012; Hole, 2008; Whitaker et al., 2017). Downstream conveniences are secondary to a strong clinical conversation (Buchanan et al., 2021). Taken together, PCC's communication attributes may be relevant when human medical resources and attention are limited. Understanding the relative importance of communication attributes to patients can help inform the design of current and future AI systems. Accordingly, we explore the following research question:

> RQ1: Which attributes reflecting key dimensions of patient-centered communication are preferred by patients as most important when looking for mental health advice?

Trade-offs between preferred feature combinations of mental health AI chatbots

AI chatbots, though unable to replicate the full human exchange of interpersonal care, can embody PCC elements via design *features*. Following Anderson and Robey (2017), these are system-level properties that, combined with the user's ability and context, shape the conversation's delivery. The strength with which features deliver intended communication qualities in context is critical. For instance, an AI chatbot may afford empathy, but only goal-aligned outcomes (e.g., feeling understood) are actualized. If generic replies result in unmet expectations, users may abandon the technology. More recently, scholars emphasize the *interplay* of multiple features in fostering patient-centered conversational AI (Angermayr et al., 2025). In this work, we refer to these as feature bundles (i.e., a coordinated combination of design features delivering a targeted user experience or task outcome). For example, empathetic style, transparent reasoning, and safe escalation combine to yield trusted guidance. The task is to explore which feature bundle configurations are perceived as delivering stronger patient-centered experiences.

Research distinguishes stated-preference and choice-based studies (Louviere et al., 2000). Evaluating system profiles allows trading off combined features, yielding weights for communication features. This addresses gaps where prior work prioritized technical features (Abd-Alrazaq et al., 2020; Ahmed et al., 2022), or examined isolated features like empathy (Seitz, 2024) via non-generalizable qualitative methods (Haque & Rubya, 2023). Unlike studies assuming additive effects, these capture feature interdependence and distinct bundles (Furnari et al., 2021; Liu et al., 2022). A preference-based approach offers a way to capture trade-offs among multiple communication features and identify preference patterns across feature combinations for actionable, patient-centered AI chatbot design.

Only a few preference studies assess trade-offs in preferences for health AI chatbots. Wang, Wang, et al. (2025) found users prioritize practical, accurate information and utility. Zhang et al. (2025) reported young adults value symptom-specific results, video content, and clear language. Zheng et al. (2025) showed South African students value language flexibility, security, and personalized advice. Mayer et al. (2024) observed that while efficiency aids appeal, face-to-face contact remains preferred for sensitive contexts as trust declines with digitalization. These findings suggest AI chatbot preferences blend universal principles with contextual factors (e.g., culture). Critically, mental-health AI chatbots must be designed to meet contextual needs and patient-driven solutions (Bond et al., 2023), requiring derivation from multiple attribute trade-offs. We therefore ask:

> RQ2: How do patients trade off combinations of preferred communication features in mental health conversations with an AI chatbot?

Communication accommodation preferences as a moderator

While the CASA paradigm explains why users apply social heuristics to AI, PCC specifies which communicative features are plausibly relevant for high-quality health communication. Both frameworks often assume uniform responses. However, users may evaluate conversational AI as bundled configurations, trading off relational, informational, and safety features under constraints of time, effort, and attention (Ramaul et al., 2024). Consistent with evidence of heterogeneity in attribute weighting (Vass et al., 2022), this variation may reflect differences in users' communicative dispositions and expectations.

In attribute testing, a "supportive" design feature (e.g., reflective listening) does not have fixed utility. Its value depends on users' desire for adaptation. To capture this, we draw on Communication Accommodation Theory (CAT; Giles, 1973) as a guiding framework to examine the fit between an AI chatbot's algorithmic adjustment and users' goals. CAT predicts communicative moves increase value when aligned with preferred styles (convergence) and decrease value when misaligned (divergence; Dragojevic et al., 2015). CAT strategies including discourse management, interpretability, approximation, interpersonal control, and emotional expression improve clinician–patient outcomes (Farzadnia & Giles, 2015) and show comparable effects in human–machine interaction (e.g., Shen & Wang, 2023).

In AI chatbots, we conceptualize accommodation as algorithmic adjustment: system-level output variations within technical constraints aligning with user needs or pace. Reflecting discourse modification rather than intentionality (Dragojevic et al., 2015), AI adjustment follows design logic or optimization patterns. Nevertheless, users apply social heuristics and respond to communicative adjustment in human–machine interaction (Shen & Wang, 2023). Accordingly, accommodation is treated as a perceived response quality. CAT explains why identical PCC-aligned features may yield divergent evaluations based on users' desire for alignment.

Accordingly, we define adaptive cues as AI output features reflecting attentiveness or guidance relative to user situations. They are not a distinct attribute set, because many features reflect structural properties rather than communicative adjustment (Sundar et al., 2015). Per CAT, cue value depends on users' accommodation preferences (Dragojevic et al., 2015): accommodation-oriented users interpret cues as uncertainty-reducing, whereas efficiency-oriented users may find them unnecessary, yielding divergent evaluations of identical PCC-aligned features. Given unclear preferences for communication attributes in AI-mediated mental health conversations, we explore:

> RQ3: Do users´ preferences for communication accommodation moderate how AI chatbot features shape users' trade-off decisions?

Ethical considerations

The university IRB approved both empirical studies (IRB-A-2025-x101).

Study 1: Best–Worst Scaling for identifying the importance of patient-centered communication attributes

The aim of Study 1 was to identify which attributes reflecting key dimensions of patient-centered communication (PCC) patients consider most important when seeking mental health advice (RQ1). Study 1 used an independent sample to identify the most relevant patient-centered communication attributes, which informed the trade-off decisions examined in Study 2. Figure 1 provides an overview of the experimental protocol.

We recruited a heterogeneous sample of $N = 461$ U.S. adults (≥18 years) for an online survey via *Prolific*, using quota sampling for sex, age, and religious affiliation.[1] Participants received compensation on the platform. No human-verification procedures were conducted. In this preregistered study, we used Best–Worst Scaling (BWS; Louviere et al., 2015; Schuster et al., 2024), in which respondents repeatedly selected the most and least important items from randomized subsets drawn from a larger attribute pool. Compared to other conjoint approaches, BWS captures both high- and low-priority elements, elicits nuanced trade-off patterns, supports diverse quantitative analyses (Cheung et al., 2019) and is widely used in health research (Hollin et al., 2022).

Method

Sample and procedure

To ensure data quality, we incorporated two attention-check items (e.g., "I would select 'strongly agree' to show that I'm paying attention to this question"). A total of $n = 47$ participants failed one of these checks and were excluded from the analysis, leading to a total sample of $N = 414$. All participants were U.S. residents with a mean age of 39.83 years (SD = 13.60). The sample comprised $n = 207$ women (50.0%) and n = 207 men (50.0%). Most participants reported holding a college degree or higher ($n = 258$; 62.3%), and the majority identified as White ($n = 265$; 64.0%) or Black/African American ($n = 52$; 12.6%). Demographic details for Study 1 are presented in Table 1.

After informed consent, demographics, and instructions, participants were randomly assigned to a mental health scenario (anxiety, depression, or bipolar disorder). They completed three modules, each with six BWS tasks, using a balanced incomplete block design (Schuster et al., 2024). Each module covered a distinct subset of communication attributes from a pool of 24, applicable across three modes

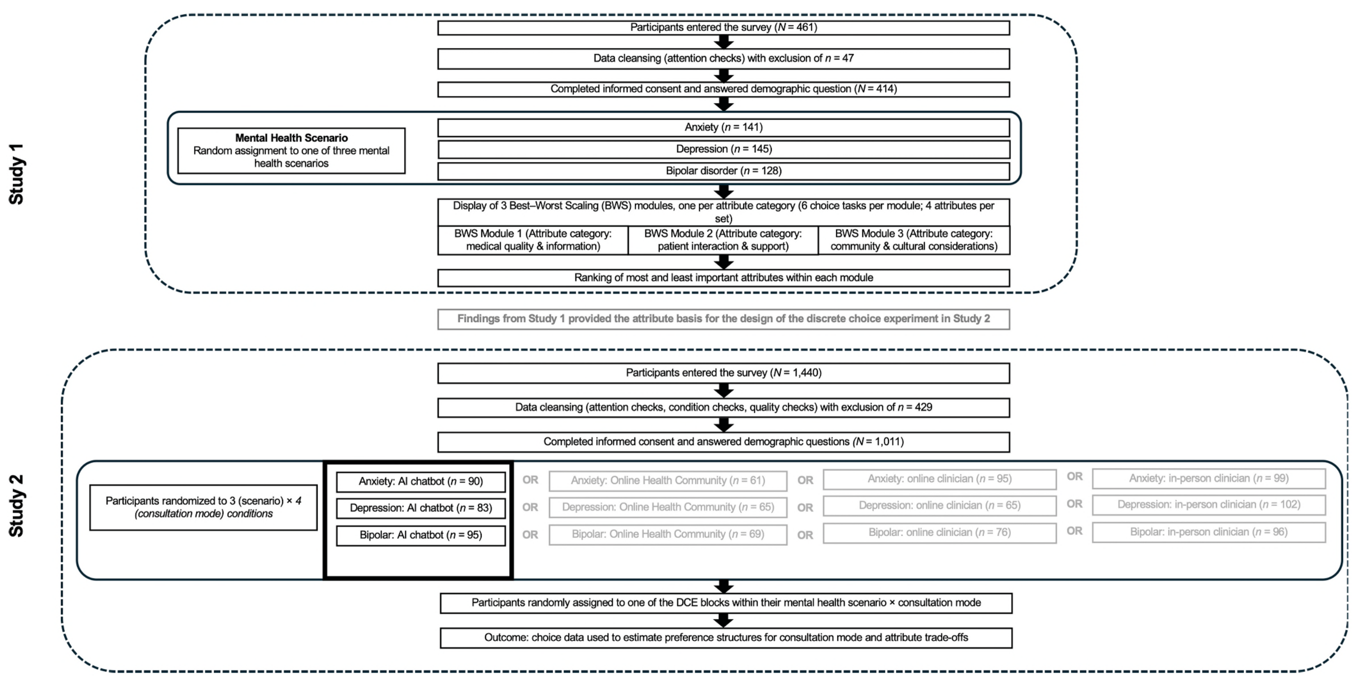

Figure 1. Experimental protocol and participant flow for Study 1 and Study 2. *Note.* The figure illustrates the experimental protocol and participant flow for Study 1 and Study 2.

Table 1. Socio-demographic statistics of Study 1 (BWS; $N = 414$) and AI chatbot subsample of Study 2 (DCE; $n = 268$).

	Study 1 (BWS)		Study 2 (DCE)	
	M (*SD*)	%	*M* (*SD*)	%
Sex				
Female		50.0%		57.1%
Male		50.0%		42.9%
Age	39.83 (13.59)		45.59 (14.24)	
18–30y		30.9%		15.7%
31–40y		25.1%		27.2%
41–50y		22.5%		21.3%
51–60y		11.6%		19.0%
61y and older		9.9%		16.8%
Race				
White/Caucasian		64.0%		67.5%
Hispanic, Latino or Spanish origin		5.1%		6.7%
Black or African American		12.6%		15.3%
Asian		10.6%		4.9%
American Indian or Alaska Native		0.2%		0.7%
Arab/West Asian		4.3%		0.4%
Native Hawaiian or Other Pacific Islander		3.1%		–
Education				
Less than a high school diploma		0.5%		0.7%
High school graduate/diploma or equivalent (e.g., GED)		9.9%		12.7%
Some college but no degree or Associate Degree		27.3%		29.9%
Bachelor's Degree (e.g., BA) or Master's Degree (e.g., MA)		48.8%		53.4%
Professional School Degree (e.g., MD) or Doctorate (e.g., PhD)		13.5%		3.4%
Mental Health Status				
Depression				
Yes		56.0%		50.4%
No		42.3%		48.1%
Prefer not to answer		1.7%		1.5%
Anxiety				
Yes		60.6%		60.4%
No		38.6%		38.8%
Prefer not to answer		0.7%		0.7%
Bipolar Disorder				
Yes		6.5%		4.9%
No		92.0%		94.0%
Prefer not to answer		1.4%		1.1%

Note. The table shows descriptive statistics for both samples after final data cleaning (Study 1, BWS: $N = 414$; Study 2, DCE AI chatbot Subsample: $n = 268$). BWS sample used quotas for sex, age, and religious affiliation. The DCE AI chatbot subsample was drawn from a larger sample that was quota-sampled to approximate U.S. population distributions for sex, age, and race/ethnicity based on simplified Census categories. Mental health status reflects whether respondents reported having ever been experienced the respective condition at any point in their lifetime. Percentages may not sum to 100% due to rounding.

(AI chatbot, online community, online clinician). Respondents selected the most and least important attribute per set (selections could not coincide), ensuring broad pairwise coverage, manageable cognitive load, and reduced response bias. The survey took about 20 minutes. Figure S1 presents an example study task [https://osf.io/t69rx/files/ameu9].

Study design

Attribute Selection. Following guidance on attribute selection (Szinay et al., 2021), we derived an initial attribute pool of 24 attributes from the relevant PCC literature synthesized in the theory section (see Table 2) applicable across three consultation modes (AI chatbot, online health community, online clinician). Attributes were organized into three categories: (1) medical quality and information, (2) patient interaction and support, and (3) community and cultural considerations. We standardized wording, removed duplicates, retained only quantifiable attributes, and excluded mode-incomparable items (e.g., AI chatbot waiting time). To capture underexplored dimensions, we added self-developed attributes on patient-centered language (e.g., medical jargon), cultural/spiritual guidance (Puchalski et al., 2014), and community awareness/connection (Laidsaar-Powell et al., 2016; Van Oerle et al., 2018). References documenting the derivation of each attribute are reported in Table S1 [https://osf.io/t69rx/files/2kzfs].

Development of mental health scenarios as BWS context. To ensure relevance and realism, we developed three mental health scenarios (anxiety, depression, bipolar disorder) reflecting the most searched U.S. psychiatric conditions (Minn et al., 2025) and corresponding to global prevalence rankings (World Health Organization, 2022). Scenarios (generated with GPT-4; OpenAI, 2025) were standardized for tone, clarity, and length, describing symptoms and care-seeking rationale. Although not analyzed as a factor, scenario variability was included to enhance real-world generalizability. Figure S2 provides all scenarios [https://osf.io/t69rx/files/gmj5c]. Formal robustness checks revealed no evidence of systematic preference differences across mental health scenarios within any of the three attribute categories tested in the BWS (all joint likelihood-ratio tests $p > .05$). Detailed analysis is reported in Table S2 [https://osf.io/t69rx/files/q2ayt] and Table S3 [https://osf.io/t69rx/files/qzjv2].

Measures

BWS attribute preferences. In each BWS task, respondents reviewed four attributes with brief descriptions and selected the *most and least important* for medical advice; selections were coded as most (+1), least (−1), and unselected (0).

Mental health status. To explore whether identified attributes vary by mental health status, we included lifetime prevalence indicators for depression, anxiety, and bipolar disorder, assessed via self-report items adapted from large-scale digital health research (Valla et al., 2025). Participants indicated whether they had ever experienced each condition (0 = no, 1 = yes) or refused to answer.

Data analysis

Data were analyzed using three complementary methods in R (4.5.0; R Core Team, 2025). *Count analysis* computes each attribute's best-worst score (ranging from −1 to 1; higher values indicating greater perceived importance) as the difference between "most" and "least important" selections divided by total appearances (Cheung et al., 2019). The *multinomial logit model* (MNL) applies dual coding (most/least important = 1; otherwise = 0) to estimate propensity scores and coefficients (β) reflecting relative preference versus a reference attribute (Cheung et al., 2019; Mühlbacher et al., 2016). The *mixed logit model* (MIXL)

Table 2. Attribute rankings from Best–Worst Scaling count analysis ($N = 414$).

Attributes	Description	Ranking	Best Count	Worst Count	BWS Score
Medical quality & information					
Explanation about medicine	**How clearly medical information (diagnoses, treatment, medication) is explained**	**1**	**631**	**65**	**.46**
Symptom assessment capabilities	**Extent to which the consultation mode can assess symptoms**	**2**	**601**	**59**	**.44**
Reputation	Quality and trustworthiness based on patient reviews and ratings	3	338	85	.20
Extent of information provided	Amount of information shared with the patient during the consultation	4	299	52	.20
Financial cost	Costs that must be paid for the consultation	5	364	225	.11
Length of response	Amount of detail provided in each response during the consultation	6	127	445	−.26
Frequency of information provided	How often information is shared with the patient during the consultation	7	97	528	−.35
Use of medical jargon	Extent to which complex medical terminology is used	8	27	1025	−.80
Patient interaction & support					
Attentiveness and listening skills	**How well the patient is listened to, how well the patient's concerns are acknowledged**	**1**	**506**	**56**	**.36**
Scope and help provided	**Extent to which the consultation includes both diagnoses and advice or only general guidance**	**2**	**504**	**205**	**.24**
Shared decision making	Extent to which the patient actively participates in treatment decisions, with concerns considered	3	404	162	.20
Personalization	Extent to which medical communication is tailored to individual patients	4	306	310	−.003
Biopsychosocial care	Extent to which not only physical symptoms but also social and emotional factors are considered	5	283	290	−.01
Empathy and emotional support	How strong the patient's emotional state is recognized and responded to	6	269	320	−.04
Interpersonal communication style	How strong the interaction with the patient is, including the manner and communicating approach	7	110	416	−.25
Familiarity with healthcare provider	Extent to which the patient knows and feels comfortable with the healthcare provider	8	102	723	−.50
Community and cultural considerations					
Extended support beyond consultation	**Availability of additional help beyond the consultation (assistance, e.g., through communities)**	**1**	**869**	**68**	**.64**
Community based assistance	**Patient connection to community resources and support networks**	**2**	**439**	**131**	**.25**
Cultural competence and awareness	Extent to which cultural factors, alongside biomedical factors, are considered and integrated	3	391	167	.18
Community awareness	Extent to which community-specific health concerns during the consultation are considered	4	257	164	.07
Family involvement	Extent to which family members are included in medical decision-making	5	242	253	−.01
Culturally sensitive language	Extent to which communication to accommodate language is adapted	6	89	336	−.20
Religious-based treatment recommendations	Whether medical recommendations are adapted to align with religious beliefs	7	81	632	−.44
Religious or spiritual guidance	Whether religious or spiritual beliefs are considered in the consultation to support decision-making	8	116	733	−.50

Note. The table lists the 24 final attributes included in the Best–Worst Scaling (BWS) experiment, each accompanied by a brief, participant-friendly description as presented to respondents. Attributes are grouped into three thematic categories, with eight attributes per category. The table reports results from the BWS count analysis; the BWS score reflects the relative frequency with which attributes were selected as "best" versus "worst."

allows normally distributed random parameters (500 Halton draws) to estimate mean coefficients (β) and standard deviations (η); heterogeneity is significant when $\eta \neq 0$ (Cheung et al., 2019). Both logit models were estimated using *Nlogit* in R (4.5.0; Greene, 2016).

Results

Preferred attributes related to patient-centered communication (RQ1)

The two highest-ranked attributes from each category were selected for inclusion in the subsequent DCE. Across all three data analytic approaches the rankings were stable with only minor variations. Table S4 provides detailed results of all models [weblink]. Given this convergence, we relied on the count analysis for interpretation as the most transparent and widely used approach in BWS (see Table 2; Cheung et al., 2019; Louviere et al., 2015).

For medical quality and information, the most important attributes were *explanation about medicine* (*BWS score* = .46) and *symptom assessment capabilities* (*BWS* score = .44). For patient interaction and support, the highest ranked attributes were *attentiveness and listening skills* (*BWS score* = .36) and *scope and help provided* (*BWS score* = .24). For community and cultural considerations, the preferred attributes were *extended support beyond consultation* (*BWS score* = .64) and *community-based assistance* (*BWS score* = .25).

Addressing RQ1, patients most strongly preferred accurate symptom assessment, clear explanation about medication, attentiveness and listening skills, and extended and community-based support.

Exploratory moderation of preferred attributes by mental health status

Exploratory moderation analyses tested whether dummy-coded lifetime experience with anxiety, depression, or bipolar disorder moderated attribute preferences. After Bonferroni correction ($\alpha = .008$), there were no moderating influences of mental health status. Full details are reported in Table S8 [https://osf.io/t69rx/files/jg9dv].

Discussion of Study 1

Study 1 utilized Best–Worst Scaling (BWS; Louviere et al., 2015; Schuster et al., 2024) to explore preferred patient-centered communication attributes. Users prioritize consultations that *assess* (multi-symptom capability), *explain* (reason-giving information), *attend* (visible listening), *guide* (concrete next steps), *continue* (aftercare), and *embed* (community resources). These findings can be organized into three themes. First, information quality tends to outrank convenience; symptom assessment reflects clinical competence and diagnostic certainty, which patients rarely trade for convenience (Cheraghi-Sohi et al., 2008; Lagarde et al., 2015; Tinelli et al., 2015), while explanation is central to shared decision-making (Burton et al., 2017; Cheraghi-Sohi et al., 2008; Gerard et al., 2012). Second, relational conduct seems specific; attentiveness and listening drive trust and adherence, often outweighing logistical factors (Cheraghi-Sohi et al., 2008; Gerard et al., 2012; Whitaker et al., 2017). Scope of help signals commitment to comprehensive care (Lagarde et al., 2015; Tinelli et al., 2015). Third, patients appear to look beyond the single encounter, expressing preferences for continuity and aftercare to build safety and trust (Cheraghi-Sohi et al., 2008; Lagarde et al., 2015; Mengoni et al., 2013), alongside community-based assistance highlighting networked social support (Laidsaar-Powell et al., 2016; Van Oerle et al., 2018).

Study 2: trade-offs between AI chatbot features

The aim of Study 2 was to explore how patients trade off combinations of preferred communication features in mental-health conversations with an AI chatbot (RQ2) and to examine whether users' preferences for communication accommodation moderate how AI chatbot feature levels shape users' trade-off decisions (RQ3). Study 2 is based on a larger Discrete Choice Experiment (DCE; Szinay et al., 2021), which included a total of four experimental conditions (AI chatbot, online health community, online clinician, and in-person clinician), of which only the AI chatbot condition is of focal interest (see Figure 1). We recruited a quota-balanced U.S. adult sample (N = 1,440) via *Prolific* to meet a priori power requirements.[2] Participants received compensation on the platform. No human-verification procedures were conducted.

Respondents repeatedly chose between systematically varied attributes and levels (i.e., alternatives), enabling estimation of trade-offs and preferences. Communicative attributes from the BWS (Study 1) were combined into feature bundles to test integrated effects in AI mental-health conversations. DCEs simulate absent real-world choices, quantify trade-offs, structures, and interactions, surpass rankings or experts to optimize chatbot feature deployment (Ho et al., 2025).

Method

Sample and procedure

Data quality procedures included three attention checks (e.g., "I would select 'strongly agree' to show that I'm paying attention to this question"), one condition check ("Earlier in the study, which healthcare provider type were you told to evaluate?"), and five response-quality items (e.g., "I clicked on anything occasionally to finish faster"). Using the overall sample, all participants who failed at least one attention check were excluded first ($n = 36$). From the remaining cases, participants who did not correctly complete the condition check were excluded next ($n = 387$). Finally, participants who did not meet the response-quality requirements (*score* ≥ *5* on the five response-quality items) were excluded ($n = 6$). These procedures resulted in a final sample of N = 1,011 participants. As the focus here is on AI chatbots in the context of mental health, we only included a subsample randomly assigned to the AI chatbot condition ($n = 268$) in Study 2.

All participants were U.S. residents between 18 and 83 years old ($M = 45.59$; $SD = 14.24$). The sample included n = 153 women (57.1%) and $n = 115$ men (42.9%). Most participants reported having a college degree or higher ($n = 152$; 56.7%), and the majority identified as White ($n = 181$; 67.5%) or Black/African American ($n = 41$; 15.3%). Regarding their self-reported mental health history, 50.4% of participants indicated having experienced depression at some point in their lives, 60.4% reported a history of anxiety, and 4.9% reported having experienced bipolar disorder. Demographics for Study 2 are also presented in Table 1. Information about the overall sample is provided in Table S5 [https://osf.io/t69rx/files/mt27x].

A dropout analysis was conducted within the AI chatbot condition. Inclusion was coded binary ($n_{excl} = 51$; $n_{incl} = 268$). Predictors of inclusion were examined via logistic regression, identifying lower age ($p = .033$) and higher discourse management ($p = .027$) as significant predictors. Full statistical results are reported in Table S6 [https://osf.io/t69rx/files/k54gh].

Individuals completed an online discrete choice survey after informed consent, demographics, and measures of technology affinity, trust in technology, and communicative preferences. In a between-subject design, participants were randomly assigned to one of three mental health scenarios (from Study 1) and one of four consultation modes (AI chatbot, online community, online clinician, in-person clinician). Each completed one randomly assigned block of eight tasks; blocking reduced burden and preserved balance (Szinay et al., 2021). Tasks presented two alternatives with six attributes at varied levels. Survey duration was approximately 20 minutes. Figure S3 shows an example study task [https://osf.io/t69rx/files/8bfsd].

Study design

Establishing attributes and levels. The initial attribute selection (Szinay et al., 2021) was completed through Study 1. In line with clinician–patient DCEs (Burton et al., 2017; Cheraghi-Sohi et al., 2008; Chudner et al., 2025; Gerard et al., 2015), we operationalized each attribute (A) at two levels (L).

The two levels were content-specific and conceptually grounded (see Table 3), consistent with DCE guidance (Mariel et al., 2025), and chosen to streamline comparisons, avoid ambiguous mid-categories, and improve attribute salience and decision quality relative to alternative DCE approaches (Hensher et al., 2015; Szinay et al., 2021). No

Table 3. The attributes and attribute levels included in the Discrete Choice Experiment.

Category	Attributes	Attribute levels
Medical quality & information	Explanation about medicine	• Brief overview (0) • Detailed explanation (1)
	Symptom assessment capabilities	• Basic screening (0) • Multi-symptom evaluation (1)
Patient interaction & support	Attentiveness and listening skills	• No user reference (0) • Reflective listening (1)
	Scope and help provided	• Low scope (0) • High scope (1)
Community and cultural considerations	Extended support beyond consultation	• One-time interaction (0) • Ongoing follow-ups/support (1)
	Community-based assistance	• No support contacts (0) • Direct contact with health community (1)

Note. The attributes and levels shown were included in a Discrete Choice Experiment (DCE; Study 2) with a U.S.-representative sample by sex, age, and race/ethnicity (simplified Census categories; $N = 1{,}011$). Attribute selection was informed by a Best-Worst Scaling (BWS; Study 1) with a U.S. quota sample ($N = 414$). Each attribute had two levels (low and high). The "(0)/(1)" labels indicate the dummy coding used (low = 0; high = 1).

additional attribute descriptions were provided to keep attribute wording concise and easily interpretable and minimize cognitive burden (Szinay et al., 2021).

Choice tasks. Choice tasks used a full profile with six attributes to enable trade-offs, as partial profiles limit observations (Szinay et al., 2021) and six attributes are cognitively manageable (Mühlbacher & Johnson, 2016). Forced-choice without opt-out preserved preference inference and efficiency (Mariel et al., 2025). Attribute sets, order, and levels were held constant across conditions for comparability.

Experimental design. We implemented an unlabeled, fractional-factorial DCE with six attributes (A) at two levels (L). Rather than the full-factorial $L^A = 2^6 = 64$ tasks, we used an efficient subset and presented alternatives as "Option A" and "Option B" to minimize label bias and isolate attribute trade-offs (Szinay et al., 2021). We generated the design in R (4.5.0) using the package idefix (Traets et al., 2020). After a pretest with $N = 10$ U.S. participants to check task comprehension and wording, we ran a pilot with $N = 150$ U.S. participants (which included the pretest sample) to estimate priors. We then produced a Bayesian D-efficient design with 2000 Halton draws that balanced attribute presentation and co-occurrence, yielding 24 choice sets arranged into three blocks of eight with random assignment. Attributes were effect-coded (±1). The design achieved global and block-level balance and met accepted quality thresholds, with a Bayesian D-error of 0.064, Bayesian D-efficiency of 15.68, utility balance of .036, no dominant or duplicate alternatives, and choice probabilities centered near 0.5.

Measures

DCE attribute preference. Preference was measured as a pairwise choice between two profiles (alternative 1 vs. alternative 2) differing on multiple communication attributes, each experimentally set at two levels (low, high). For each choice set, respondents chose their preferred option for medical advice. Responses were coded binary (0 = option 1, 1 = option 2). Attributes were coded within alternatives (low [reference] = 0, high = 1). Choosing option 2 does not assign attributes to 1; models estimate how high levels affect choice probability.

Patient preference for communication accommodation. Individuals' preferences for communication accommodation during medical consultations were measured with 26 items on a 7-point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*) reflecting five CAT sociolinguistic strategies (Giles, 1973): approximation (5 items; e.g., "Adapts their communication style to my needs"), discourse management (6 items; e.g., "Treats me as an equal"), emotional expression (4 items; e.g., "Reassures me"), interpretability (5 items; e.g., "Handles the conversation competently"), and interpersonal control (6 items; e.g., "Does not intrude on my privacy"). Sixteen items were adapted from Watson and Gallois (1998), six from the Patient-Centered Communication Scale (Moser et al., 2022), informed by Epstein and Street (2007), and four approximation items were self-developed (Dragojevic et al., 2015; Farzadnia & Giles, 2015). Mean scores were computed, with higher values indicating stronger accommodation preferences.

Internal consistency was high for *approximation* ($\alpha = .877$, $M = 5.54$, $SD = 1.12$), *discourse management* ($\alpha = .862$, $M = 6.31$, $SD = 0.77$), *emotional expression* ($\alpha = .870$, $M = 5.54$, $SD = 1.21$), and *interpretability* ($\alpha = .864$, $M = 6.32$, $SD = 0.77$). Due to the low internal consistency of the *interpersonal control* subscale (Cronbach's $\alpha = .455$), we excluded it from the primary analyses. We did not present exploratory factor analyses as these were not preregistered.

Mental health status. We included the same dummy indicators for mental health status (lifetime depression, anxiety, and bipolar disorder) as in Study 1.

Data analysis

Choice models were estimated in R (4.5.0; R Core Team, 2025). We first estimated a conditional logit model (CLM) with fixed coefficients (i.e., one common set of preference weights for all respondents) using clogit (Therneau, 2024) with dummy-coded attributes (high = 1, low (reference)). Preference heterogeneity was modeled with a mixed logit (MIXL) with random coefficients (allowing them to vary across individuals and reporting the population mean) using the mlogit (Croissant, 2025), specifying normally distributed, correlated random coefficients for the six attributes and no intercept, so coefficients represent high vs. low. Estimation used 2,000 Halton-free draws. Thus, the CLM identifies average preference weights, whereas the MIXL captures between-person heterogeneity. Interaction effects between attributes and z-standardized CAT subscale indices were estimated within the MIXL.

Results

Preference structure for AI chatbot communication attributes (RQ2)

To address RQ2, we estimated choice models to quantify how patients trade off combinations of preferred communication features in AI mental-health chatbot conversations. We first estimated a conditional logit model (CLM) with dummy-coded attributes (0 = low (reference), 1 = high) as a baseline. Relative to the reference levels, each higher-level attribute was positively associated with choice in the CLM. Model fit improved on *AIC* with a mixed logit (MIXL; $AIC = 2737.50$ vs. CLM $AIC = 2741.32$), while BIC was unchanged (both 2764.10). Estimated main effects are the unstandardized logit coefficients (β) for the high level of each attribute relative to its reference (attributes coded 0 = reference and 1 = high); a positive β indicates a higher probability of choosing options that include that level, and values may exceed one.

Individuals showed the strongest preference for *reflective listening* ($\beta = 1.26$, $SE = 0.18$, $p < .001$) and *multi-symptom evaluation* ($\beta = 1.25$, $SE = 0.19$, $p < .001$). *Direct contact with a health community* also increased choice likelihood ($\beta = 1.09$, $SE = 0.17$, $p < .001$), as did *ongoing follow-ups/support* ($\beta = 0.87$, $SE = 0.16$, $p < .001$), *detailed explanation* ($\beta = 0.83$, $SE = 0.15$, $p < .001$), and *high scope and help provided* ($\beta = 0.62$, $SE = 0.13$, $p < .001$). Figure 2 shows the MIXL mean coefficients (β) with 95% CIs for the AI chatbot subgroup ($n = 268$). Table S7 reports CLM and MIXL estimates for the AI chatbot condition in detail [https://osf.io/t69rx/files/dcwq2].

Finally, Utility-equivalent trade-off ratios (β_i/β_j) were computed from MIXL mean coefficients to quantify attribute-on-attribute trade-offs (Figure 3). Interpreted as marginal rates of substitution (MRS), these define the amount of one attribute required to compensate for losing one unit of another (Hensher et al., 2015). Ratios indicate lower relative value, while ratios indicate equal importance, showing how many level changes of the traded attribute (0 → 1) equal one change in the preferred attribute.

Focusing on the strongest trade-offs, respondents were willing to trade 2.02 units of *scope & help* for one unit of *attentiveness & listening*, and 2.01 units of *scope and help* for one unit of *symptom assessment capabilities*.

CAT-Based interaction effects on attribute utilities (RQ3)

RQ3 asks whether users' preferences for communication accommodation moderate how AI chatbot feature levels

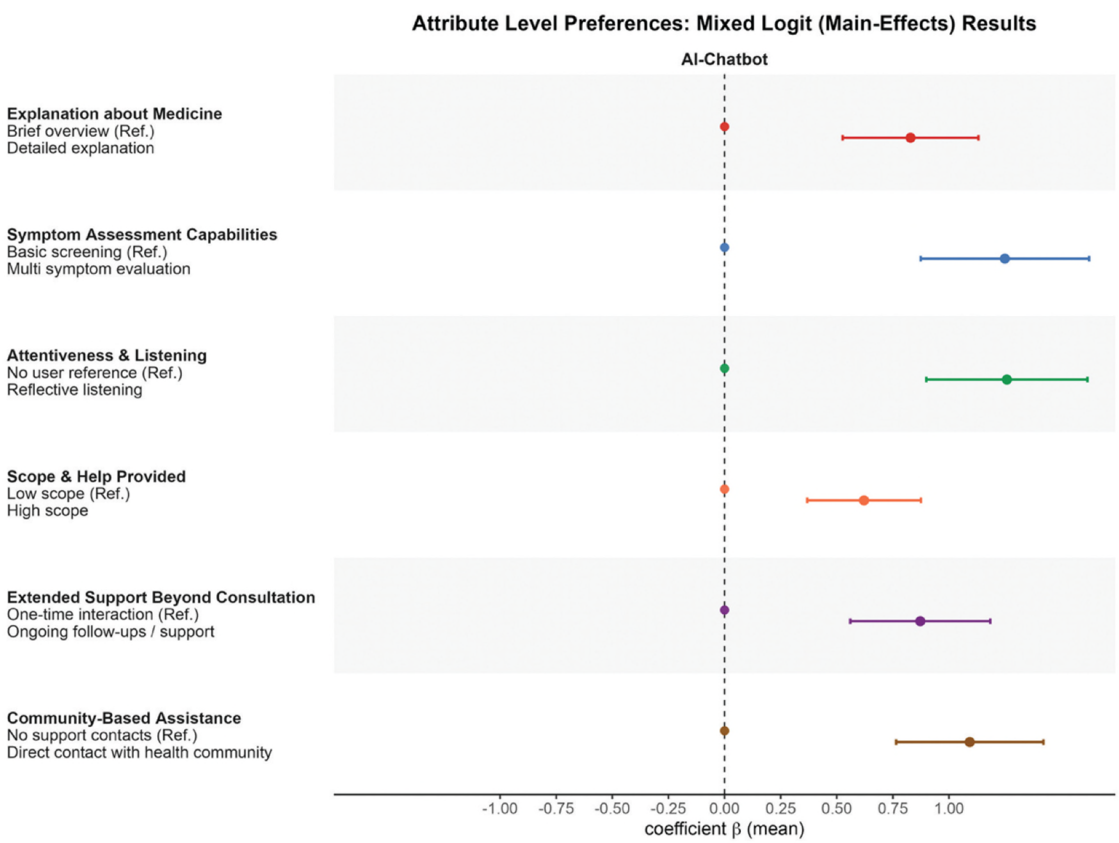

Figure 2. Estimated attribute-level preferences from the mixed logit model (main effects) for AI chatbot. *Note.* ($n = 268$). The figure shows estimated main effects (β) from a Mixed Logit (MIXL) for the AI chatbot subgroup. The dependent variable is the chosen alternative (1 vs. 2). Each attribute had two levels and was dummy-coded (Low = 0, High = 1). Points are the mean random-coefficient estimates (β) for the high level relative to the low (reference) level with 95% confidence intervals; the dashed vertical line marks $\beta = 0$ (no difference between high and low). Positive β values indicate greater log-odds of choosing alternatives featuring the high level; odds ratios = exp(β) give the multiplicative change in those odds. Figure created in R 4.5.0 using ggplot2; layout adapted from Reinders et al. (2025).

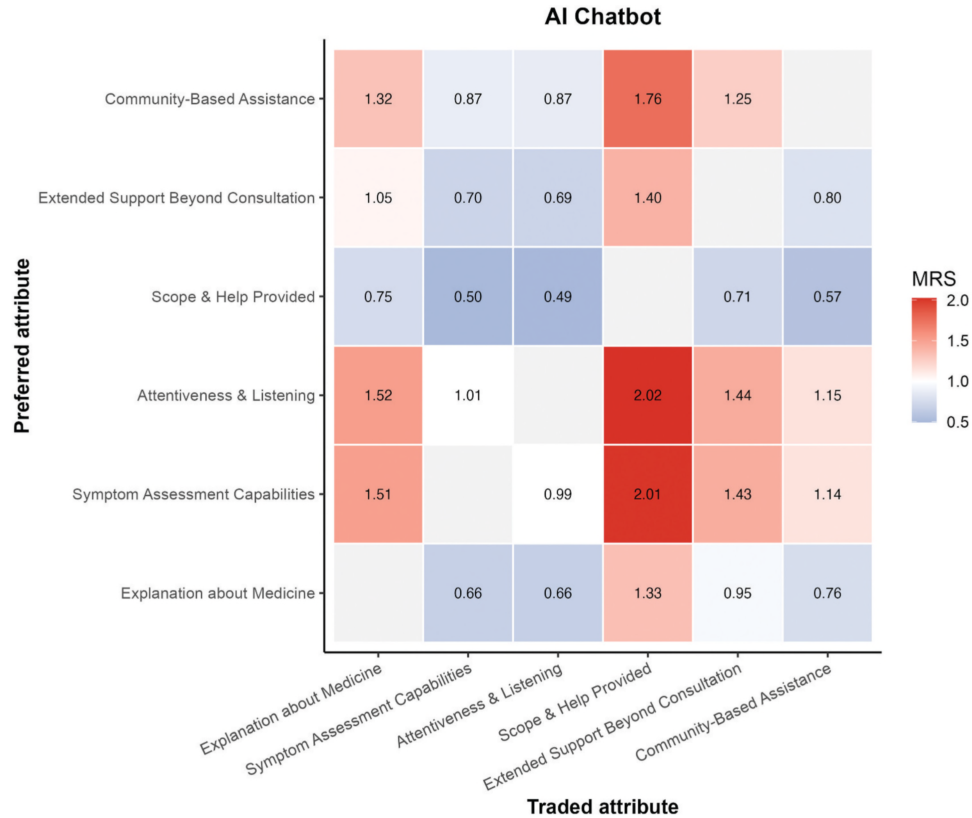

Figure 3. Trade-off matrix based on mixed logit estimates for AI chatbot. *Note.* ($n = 268$). The figure shows trade-off ratios computed as β_i / β_j from MIXL mean coefficients for the AI chatbot subgroup. We interpret these ratios as marginal rates of substitution (MRS), defined as the amount of one attribute required to exactly compensate for the loss of one unit of another (Hensher et al., 2015). Attributes correspond to the DCE attributes. Estimates come from a mixed logit with random main effects. Each cell indicates how many level changes of the traded attribute (0 → 1) respondents would give up obtaining one level change of the preferred attribute (0 → 1). Higher values mean a stronger relative preference for the preferred attribute. Values are signed; colors encode magnitude. Figure created in R 4.5.0 with ggplot2.

shape users' trade-off decisions. We estimated interactions between the four CAT preference indices with sufficient reliability and the six AI chatbot attributes, yielding 24 interaction tests (Table 4).

After Bonferroni correction for multiple comparisons ($\alpha\_adj = .002$), none of the interaction effects remained statistically significant; results were consistent when applying Benjamini–Hochberg false discovery rate correction (all $p > .05$).

Exploratory moderation analyses of trade-off decisions by mental health status

An exploratory moderation analysis tested whether dummy-coded lifetime experience with anxiety, depression, or bipolar disorder moderated the attribute selection in Study 2. Findings showed no significant differences after Bonferroni correction ($\alpha = .008$). Full details are reported in Table S9 [https://osf.io/t69rx/files/gmytu].

Table 4. Interaction matrix: CAT preference indices × AI chatbot attributes ($n = 268$).

Attribute *CAT preferences*	Explanation about medicine	Symptom assessment capabilities	Attentiveness and listening skills	Scope and help provided	Extended support beyond consultation	Community-based assistance
Approximation	−.17	−.08	.05	.07	−.06	−.08
Discourse management	.08	.66	.14	.41	.18	.02
Emotional expression	.22	−.32	.10	−.23	.16	.05
Interpretability	.23	−.16	−.13	−.20	.12	.13

Note. $n = 268$. The table reports interaction effects from a mixed logit (MIXL) model estimating profile choice in two-option tasks. Each AI chatbot attribute had two levels and was dummy coded (low/reference = 0; high = 1). Random main effects were specified for the six attributes (normally distributed, correlated random coefficients; no intercept), and fixed interaction terms captured moderation by users' Communication Accommodation Theory (CAT) strategy indices. CAT strategies were assessed using 1–7 Likert scales (mean score) with good internal consistency: approximation (Cronbach's $\alpha = .877$), discourse management ($\alpha = .862$), emotional expression ($\alpha = .870$), and interpretability ($\alpha = .864$). Strategy indices were entered as z-standardized continuous moderators. Rows list the CAT strategy indices; columns list the AI chatbot attributes. Cells show the interaction coefficient β for the high attribute level versus its low (reference) level. A positive β indicates that, as the CAT strategy index increases by one *SD*, respondents are more likely (higher log-odds) to choose options featuring the high level of that attribute; a negative β indicates the opposite. Coefficients are presented without standard errors in the table; asterisks mark Bonferroni-corrected significance (* $p < .05$, ** $p < .01$, *** $p < .001$). Prior to correction for multiple testing, significant interaction effects were observed for discourse management × symptom assessment capabilities and discourse management × scope and help provided. Given 24 interaction tests (6 attributes × 4 CAT strategies), none of the interaction effects remained statistically significant after Bonferroni correction ($\alpha\_adj = .05/24 = .00208$). Results were consistent when applying Benjamini–Hochberg false discovery rate (FDR) correction (all $p > .05$).

Discussion of Study 2

Study 2 suggests *reflective listening* and *multi-symptom evaluation* may form a single preferred value, conceptualized as dual assurance: relational assurance (recognition) and epistemic assurance (holistic assessment), which may work together to reduce social and informational uncertainty (Epstein & Street, 2007). Finally, no significant moderation effects were observed for any CAT preference dimension or mental health status. Dropout analysis showed higher completion among younger participants and those with higher discourse management, a core CAT strategy in which alignment of conversational flow with users' expectations facilitates more coherent and well-structured interactions. Older users' attrition may reflect lower technology familiarity (Poli et al., 2019) while discourse management likely supports coherent interaction and sustained engagement (Marconi et al., 2026).

General discussion

In two independent, preregistered U.S. studies, we identified PCC-aligned communication features that users tend to prefer for AI mental-health chatbot interactions and estimated how users trade off combinations of these features when seeking mental health advice. We further examined whether preferences for accommodative communication styles during medical consultations moderate observed preferences. Using a quota sample of U.S. adults and a subsample from an independent, nationally representative panel, we conducted a Best–Worst Scaling (BWS; Louviere et al., 2015; Schuster et al., 2024) study and a Discrete Choice Experiment (DCE; Szinay et al., 2021).

Across both studies, *reflective listening* and *multi-symptom evaluation* were the most preferred patient-centered features, reflecting preferences for both relational and epistemic assurance. Findings are broadly consistent with PCC principles (Epstein & Street, 2007), suggesting participants value relational and clinical features when evaluating hypothetical AI chatbot interactions. In conversational AI, *reflective listening* refers to an AI chatbot's ability to interpret open-ended user input and explicitly reflect its semantic content and implied meaning before offering guidance. This involves paraphrasing user statements, articulating inferred emotions or needs, summarizing key themes, and inviting further elaboration, thereby signaling understanding, conveying empathy, and helping users feel heard (Xiao et al., 2020). A practical implication is to design AI chatbots to initiate with a brief, structured symptom assessment covering core symptoms, duration, severity, and basic risk indicators, followed by reflective listening with a concise paraphrase and explicit verification prompt. This approach aligns with symptom reductions (Fitzpatrick et al., 2017), readable and empathetic responses (Ayers et al., 2023), and plain-language structure (Saha et al., 2025).

Both studies suggest that combined feature profiles—rather than isolated elements—play an important role in how users evaluate. Prior work often examines single features (e.g., factural accuracy (Cornelison et al., 2024); nonjudgmental style (Jang et al., 2021; empathetic tone (Kang & Ki, 2025)), whereas our results suggest preferences for profiles including reflective listening and multi-symptom assessment (see also Angermayr et al., 2025). Users trade off attributes, prioritizing these relational and clinical capabilities over breadth of scope and other presentation features, while some elements (e.g., detailed explanation vs. direct community contact) appear roughly equivalent. In practice, users may encounter highly empathetic AI chatbots (Ayers et al., 2023) yet accept reduced breadth or general support to secure high-quality relational attentiveness and clinical multi-symptom assessment. This suggests prioritization of relational and clinical competence (Chen et al., 2025; Laymouna et al., 2024) and that evaluations reflect a complex exchange across features rather than any single capability. These patterns may offer an initial basis for informing message design, evaluation, usability, and trust.

Moderation analyses indicated no significant moderation by accommodative preferences or mental health status, indicating broadly consistent preferences. This may reflect strong main effects for core PCC-aligned features and shared expectations (Epstein & Street, 2007). Because CAT concerns perceived cues rather than forced trade-offs, CAT-based preferences may be more evident in subjective evaluations than in discrete choice behavior (Dragojevic et al., 2015; Vass et al., 2022). Consistency aligns with eHealth evidence of stable communication preferences across conditions and treatment histories (Phillips et al., 2021).

However, the present studies assess stated preferences, not downstream PCC outcomes, and do not establish human-AI equivalence. While DCEs infer preference structures, participants made no conscious trade-offs. Future research should examine whether PCC-aligned AI features—configured with transparency and escalation safeguards—translate into communication processes and therapeutic outcomes across settings and modalities.

Limitations

This research has several limitations. First, it relies on a scenario-based design covering only three mental health conditions, constraining external generalizability; future studies could include physical or chronic diseases. Second, data were based on hypothetical decision situations; although DCEs approximate real preferences, they cannot capture real-life behavior (Quaife et al., 2018). Third, the study focused on interactive text-based AI chatbots, excluding audiovisual formats (e.g., voice assistants). Fourth, replications beyond the U. S., especially in the Global South and outside health contexts, are essential (Scherr et al., 2025). Additionally, missing DCE attribute definitions may have caused partially uninformed attribute trade-offs. Further, potential construct overlap between CAT and PCC may mean some indices reflect general communication quality rather than accommodation-specific preferences (Wang, Min, et al., 2025); consistent with this, interpersonal control items did not form a reliable scale. Finally, while attention checks were used to ensure data quality, no human-verification procedures (e.g., Affonso, 2026) were implemented to prevent bot-generated responses.

Conclusion

In two preregistered studies, we identified *reflective listening* and *multi-symptom evaluation* as preferred attributes when seeking mental health advice from an AI chatbot. BWS established the preferred patient-centered attributes, and DCE then examined trade-offs between features, showing that users weigh combinations depending on what a specific AI chatbot offers. Importantly, these preferences remained largely consistent across users and their preferences for communication accommodation. Taken together, these results suggest that designers of mental health AI systems should assume that users evaluate clusters of attributes at the same time and trade them off, so testing single features in isolation misses how people actually decide.

Notes

1. The sample size was determined in an a priori power analysis using the pwr package in R 4.5.0 (Champely et al., 2020), assuming an effect size of $d = 0.68$ (Burton et al., 2019, 2021), $\alpha = .05$, and 95% power, which indicated $N = 414$, accounting for random assignment to one of three scenarios.
2. A power analysis specifically for DCEs determined a required sample size of $N = 756$ individuals. We used the Johnson and Orme (2003) formula $N > 500c$ / $(t \times a)$, with $t = 8$ choice task, $a = 2$ alternatives, and $c = 2$ analysis cells, which yielded a minimum of 63 respondents per experimental group. For our 3×4 design with three mental health scenarios (depression, anxiety, bipolar disorder) and four consultation modes (AI chatbot, online health community, online clinician, in person clinician) of which we will only focus on AI chatbots in the present paper. The design required a sample of at least $12 \times 63 = 756$ respondents (Mariel et al., 2025; Szinay et al., 2021). Therefore, given our achieved sample, we are able detect even small effects, even in analyses performed within subgroups.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the University of Augsburg, the Faculty of Philosophy and Social Sciences [Grant VAF25], and the Graduate School Center.

ORCID

Katharina Angermayr http://orcid.org/0009-0003-5816-426X
Nathalie Laura Neuendorf http://orcid.org/0000-0002-5938-8578
Sebastian Scherr http://orcid.org/0000-0003-4730-1575

Data statement

Data is available from the authors upon reasonable request.

Ethical considerations

Ethical approval for both studies was obtained from the Ethics Committee of the University of Augsburg (IRB-A-2025-x101).

References

Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and safety of using chatbots to improve mental health: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *22*(7), e16021. https://doi.org/10.2196/16021

Affonso, F. M. (2026). Brief commentary: A framework for detecting AI agents in online research. *Journal of Consumer Research*. Advance online publication. https://doi.org/10.1093/jcr/ucag006

Ahmed, A., Aziz, S., Khalifa, M., Shah, U., Hassan, A., Abd-Alrazaq, A., & Househ, M. (2022). Thematic analysis on user reviews for depression and anxiety chatbot apps: Machine learning approach. *JMIR Formative Research*, *6*(3), e27654. https://doi.org/10.2196/27654

Anderson, C., & Robey, D. (2017). Affordance potency: Explaining the actualization of technology affordances. *Information and Organization*, *27*(2), 100–115. https://doi.org/10.1016/j.infoandorg.2017.03.002

Angermayr, K., Street, R. L., Jr., & Scherr, S. (2025). *Communication accommodation with AI-chatbots: Improving patient-centered communication in times of conversational AI* [Conference presentation]. 108th Annual AEJMC Conference, San Francisco, CA, United States.

Asgari, E., Montaña-Brown, N., Dubois, M., Khalil, S., Balloch, J., Yeung, J. A., & Pimenta, D. (2025). A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. *NPJ Digital Medicine*, *8*(1), 274. https://doi.org/10.1038/s41746-025-01670-7

Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., Hogarth, M., & Smith, D. M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Internal Medicine*, *183*(6), 589–596. https://doi.org/10.1001/jamainternmed.2023.1838

Bhatt, S. (2024). Digital mental health: Role of artificial intelligence in psychotherapy. *Annals of Neurosciences*, *32*(2), 117–127. https://doi.org/10.1177/09727531231221612

Bond, R. R., Mulvenna, M. D., Potts, C., O'Neill, S., Ennis, E., & Torous, J. (2023). Digital transformation of mental health services. *NPJ Mental Health Research*, *2*(1), 13. https://doi.org/10.1038/s44184-023-00033-y

Buchanan, J., Roope, L. S. J., Morrell, L., Pouwels, K. B., Robotham, J. V., Abel, L., Crook, D. W., Peto, T., Butler, C. C., Walker, A. S., & Wordsworth, S. (2021). Preferences for medical consultations from online providers: Evidence from a discrete choice experiment in the United Kingdom. *Applied Health Economics and Health Policy*, *19*(4), 521–535. https://doi.org/10.1007/s40258-021-00642-8

Burton, C. D., Entwistle, V. A., Elliott, A. M., Krucien, N., Porteous, T., & Ryan, M. (2017). The value of different aspects of person-centred care: A series of discrete choice experiments in people with long-term conditions. *BMJ Open*, *7*(4), e015689. https://doi.org/10.1136/bmjopen-2016-015689

Burton, N., Burton, M., Fisher, C., Peña, P. G., Rhodes, G., & Ewing, L. (2021). Beyond Likert ratings: Improving the robustness of developmental research measurement using best-worst scaling. *Behavior Research Methods*, *53*(5), 2273–2279. https://doi.org/10.3758/s13428-021-01566-w

Burton, N., Burton, M., Rigby, D., Sutherland, C. A. M., & Rhodes, G. (2019). Best-worst scaling improves measurement of first impressions. *Cognitive Research: Principles & Implications*, *4*(1), 36. https://doi.org/10.1186/s41235-019-0183-2

Calvin, A., Hasse, A., & Madden, M. (2024). *Getting help online: How young people find, evaluate, and use mental health apps, online therapy, and behavioral health information*. Hopelab. https://www.commonsensemedia.org/sites/default/files/research/report/2024-getting-help-online-hopelab-report_final-release-for-web.pdf

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., & Rosario, H. D. (2020). *Pwr:*

Basic functions for power analysis (Version 1.3-0) [Computer software]. https://CRAN.R-project.org/package=pwr

Chen, D., Chauhan, K., Parsa, R., Liu, Z. A., Liu, F.-F., Mak, E., Eng, L., Hannon, B. L., Croke, J., Hope, A., Fallah-Rad, N., Wong, P., & Raman, S. (2025). Patient perceptions of empathy in physician and artificial intelligence chatbot responses to patient questions about cancer. *NPJ Digital Medicine*, *8*(1), 275. https://doi.org/10.1038/s41746-025-01671-6

Cheraghi-Sohi, S., Hole, A. R., Mead, N., McDonald, R., Whalley, D., Bower, P., & Roland, M. (2008). What patients want from primary care consultations: A discrete choice experiment to identify patients' priorities. *Annals of Family Medicine*, *6*(2), 107–115. https://doi.org/10.1370/afm.816

Cheung, K. L., Mayer, S., Simon, J., De Vries, H., Evers, S. M. A. A., Kremer, I. E. H., & Hiligsmann, M. (2019). Comparison of statistical analysis methods for object case best-worst scaling. *Journal of Medical Economics*, *22*(6), 509–515. https://doi.org/10.1080/13696998.2018.1553781

Chudner, I., Drach-Zahavy, A., Madjar, B., Gelman, L., & Habib, S. (2025). Unveiling preferences in closed communities: Development of a discrete choice experiment (DCE) questionnaire to elicit Ultra-Orthodox women preferences for video consultations in primary care. *The Patient*, *18*(3), 263–277. https://doi.org/10.1007/s40271-025-00734-w

Cornelison, B. R., Erstad, B. L., & Edwards, C. (2024). Accuracy of a chatbot in answering questions that patients should ask before taking a new medication. *Journal of the American Pharmacists Association*, *64*(4), 102110. https://doi.org/10.1016/j.japh.2024.102110

Croissant, Y. (2025). *Mlogit: Multinomial logit models* (Version 1.1-3) [Computer software]. https://CRAN.R-project.org/package=mlogit

Dragojevic, M., Gasiorek, J., & Giles, H. (2015). Communication accommodation theory. In C. R. Berger, M. E. Roloff, J. Caughlin, J. P. Dillard, D. Solomon, & S. R. Wilson (Eds.), *The international encyclopedia of interpersonal communication* (pp. 1–21). Wiley. https://doi.org/10.1002/9781118540190.wbeic006

Edgman-Levitan, S., & Schoenbaum, S. C. (2021). Patient-centered care: Achieving higher quality by designing care through the patient's eyes. *Israel Journal of Health Policy Research*, *10*(1), 21. https://doi.org/10.1186/s13584-021-00459-9

Elwyn, G., Durand, M. A., Song, J., Aarts, J., Barr, P. J., Berger, Z., Cochran, N., Frosch, D., Galasiński, D., Gulbrandsen, P., Han, P. K. J., Härter, M., Kinnersley, P., Lloyd, A., Mishra, M., Perestelo-Perez, L., Scholl, I., Tomori, K., Trevena, L., & van der Weijden, T. (2017). A three-talk model for shared decision making: Multistage consultation process. *BMJ*, *359*, j4891. https://doi.org/10.1136/bmj.j4891

Epstein, R. M., & Street, R. L. (2007). *Patient-centered communication in cancer care: Promoting healing and reducing suffering*. National Cancer Institute. https://cancercontrol.cancer.gov/sites/default/files/2020-06/pcc_monograph.pdf

Esmaeilzadeh, P., Maddah, M., & Mirzaei, T. (2025). Using AI chatbots (e.g. ChatGPT) in seeking health-related information online: The case of a common ailment. *Computers in Human Behavior: Artificial Humans*, *3*, 100127. https://doi.org/10.1016/j.chbah.2025.100127

Farber, B. A., & Ort, D. (2024). Clients' perceptions of changes in their therapists' positive regard in transitioning from in-person therapy to teletherapy. *Psychotherapy Research*, *34*(5), 601–610. https://doi.org/10.1080/10503307.2022.2146544

Farzadnia, S., & Giles, H. (2015). Patient-provider interaction: A communication accommodation theory perspective. *International Journal of Society, Culture & Language*, *3*(2), 17–34.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, *4*(2), e7785. https://doi.org/10.2196/mental.7785

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, *55*(4), 316–340. https://doi.org/10.1037/pst0000172

Furnari, S., Crilly, D., Misangyi, V. F., Greckhamer, T., Fiss, P. C., & Aguilera, R. V. (2021). Capturing causal complexity: Heuristics for configurational theorizing. *Academy of Management Review*, *46*(4), 778–799. https://doi.org/10.5465/amr.2019.0298

Gerard, K., Tinelli, M., Latter, S., Blenkinsopp, A., & Smith, A. (2012). Valuing the extended role of prescribing pharmacist in general practice: Results from a discrete choice experiment. *Value in Health*, *15*(5), 699–707. https://doi.org/10.1016/j.jval.2012.02.006

Gerard, K., Tinelli, M., Latter, S., Smith, A., & Blenkinsopp, A. (2015). Patients' valuation of the prescribing nurse in primary care: A discrete choice experiment. *Health Expectations*, *18*(6), 2223–2235. https://doi.org/10.1111/hex.12193

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, *15*(2), 87–105. https://www.jstor.org/stable/30029508

Goodman, R. S., Patrinely, J. R., Stone, C. A., Zimmerman, E., Donald, R. R., Chang, S. S., Berkowitz, S. T., Finn, A. P., Jahangir, E., Scoville, E. A., Reese, T. S., Friedman, D. L., Bastarache, J. A., van der Heijden, Y. F., Wright, J. J., Ye, F., Carter, N., Alexander, M. R., Choe, J. H., & Johnson, D. B. (2023). Accuracy and reliability of chatbot responses to physician questions. *JAMA Network Open*, *6*(10), e2336483. https://doi.org/10.1001/jamanetworkopen.2023.36483

Greene, W. H. (2016). *NLOGIT* (Version 6) [Computer software]. Econometric Software, Inc.

Guo, Z., Lai, A., Thygesen, J. H., Farrington, J., Keen, T., & Li, K. (2024). Large language models for mental health applications: Systematic review. *JMIR Mental Health*, *11*(1), e57400. https://doi.org/10.2196/57400

Haque, M. D. R., & Rubya, S. (2023). An overview of chatbot-based mobile mental health apps: Insights from app description and user reviews. *JMIR mHealth and uHealth*, *11*, e44838. https://doi.org/10.2196/44838

Hasanzadeh, F., Josephson, C. B., Waters, G., Adedinsewo, D., Azizi, Z., & White, J. A. (2025). Bias recognition and mitigation strategies in artificial intelligence healthcare applications. *NPJ Digital Medicine*, *8*(1), 154. https://doi.org/10.1038/s41746-025-01503-7

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). *Applied choice analysis* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781316136232

Ho, T. Q. A., Engel, L., Le, L. K.-D., Melvin, G., Ride, J., Le, H. N. D., & Mihalopoulos, C. (2025). Discrete choice experiment versus best-worst scaling: An empirical comparison in eliciting young people's preferences for web-based mental health interventions. *The Patient*, *18*(4), 357–372. https://doi.org/10.1007/s40271-025-00739-5

Hole, A. R. (2008). Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of Health Economics*, *27*(4), 1078–1094. https://doi.org/10.1016/j.jhealeco.2007.11.006

Hollin, I. L., Paskett, J., Schuster, A. L. R., Crossnohere, N. L., & Bridges, J. F. P. (2022). Best-worst scaling and the prioritization of objects in health: A systematic review. *PharmacoEconomics*, *40*(9), 883–899. https://doi.org/10.1007/s40273-022-01167-1

Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, *6*(11), e12106. https://doi.org/10.2196/12106

Jang, S., Kim, J.-J., Kim, S.-J., Hong, J., Kim, S., & Kim, E. (2021). Mobile app-based chatbot to deliver cognitive behavioral therapy and psychoeducation for adults with attention deficit: A development and feasibility/usability study. *International Journal of Medical Informatics*, *150*, 104440. https://doi.org/10.1016/j.ijmedinf.2021.104440

Johnson, R., & Orme, B. (2003). *Getting the most from CBC*. Sawtooth Software. https://sawtoothsoftware.com/resources/technical-papers/getting-the-most-from-cbc

Kang, D., & Ki, E.-J. (2025). User needs and benefits of mental health chatbots: Text-mining analysis of mobile apps reviews. *International Journal of Human-Computer Interaction*. Advance online publication. https://doi.org/10.1080/10447318.2025.2490866

Klein, S. H. (2025). The effects of human-like social cues on social responses towards text-based conversational agents-A meta-analysis. *Humanities and Social Sciences Communications*, *12*(1), 1322. https://doi.org/10.1057/s41599-025-05618-w

Kong, M., Wang, Y., Li, M., & Yao, Z. (2025). Mechanism assessment of physician discourse strategies and patient consultation behaviors on online health platforms: Mixed methods study. *Journal of Medical Internet Research*, *27*(1), e54516. https://doi.org/10.2196/54516

Kuhail, M. A., Alturki, N., Thomas, J., Alkhalifa, A. K., & Alshardan, A. (2025). Human-human vs human-AI therapy: An empirical study. *International Journal of Human-Computer Interaction*, *41*(11), 6841–6852. https://doi.org/10.1080/10447318.2024.2385001

Lagarde, M., Erens, B., & Mays, N. (2015). Determinants of the choice of GP practice registration in England: Evidence from a discrete choice experiment. *Health Policy*, *119*(4), 427–436. https://doi.org/10.1016/j.healthpol.2014.10.008

Laidsaar-Powell, R., Butow, P., Bu, S., Charles, C., Gafni, A., Fisher, A., & Juraskova, I. (2016). Family involvement in cancer treatment decision-making: A qualitative study of patient, family, and clinician attitudes and experiences. *Patient Education & Counseling*, *99*(7), 1146–1155. https://doi.org/10.1016/j.pec.2016.01.014

Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S., & Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association: JAMIA*, *25*(9), 1248–1258. https://doi.org/10.1093/jamia/ocy072

Lawrence, H. R., Schneider, R. A., Rubin, S. B., Matarić, M. J., McDuff, D. J., & Bell, M. J. (2024). The opportunities and risks of large language models in mental health. *JMIR Mental Health*, *11*(1), e59479. https://doi.org/10.2196/59479

Laymouna, M., Ma, Y., Lessard, D., Schuster, T., Engler, K., & Lebouché, B. (2024). Roles, users, benefits, and limitations of chatbots in health care: Rapid review. *Journal of Medical Internet Research*, *26*(1), e56930. https://doi.org/10.2196/56930

Liao, F., Murphy, D., Wu, J.-C., Chang, C.-C., & Tsai, P.-F. (2022). How technology-enhanced experiential e-learning can facilitate the development of person-centred communication skills online for health-care students: A qualitative study. *BMC Medical Education*, *22*(1), 60. https://doi.org/10.1186/s12909-022-03127-x

Lim, S. M., Shiau, C. W. C., Cheng, L. J., & Lau, Y. (2022). Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: A systematic review and meta-regression. *Behavior Therapy*, *53*(2), 334–347. https://doi.org/10.1016/j.beth.2021.09.007

Liu, J., & Wang, J. (2021). Users' intention to continue using online mental health communities: Empowerment theory perspective. *International Journal of Environmental Research and Public Health*, *18*(18), 9427. https://doi.org/10.3390/ijerph18189427

Liu, X., Feng, Y., Tang, J., Hu, C., & Zhao, D. (2022). Counterfactual recipe generation: Exploring compositional generalization in a realistic scenario. *arXiv*. https://doi.org/10.48550/arXiv.2210.11431

Lopes, E., Jain, G., Carlbring, P., & Pareek, S. (2024). Talking mental health: A battle of wits between humans and AI. *Journal of Technology in Behavioral Science*, *9*(4), 628–638. https://doi.org/10.1007/s41347-023-00359-6

Lopez-Lopez, E., Abels, C. M., Holford, D., Herzog, S. M., & Lewandowsky, S. (2025). Generative artificial intelligence-mediated confirmation bias in health information seeking. *Annals of the New York Academy of Sciences*, *1550*(1), 23–36. https://doi.org/10.1111/nyas.15413

Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling: Theory, methods and applications*. Cambridge University Press. https://doi.org/10.1017/CBO9781107337855

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: Analysis and applications*. Cambridge University Press.

Marconi, L., Longo, L., & Cabitza, F. (2026). Assessing interaction quality in human-AI dialogue: An integrative review and multi-layer framework for conversational agents. *Machine Learning and Knowledge Extraction*, *8*(2), 28. https://doi.org/10.3390/make8020028

Mariel, P., Campbell, D., Sandorf, E. D., Meyerhoff, J., Vega-Bayo, A., & Blevins, R. (2025). Steps of a discrete choice experiment. In P. Mariel, D. Campbell, E. D. Sandorf, J. Meyerhoff, A. Vega-Bayo & R. Blevins (Eds.), *Environmental valuation with discrete choice experiments in R: A guide on design, implementation, and data analysis* (pp. 9–75). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-89338-4_2

Mayer, C. J., Mahal, J., Geisel, D., Geiger, E. J., Staatz, E., Zappel, M., Lerch, S. P., Ehrenthal, J. C., Walter, S., & Ditzen, B. (2024). User preferences and trust in hypothetical analog, digitalized and AI-based medical consultation scenarios: An online discrete choice survey. *Computers in Human Behavior*, *161*, 108419. https://doi.org/10.1016/j.chb.2024.108419

Mengoni, A., Seghieri, C., & Nuti, S. (2013). Heterogeneity in preferences for primary care consultations: Results from a discrete choice experiment. *International Journal of Statistics in Medical Research*, *2*(1), 67–75. https://doi.org/10.6000/1929-6029.2013.02.01.08

Minn, S. W., Tariq, D., Ndubueze, C., Paul, P. M., & See, J. W. (2025). A Google Trends analysis exploring public interest in common psychiatric conditions and non-pharmacological interventions. *Asian Journal of Psychiatry*, *107*, 104482. https://doi.org/10.1016/j.ajp.2025.104482

Moser, R. P., Trivedi, N., Murray, A., Jensen, R. E., Willis, G., & Blake, K. D. (2022). Patient-centered communication (PCC) scale: Psychometric analysis and validation of a health survey measure. *PLOS ONE*, *17*(12), e0279725. https://doi.org/10.1371/journal.pone.0279725

Mühlbacher, A. C., Kaczynski, A., Zweifel, P., & Johnson, F. R. (2016). Experimental measurement of preferences in health and healthcare using best-worst scaling: An overview. *Health Economics Review*, *6*(1), 2. https://doi.org/10.1186/s13561-015-0079-x

Mühlbacher, A., & Johnson, F. R. (2016). Choice experiments to quantify preferences for health and healthcare: State of the practice. *Applied Health Economics and Health Policy*, *14*(3), 253–266. https://doi.org/10.1007/s40258-016-0232-7

Muir, H. J., Constantino, M. J., Coyne, A. E., Westra, H. A., & Antony, M. M. (2021). Integrating responsive motivational interviewing with cognitive-behavioral therapy for generalized anxiety disorder: Direct and indirect effects on interpersonal outcomes. *Journal of Psychotherapy Integration*, *31*(1), 54–69. https://doi.org/10.1037/int0000194

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *The Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Niu, Y., Sun, J., Zhu, K., Xu, B., Zhang, Y.-P., & Peng, M. (2025). The critical role and effects of patient-centered communication in psychotherapy: A narrative review. *Psychology Research and Behavior Management*, *18*, 1657–1671. https://doi.org/10.2147/PRBM.S528343

Olawade, D. B., Wada, O. Z., Odetayo, A., David-Olawade, A. C., Asaolu, F., & Eberhardt, J. (2024). Enhancing mental health with artificial intelligence: Current trends and future prospects. *Journal of Medicine, Surgery, and Public Health*, *3*, 100099. https://doi.org/10.1016/j.glmedi.2024.100099

OpenAI. (2025). *ChatGPT (GPT-4)* [Large language model].

Phillips, E. A., Himmler, S. F., & Schreyögg, J. (2021). Preferences for e-mental health interventions in Germany: A discrete choice experiment. *Value in Health*, *24*(3), 421–430. https://doi.org/10.1016/j.jval.2020.09.018

Poli, A., Kelfve, S., & Motel-Klingebiel, A. (2019). A research tool for measuring non-participation of older people in research on digital health. *BMC Public Health*, *19*(1), 1487. https://doi.org/10.1186/s12889-019-7830-x

Puchalski, C. M., Vitillo, R., Hull, S. K., & Reller, N. (2014). Improving the spiritual dimension of whole person care: Reaching national and international consensus. *Journal of Palliative Medicine*, *17*(6), 642–656. https://doi.org/10.1089/jpm.2014.9427

Qin, J., Nan, Y., Li, Z., & Meng, J. (2025). Effectiveness of communication competence in AI conversational agents for health: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *27*(1), e76296. https://doi.org/10.2196/76296

Quaife, M., Terris-Prestholt, F., DiTanna, G. L., & Vickerman, P. (2018). How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. *The European Journal of Health Economics*, *19*(8), 1053–1066. https://doi.org/10.1007/s10198-018-0954-6

Ramaul, L., Ritala, P., & Ruokonen, M. (2024). Creational and conversational AI affordances: How the new breed of chatbots is revolutionizing knowledge industries. *Business Horizons*, *67*(5), 615–627. https://doi.org/10.1016/j.bushor.2024.05.006

R Core Team. (2025). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Reinders, P., Augustin, M., Stephan, B., & Otten, M. (2025). Understanding patient priorities in teledermatology for psoriasis: A discrete choice experiment. *Journal of the European Academy of Dermatology and Venereology*, *39*(10), 1773–1784. https://doi.org/10.1111/jdv.20701

Saha, K., Jain, Y., & Choudhury, M. D. (2025). Linguistic comparison of AI- and human-written responses to online mental health queries (arXiv: 2504.09271) [preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2504.09271

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, *11*(6), 887. https://doi.org/10.3390/healthcare11060887

Scherr, S., Cao, B., Jiang, L. C., & Kobayashi, T. (2025). Explaining the use of AI chatbots as context alignment: Motivations behind the use of AI chatbots across contexts and culture. *Computers in Human Behavior*, *172*, 108738. https://doi.org/10.1016/j.chb.2025.108738

Scholich, T., Barr, M., Stirman, S. W., & Raj, S. (2025). A comparison of responses from human therapists and large language model-based chatbots to assess therapeutic communication: Mixed methods study. *JMIR Mental Health*, *12*(1), e69709. https://doi.org/10.2196/69709

Schuster, A. L. R., Crossnohere, N. L., Campoamor, N. B., Hollin, I. L., & Bridges, J. F. P. (2024). The rise of best-worst scaling for prioritization: A transdisciplinary literature review. *Journal of Choice Modelling*, *50*, 100466. https://doi.org/10.1016/j.jocm.2023.100466

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

Shen, H., & Wang, M. (2023). Effects of social skills on lexical alignment in human-human interaction and human-computer interaction. *Computers in Human Behavior*, *143*, 107718. https://doi.org/10.1016/j.chb.2023.107718

Stephan, D., Bertsch, A. S., Schumacher, S., Puladi, B., Burwinkel, M., Al-Nawas, B., Kämmerer, P. W., & Thiem, D. G. (2025). Improving patient communication by simplifying AI-generated dental radiology reports with ChatGPT: Comparative study. *Journal of Medical Internet Research*, *27*(1), e73337. https://doi.org/10.2196/73337

Sundar, S. S., Jia, H., Waddell, T. F., & Huang, Y. (2015). Toward a theory of interactive media effects (TIME). In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 47–86). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118426456.ch3

Swift, J. K., Callahan, J. L., Cooper, M., & Parkin, S. R. (2018). The impact of accommodating client preference in psychotherapy: A meta-analysis. *Journal of Clinical Psychology*, *74*(11), 1924–1937. https://doi.org/10.1002/jclp.22680

Szinay, D., Cameron, R., Naughton, F., Whitty, J. A., Brown, J., & Jones, A. (2021). Understanding uptake of digital health products: Methodology tutorial for a discrete choice experiment using the Bayesian efficient design. *Journal of Medical Internet Research*, *23*(10), e32365. https://doi.org/10.2196/32365

Therneau, T. M. (2024). *Survival* (Version 3.8-3) [Computer software]. https://doi.org/10.32614/CRAN.package.survival

Tinelli, M., Nikoloski, Z., Kumpunen, S., Knai, C., Pribakovic Brinovec, R., Warren, E., Wittgens, K., & Dickmann, P. (2015). Decision-making criteria among European patients: Exploring patient preferences for primary care services. *European Journal of Public Health*, *25*(1), 3–9. https://doi.org/10.1093/eurpub/cku082

Tinetti, M. E., Naik, A. D., Dindo, L., Costello, D. M., Esterson, J., Geda, M., Rosen, J., Hernandez-Bigos, K., Smith, C. D., Ouellet, G. M., Kang, G., Lee, Y., & Blaum, C. (2019). Association of patient priorities-aligned decision-making with patient outcomes and ambulatory health care burden among older adults with multiple chronic conditions: A nonrandomized clinical trial. *JAMA Internal Medicine*, *179*(12), 1688–1697. https://doi.org/10.1001/jamainternmed.2019.4235

Traets, F., Sanchez, D. G., & Vandebroek, M. (2020). Generating optimal designs for discrete choice experiments in r: The idefix package. *Journal of Statistical Software*, *96*(3), 1–41. https://doi.org/10.18637/jss.v096.i03

Valla, L., Slinning, K., Wentzel-Larsen, T., Røsand, G.-M., & Arnardóttir, S. B. (2025). The newborn behavioral observations (NBO) system embedded in routine postpartum care in at-risk families in Iceland: A randomized controlled trial. *BMC Pregnancy and Childbirth*, *25*(1), 13. https://doi.org/10.1186/s12884-024-07128-0

van der Eijk, M., Faber, M. J., Aarts, J. W. M., Kremer, J. A. M., Munneke, M., & Bloem, B. R. (2013). Using online health communities to deliver patient-centered care to people with chronic conditions. *Journal of Medical Internet Research*, *15*(6), e115. https://doi.org/10.2196/jmir.2476

Van Oerle, S., Lievens, A., & Mahr, D. (2018). Value co-creation in online healthcare communities: The impact of patients' reference frames on cure and care. *Psychology and Marketing*, *35*(9), 629–639. https://doi.org/10.1002/mar.21111

Vass, C., Boeri, M., Karim, S., Marshall, D., Craig, B., Ho, K.-A., Mott, D., Ngorsuraches, S., Badawy, S. M., Mühlbacher, A., Gonzalez, J. M., & Heidenreich, S. (2022). Accounting for preference heterogeneity in discrete-choice experiments: An ISPOR special interest group report. *Value in Health*, *25*(5), 685–694. https://doi.org/10.1016/j.jval.2022.01.012

von Lützow, U., Neuendorf, L. N., & Scherr, S. (2025). Effectiveness of just-in-time adaptive interventions for improving mental health and psychological well-being: A systematic review and meta-analysis. *BMJ Mental Health*, *28*(1), e301641. https://doi.org/10.1136/bmjment-2025-301641

von Weinrich, P., Kong, Q., & Liu, Y. (2024). Would you zoom with your doctor? A discrete choice experiment to identify patient preferences for video and in-clinic consultations in German primary care. *Journal of Telemedicine and Telecare*, *30*(6), 969–992. https://doi.org/10.1177/1357633X221111975

Walker, H. L., Ghani, S., Kuemmerli, C., Nebiker, C. A., Müller, B. P., Raptis, D. A., & Staubli, S. M. (2023). Reliability of medical information provided by ChatGPT: Assessment against clinical guidelines and patient information quality instrument. *Journal of Medical Internet Research*, *25*, e47479. https://doi.org/10.2196/47479

Wang, F., Wang, J., Hu, H., & Shi, W. (2025). When language heals: Evaluating patient-centered communication in Chinese telemedicine through communication accommodation theory. *Digital Health*, *11*, 20552076251411220. https://doi.org/10.1177/20552076251411220

Wang, J., Min, H., Li, T., Li, J., Jiang, Y., Zhang, J., Wu, Y., & Sun, X. (2025). Women's preferences and willingness to pay for AI chatbots in women's health: Discrete choice experiment study. *Journal of Medical Internet Research*, *27*(1), e67303. https://doi.org/10.2196/67303

Watson, B., & Gallois, C. (1998). Nurturing communication by health professionals toward patients: A communication accommodation theory approach. *Health Communication*, *10*(4), 343–355. https://doi.org/10.1207/s15327027hc1004_3

Whitaker, K. L., Ghanouni, A., Zhou, Y., Lyratzopoulos, G., & Morris, S. (2017). Patients' preferences for GP consultation for perceived cancer risk in primary care: A discrete choice experiment. *British Journal of General Practice*, *67*(659), e388–e395. https://doi.org/10.3399/bjgp17X690905

World Health Organization. (2022). *World mental health report: Transforming mental health for all*. https://www.who.int/publications/i/item/9789240049338

Xiao, Z., Zhou, M. X., Chen, W., Yang, H., & Chi, C. (2020). If I hear you correctly: Building and evaluating interview chatbots with active listening skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). Association for Computing Maschinery. https://doi.org/10.1145/3313831.3376131

Xue, J., Zhang, B., Zhao, Y., Zhang, Q., Zheng, C., Jiang, J., Li, H., Liu, N., Li, Z., Fu, W., Peng, Y., Logan, J., Zhang, J., & Xiang, X. (2023). Evaluation of the current state of chatbots for digital health: Scoping review. *Journal of Medical Internet Research*, *25*(1), e47217. https://doi.org/10.2196/47217

Zhang, J., Wang, J., Zhang, J., Xia, X., Zhou, Z., Zhou, X., & Wu, Y. (2025). Young adult perspectives on artificial intelligence-based medication counseling in China: A discrete choice experiment. *Journal of Medical Internet Research*, *27*, e67744. https://doi.org/10.2196/67744

Zheng, A., Long, L., Govathson, C., Chetty-Makkan, C., Morris, S., Rech, D., Fox, M. P., & Pascoe, S. (2025). Designing AI-powered healthcare assistants to effectively reach vulnerable populations with health care services: A discrete choice experiment among South African university students [preprint]. *medRxiv*. https://doi.org/10.1101/2025.01.30.25321409

Zhong, W., Luo, J., & Zhang, H. (2024). The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: A systematic review and meta-analysis. *Journal of Affective Disorders*, *356*, 459–469. https://doi.org/10.1016/j.jad.2024.04.057