"I Am the Gatekeeper": Clinician Visions of AI Collaboration in Exposure Therapy

Sam Moradzadeh
College of Information Sciences and Technology
The Pennsylvania State University
University Park, Pennsylvania, USA
szm6500@psu.edu

Andrew M Sherrill
Department of Psychiatry and Behavioral Sciences
Emory University School of Medicine
Atlanta, Georgia, USA
andrew.m.sherrill@emory.edu

Christopher W Wiese
School of Psychology
Georgia Instittue of Technology
Atlanta, Georgia, USA
chriswiese@gatech.edu

Rosa I. Arriaga
School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia, USA
arriaga@cc.gatech.edu

Saeed Abdullah
Information Sciences and Technology
Pennsylvania State University
University Park, Pennsylvania, USA
saeed@psu.edu

Abstract

Exposure therapy is one of the most effective treatments for anxiety and trauma-related disorders, yet many trained clinicians avoid it in practice. Its delivery is cognitively, relationally, and creatively demanding in ways that existing AI tools have not addressed. Research on AI in mental healthcare has focused on evaluative systems that score, flag, and recommend, leaving clinicians as consumers of AI outputs rather than partners in clinical reasoning. To understand how clinicians themselves envision AI collaboration, we conducted semi-structured interviews with twelve licensed exposure therapy specialists and analyzed the data through reflexive thematic analysis. We identified four attributes of the ideal AI teammate: a silent in-session observer, a creative partner in hierarchy design, a transparent reasoner, and a collaborator whose autonomy expands incrementally under clinician gatekeeping. We discuss how these attributes form a coherent relational model of human-AI collaboration, which we term relational AI architecture, and conclude with design and research implications for clinical AI.

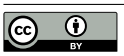
CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

Human-AI Collaboration, Mental Health, Clinical Decision Support, Exposure Therapy

ACM Reference Format:
Sam Moradzadeh, Andrew M Sherrill, Christopher W Wiese, Rosa I. Arriaga, and Saeed Abdullah. 2026. "I Am the Gatekeeper": Clinician Visions of AI Collaboration in Exposure Therapy. In *Companion of the Computer-Supported Cooperative Work and Social Computing (CSCW Companion '26), October 10–14, 2026, Salt Lake City, UT, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3785651.3831522



This work is licensed under a Creative Commons Attribution 4.0 International License.
CSCW Companion '26, Salt Lake City, UT, USA
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2378-0/26/10
https://doi.org/10.1145/3785651.3831522

1 Introduction

Anxiety and trauma-related disorders affect hundreds of millions of people worldwide [34], yet fewer than a third of those who need mental health treatment receive care that meets minimal adequacy standards [26, 32]. Exposure therapy stands among the most rigorously validated treatments for this population [13, 16, 23], and yet it remains one of the most underutilized in clinical practice [4, 19]. This underutilization is not primarily a matter of access or awareness. Clinicians who are trained in exposure therapy frequently avoid it, citing the sustained cognitive load of in-session monitoring, the creative demands of hierarchy design[1], and the relational weight of guiding patients through distress [1, 17]. The workflow is, in other words, the barrier.

In response to the broader mental health treatment gap, researchers and developers have introduced AI tools that have meaningfully advanced clinical practice, with systems capable of analyzing session transcripts, flagging risk indicators, and surfacing treatment recommendations with promising results [5, 15, 30]. Yet most remain oriented toward unidirectional evaluation, affording clinicians limited opportunity to contest, refine, or think alongside their outputs. Recent work has begun to challenge this orientation: Swinger et al. [29] found that clinicians hold conditional, context-dependent criteria for AI collaboration that existing tools are not designed to meet. Yet how clinicians conceptualize AI collaboration within the active delivery of treatment, rather than its supervision or training, remains unaddressed. Exposure therapy presents a particularly demanding test case: its in-session cognitive load, creative hierarchy design, and relational weight impose constraints on AI integration that no prior work has mapped.

To address this gap, we conducted semi-structured interviews with 12 licensed clinicians specializing in exposure therapy and analyzed the data using reflexive thematic analysis. We ask:

RQ: How do exposure therapy clinicians conceptualize their ideal AI teammate, and what attributes must it embody to earn a collaborative role in clinical practice?

[1] In exposure therapy, a hierarchy refers to a collaboratively constructed list of feared situations or stimuli, ranked from least to most distressing, that guides the order and intensity of exposure exercises across treatment.

Our findings reveal four defining attributes of the ideal clinical AI teammate, which together describe what we term *relational AI architecture*: a design orientation that reframes clinical AI from evaluator to earned collaborator. This work contributes to HCI and CSCW in three ways. First, it provides empirical insight into how licensed clinicians in a cognitively demanding psychotherapy modality conceptualize human-AI collaboration, expanding our understanding of AI teaming beyond general healthcare and data science contexts. Second, it introduces the concept of relational AI architecture as a design framework for clinical AI, grounded in clinician-identified attributes rather than system capabilities. Third, it offers a clinician-centered account of trust as developmental and conditional, complicating existing teaming models that treat trust as a configuration parameter rather than a relational achievement.

2 Related Work

We situate this work across three bodies of literature: AI applications in mental healthcare, the design of human-AI collaboration, and the clinical demands of exposure therapy delivery. Together, they converge on a shared gap: the absence of a clinician-centered account of what an AI teammate should look like within one of the most cognitively and relationally demanding forms of psychotherapy.

2.1 AI in Mental Healthcare

The past decade has produced a growing body of work on technology-mediated mental health support. Conversational agents and mobile applications have been explored as mechanisms for symptom tracking [24], psychoeducation [6], and between-session reinforcement [25]. More recently, clinical decision support systems have been developed to assist practitioners in treatment selection, risk stratification, and progress monitoring [12, 22]. However, they share a common design orientation: they position AI as an evaluator that outputs scores, flags, and recommendations to a clinician who receives but cannot meaningfully contest or refine them. This unidirectional framing leaves the clinician as a consumer of AI outputs rather than a partner in clinical reasoning. Our study addresses this gap by asking clinicians themselves what a bidirectional, collaborative AI role would require.

2.2 Human-AI Collaboration

HCI researchers have developed foundational frameworks for understanding how humans and AI systems work together [35]. Amershi et al. [2] established design guidelines for human-AI interaction that center intelligibility, calibrated trust, and appropriate reliance. Wang et al. [31] examined collaboration between data scientists and automated AI systems, finding that effective partnership depends on transparency about model limitations and opportunities for human contestation. Liao et al. [18] further showed that users' ability to interrogate AI reasoning is a prerequisite for trust, not a feature added after the fact. A growing body of work has extended these principles to Human-Autonomy Teams (HATs), where humans and autonomous AI agents work toward shared goals with interdependent roles [28, 36], though applications to high-stakes clinical contexts remain limited [27]. Swinger et al. [29] found that clinicians hold conditional, expertise-dependent criteria for AI collaboration that existing teaming models do not fully capture. Yet the collaborative demands of active treatment delivery remain unmapped; our study grounds human-AI collaboration in clinical practice rather than its training or supervision.

2.3 Exposure Therapy as a Clinical Workflow

Exposure therapy is among the most efficacious treatments for anxiety and trauma-related disorders [14], yet it remains one of the most underutilized in routine care [4, 19]. Abramowitz et al. [1] identified clinician-level barriers as a primary driver of this gap, including discomfort with inducing distress, uncertainty in hierarchy design, and the sustained cognitive load of in-session monitoring. The therapeutic relationship has also been consistently identified as a central mechanism of change [17], meaning that anything competing for the clinician's attention during a session carries real clinical cost. These features make exposure therapy an unusually demanding test case for AI integration. No prior work has examined how clinicians delivering exposure therapy envision AI collaboration within this specific context.

3 Methods

We recruited 12 licensed clinicians through purposive sampling from a registry of mental health professionals, targeting self-identified CBT [3] practitioners across professional directories, APA-accredited program listings, and conference networks. Participants ranged from master's-level student clinicians to licensed psychologists with over two decades of experience, spanning private practice, VA hospital, academic medical center, community mental health, and primary care settings. Nine identified as women, two as men, and one as non-binary; eight identified as White, three as Asian, and one as Hispanic or Latino. Prior to each interview, participants completed a short questionnaire covering clinical role, setting, experience, and technology use (see Appendix A). This study received IRB approval (STUDY00028090); participants were compensated $50.

3.1 Data Collection

We conducted 12 semi-structured interviews via Zoom between December 2025 and March 2026, each lasting approximately 45 minutes (M = 47 minutes). Interviews were audio- and video-recorded and transcribed using Zoom's transcription service. All transcripts were subsequently reviewed and corrected against the recordings for accuracy.

Our interview protocol began with questions about participants' current clinical workflows and the particular demands of delivering exposure therapy, before moving into how they envisioned AI fitting within those workflows. Consistent with our research questions, we did not frame AI as a replacement for clinical judgment. Instead, interviews focused on collaboration rather than automation, helping avoid priming participants toward either enthusiasm or resistance.

3.2 Data Analysis

We conducted reflexive thematic analysis following Braun and Clarke's six-phase framework [7, 8], approaching the data inductively. The first author served as primary analyst and the fifth author

as secondary coder, meeting regularly to review emerging codes and resolve disagreements through discussion. Code-to-theme development was iterative: for example, codes such as "evidence sourcing," "reasoning visibility," and "recommendation skepticism" converged into the theme *Transparent Reasoning as a Condition for Clinical Trust* once the research team identified that participants were articulating transparency as a precondition for action rather than a desirable feature. The analysis produced eight candidate themes; we selected the four addressing attributes of an ideal AI teammate and excluded four characterizing current workflows, documentation burden, and technology attitudes, which described present practice rather than envisioned collaboration. A condensed codebook is provided as supplementary material. All participants are referenced using assigned P-numbers.

4 Findings

Our findings reveal four attributes that clinicians across the sample consistently described as defining their ideal AI teammate: themes that characterize not what they want AI to do, but what they need AI to be. Taken together, they describe a coherent relational model of human-AI collaboration specific to the demands of exposure therapy practice.

4.1 AI as a Silent Second Observer

The most consistent image across all twelve interviews was not AI as advisor or evaluator, but as a presence-enabling companion: a background observer that absorbs the mechanical demands of a session so the clinician can attend fully to the patient. The clearest bottleneck was SUDS[2] monitoring. P3 described the bind directly: *"The most annoying part of exposures is SUDS monitoring. I don't ask during the exposure because I don't want to interrupt."* Collecting the data requires disrupting the conditions that make the data meaningful. The problem extended beyond physiological tracking. P1 noted: *"When a patient brings up something unexpected, I'm on my laptop mid-session looking for a resource. It breaks my presence."* Clinicians were not asking AI to contribute to the clinical conversation but to operate beneath it: passively monitoring vocal and behavioral distress markers to estimate SUDS, taking no action during the exposure itself, and producing a post-session distress trajectory for review. This is not a passive tool function but active collaboration, requiring continuous, context-sensitive judgment about when to act and when to recede.

4.2 Creative Partner in Exposure Design

When the creative process of building a hierarchy stalls, clinicians identified AI as a generative collaborator rather than a reference tool. P3 named the specific gap: *"I want AI to generate a larger menu of exposure options when both the patient and I feel stuck."* The request was not for AI to design the hierarchy, but to expand the option space when clinical creativity runs thin. P9 extended this to hierarchy navigation: *"AI could suggest intermediate exposure steps between two hierarchy options. What is something in between?"* This generative orientation is distinct from how current clinical AI tools are designed, which primarily evaluate or classify existing data rather than produce new material [9]. The clinician remains the architect. What clinicians wanted was something to broaden what there is to build with.

4.3 Transparency as the Price of Entry

Across the dataset, transparency was not described as a desirable feature. It was described as a prerequisite for action. P2 articulated the condition clearly: *"If the AI weighs pros and cons and links its reasoning to the PE manual*[3] *or a study, I would trust it."* The inversion was equally explicit for P8: *"I would be very hesitant to act on a recommendation without seeing its reasoning and data."* Several clinicians reframed what they wanted from AI altogether. P5 described it as: *"Less about getting an answer. More about having a thinking partner to test my own reasoning."* This aligns with Liao et al.'s [18] finding that the ability to interrogate AI reasoning is a precondition for trust, not an optional affordance. The ideal AI does not replace clinical judgment. It makes it more rigorous.

4.4 Graduated Trust and the Clinician as Gatekeeper

All twelve clinicians independently described a conditional, incremental model of AI trust, with the clinician serving as a non-negotiable gatekeeper between AI output and patient contact. The most common analogy was supervision. P3 articulated the trajectory: *"I would start in subordinate mode with aspiration to move toward collaboration as trust develops, like supervising a doctoral student."* The gatekeeper boundary was stated with particular clarity by P12: *"I am the gatekeeper. I would not use AI as a tool that communicates directly with the client."* P11 reinforced the same partition: *"I trust my own therapeutic and clinical skills to navigate as I see fit."* Documentation, data synthesis, and brainstorming were available for delegation. Clinical reasoning and direct patient contact were not. Trust, in this model, is not assigned at deployment. It is earned incrementally through demonstrated reliability.

5 Discussion

Our findings reveal that clinicians do not conceptualize AI as a smarter tool but as a particular kind of colleague: present but unobtrusive, generative but not authoritative, transparent about its reasoning, and willing to earn greater responsibility over time rather than assume it. We term this orientation *relational AI architecture*: a design framework that reframes the central question from what clinical AI should output to what kind of collaborative relationship it should sustain. Unlike interaction guidelines specifying momentary behavior [2] or teaming models specifying role allocation [28], it specifies the collaboration's trajectory: an AI that estimates distress silently, proposes evidence-linked exposures on request, and gains autonomy only through demonstrated reliability, never direct patient contact.

Each of our four themes speaks to a gap in an existing literature. The silent observer theme challenges the dominant design logic of AI in mental healthcare [12, 22], where tools are built to

[2]Subjective Units of Distress Scale (SUDS): a self-report measure where patients rate their distress level on a scale from 0 (no distress) to 100 (maximum distress), used during exposure therapy to track anxiety habituation in real time.

[3]The PE (Prolonged Exposure) manual refers to the standardized treatment protocol that guides clinicians in the structured delivery of exposure therapy for PTSD, including session-by-session procedures, psychoeducation scripts, and homework assignments.

produce outputs for clinicians to receive. Our participants did not want more output. They wanted reduced cognitive load during a session, a need that existing evaluative tools are not designed to address. The transparency theme extends Liao et al.'s [18] finding that interrogability is a precondition for trust by showing what interrogability must look like in a high-stakes clinical context: not an explainability dashboard, but a reasoning process that a clinician can test their own judgment against. The graduated trust theme complicates human-AI teaming frameworks [2, 31], which tend to treat trust as a configuration parameter. Our clinicians described trust as developmental and relational, closer to a supervisory relationship than a system setting. This suggests that teaming models built for data science or manufacturing contexts require significant translation before they can serve clinical psychotherapy.

The creative partner theme describes a need that existing clinical AI tools were never designed to meet. Current systems are evaluative by orientation: they classify, flag, and recommend based on existing data [12], whereas hierarchy construction requires producing novel combinations of stimuli tailored to a specific patient's fear structure. When P3 described wanting AI to generate *"a larger menu of exposure options when both the patient and I feel stuck,"* they were describing co-creative ideation [20] — a mode only beginning to emerge in clinical psychotherapy. Recent work shows LLMs can generate clinically appropriate exposure hierarchies [5] and personalized imaginal exposure scripts [33], yet what remains missing is a collaborative framework for integrating generative capability into practice, grounded in how clinicians themselves conceptualize that partnership.

Our findings also carry implications that extend beyond exposure therapy. The relational model clinicians described, where an AI earns trust incrementally, makes its reasoning visible, and accepts structural limits on its autonomy, maps onto a broader set of design challenges that CSCW researchers have identified across high-stakes collaborative work [9, 31]. In fields ranging from surgical assistance [21] to legal decision support [10, 11], the core tension is the same: how do human professionals incorporate AI into workflows defined by irreducible expertise, ethical accountability, and relational complexity [9]? Our participants' answer to this question was not to resist AI but to insist on the conditions under which it could be a legitimate participant. This distinction matters for how the field frames the problem. The question is not whether AI belongs in high-stakes clinical or professional work. It is what design commitments are required to make that belonging earned rather than assumed.

These findings carry concrete design implications. Clinical AI should reduce in-session cognitive demands without inserting itself into the relational space between clinician and patient, expose its reasoning rather than summarize it, and support graduated permission structures that allow trust to accumulate through demonstrated reliability. The gatekeeper boundary our participants drew offers designers a principled starting point for scoping what any clinical AI system should and should not be permitted to do.

6 Limitations

This study reflects a single modality, and findings may not generalize to other evidence-based psychotherapies. Our sample of 12 is modest, though licensed clinicians with direct exposure therapy experience remain a specialized, hard-to-access population; it also skewed toward clinicians identifying as women (9 of 12), which may bound the generalizability of our claims. Our data captures vision rather than behavior, and clinicians' stated ideals may not match their responses to an actual system under live clinical conditions. Finally, participants framed graduated trust as uniformly beneficial, but trust that grows with familiarity can drift toward overreliance and automation bias, a risk future work should examine.

Acknowledgments

This work was supported by the National Science Foundation under Award #2326145. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] Jonathan S Abramowitz, Brett J Deacon, and Stephen PH Whiteside. 2019. *Exposure therapy for anxiety: Principles and practice.* Guilford Publications.

[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems.* 1–13.

[3] APA Division 12 (Society of Clinical Psychology). 2017. *What is Cognitive Behavioral Therapy?* American Psychological Association. https://www.apa.org/ptsd-guideline/patients-and-families/cognitive-behavioral

[4] Emily M Becker-Haimes, Rebecca E Stewart, and Hannah E Frank. 2023. It's all in the name: why exposure therapy could benefit from a new one. *Current Psychology* 42, 25 (2023), 21641–21647.

[5] Emily E Bernstein, Adam C Jaroszewski, Ryan J Jacoby, Natasha H Bailen, Jennifer Ragan, Aisha Usmani, and Sabine Wilhelm. 2025. Feasibility of using ChatGPT to generate exposure hierarchies for treating obsessive-compulsive disorder. *Behavior Therapy* 56, 4 (2025), 680–688.

[6] Ananya Bhattacharjee, Joseph Jay Williams, Karrie Chou, Justice Tomlinson, Jonah Meyerhoff, Alex Mariakakis, and Rachel Kornfield. 2022. " I kind of bounce off it": translating mental health principles into real life through story-based text messages. *Proceedings of the ACM on Human-computer Interaction* 6, CSCW2 (2022), 1–31.

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[8] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.

[9] Carrie J Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. " Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. *Proceedings of the ACM on Human-computer Interaction* 3, CSCW (2019), 1–24.

[10] Quan Ze Chen and Amy Xian Zhang. 2025. Case law grounding: Using precedents to align decision-making for humans and AI. In *Proceedings of the ACM Collective Intelligence Conference.* 226–238.

[11] Fernando Delgado, Solon Barocas, and Karen Levy. 2022. An uncommon task: Participatory design in legal AI. *Proceedings of the ACM on human-computer interaction* 6, CSCW1 (2022), 1–23.

[12] Hayley I Evans, Myeonghan Ryu, Theresa Hsieh, Jiawei Zhou, Kefan Xu, Kenneth W Akers, Andrew M Sherrill, and Rosa I Arriaga. 2024. Using Sensor-Captured Patient-Generated Data to Support Clinical Decision-making in PTSD Therapy. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–28.

[13] Edna B Foa. 2011. Prolonged exposure therapy: past, present, and future. *Depression and anxiety* (2011).

[14] Edna B. Foa, Elizabeth Hembree, and Barbara Rothbaum. 2007. *Prolonged Exposure Therapy for PTSD: Emotional Processing of Traumatic Experiences, Therapist Guide.* Oxford University Press. doi:10.1093/med:psych/9780195308501.001.0001

[15] Arash Javanbakht, Liza Hinchey, Kathleen Gorski, Alex Ballard, Luke Ritchie, and Alireza Amirsadri. 2024. Unreal that feels real: artificial intelligence-enhanced augmented reality for treating social and occupational dysfunction in post-traumatic stress disorder and anxiety disorders. *European Journal of Psychotraumatology* 15, 1 (2024), 2418248.

[16] Johanna S Kaplan and David F Tolin. 2011. Exposure therapy for anxiety disorders: theoretical mechanisms of exposure and treatment strategies. *Psychiatric Times* 28, 9 (2011), 33–33.

[17] Nikolaos Kazantzis, Craig Whittington, and Frank Dattilio. 2010. Meta-analysis of homework effects in cognitive and behavioral therapy: A replication and extension. *Clinical Psychology: Science and Practice* 17, 2 (2010), 144.
[18] Vivian Lai, Yiming Zhang, Chacha Chen, Q Vera Liao, and Chenhao Tan. 2023. Selective explanations: Leveraging human input to align explainable ai. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–35.
[19] Diane Langthorne, Jessica Beard, and Glenn Waller. 2023. Therapist factors associated with intent to use exposure therapy: a systematic review and meta-analysis. *Cognitive Behaviour Therapy* 52, 4 (2023), 347–379.
[20] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J Cai. 2020. Novice-AI music co-creation via AI-steering tools for deep generative models. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.
[21] Angela Mastrianni, Mary Suhyun Kim, Travis M Sullivan, Genevieve Jayne Sippel, Randall S Burd, Krzysztof Z Gajos, and Aleksandra Sarcevic. 2025. To Recommend or Not to Recommend: Designing and Evaluating AI-Enabled Decision Support for Time-Critical Medical Events. *Proceedings of the ACM on human-computer interaction* 9, 7 (2025), 1–33.
[22] Angela Mastrianni, Paige Kmetz-Cutrone, Kathryn Chang, Jonathan Y Stein, and Aleksandra Sarcevic. 2025. Beyond Decision Making: Considering Collaboration and Agency in the Design of AI-Based Decision-Support Systems for Fast-Response Medical Teams. *Proceedings of the ACM on human-computer interaction* 9, 7 (2025), 1–28.
[23] Carmen P McLean, Hannah C Levy, Madeleine L Miller, and David F Tolin. 2022. Exposure therapy for PTSD: A meta-analysis. *Clinical psychology review* 91 (2022), 102115.
[24] Elizabeth L Murnane, Tara G Walker, Beck Tench, Stephen Voida, and Jaime Snyder. 2018. Personal informatics in interpersonal contexts: towards the design of technology that supports the social ecologies of long-term mental health management. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–27.
[25] Bruna Oewel, Patricia Anne Arean, and Elena Agapie. 2024. Approaches for tailoring between-session mental health therapy activities. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19.
[26] Julia R Pozuelo, Daniel V Vigo, Alan E Kazdin, Meredith G Harris, Dan J Stein, Maria Carmen Viana, Irving Hwang, Timothy L Kessler, Sophie M Manoukian, Nancy A Sampson, et al. 2025. Predictors and barriers to minimally adequate treatment among treated individuals with mental disorders: results from the World Mental Health Surveys. *International journal of mental health systems* 19, 1 (2025), 34.
[27] Beau Schelble, Amy Barron, Nathan McNeese, and Thomas O'Neill. 2020. Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature. *All Data Sets* (Jan. 2020). doi:10.25384/sage.c.5184895.v1
[28] Beau G Schelble, Christopher Flathmann, Nathan J McNeese, Guo Freeman, and Rohit Mallick. 2022. Let's think together! Assessing shared mental models, performance, and trust in human-agent teams. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–29.
[29] Nathaniel Swinger, Cynthia M Baseman, Myeonghan Ryu, Saeed Abdullah, Christopher W Wiese, Andrew M Sherrill, and Rosa I Arriaga. 2025. There's No" I" in TEAMMAIT: Impacts of Domain and Expertise on Trust in AI Teammates for Mental Health Work. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (2025), 1–36.
[30] Anja Thieme, Maryann Hanratty, Maria Lyons, Jorge Palacios, Rita Faia Marques, Cecily Morrison, and Gavin Doherty. 2023. Designing human-centered AI for mental health: Developing clinically relevant applications for online CBT treatment. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–50.
[31] Dakuo Wang, Justin D Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. 2019. Human-AI collaboration in data science: Exploring data scientists' perceptions of automated AI. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–24.
[32] Philip S Wang, Sergio Aguilar-Gaxiola, Jordi Alonso, Matthias C Angermeyer, Guilherme Borges, Evelyn J Bromet, Ronny Bruffaerts, Giovanni De Girolamo, Ron De Graaf, Oye Gureje, et al. 2007. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. *The Lancet* 370, 9590 (2007), 841–850.
[33] Yimeng Wang, Yinzhou Wang, Alicia Hong, and Yixuan Zhang. 2026. Explore LLM-enabled Tools to Facilitate Imaginal Exposure Exercises for Social Anxiety. In *Proceedings of the 2026 ACM Interactive Health Conference (IH '26)*. Association for Computing Machinery, New York, NY, USA, Article 26, 8 pages. doi:10.1145/3786579.3804922
[34] World Health Organization. 2025. Anxiety Disorders. Fact Sheet. Accessed: May 14, 2026. https://www.who.int/news-room/fact-sheets/detail/anxiety-disorders
[35] Wei Xu. 2019. Toward human-centered AI: a perspective from human-computer interaction. *interactions* 26, 4 (2019), 42–46.
[36] Rui Zhang, Nathan J McNeese, Guo Freeman, and Geoff Musick. 2021. " An ideal human" expectations of AI teammates in human-AI teaming. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–25.

A Participant Demographics

Participants represented a range of clinical backgrounds, training stages, and practice contexts. Seven held expert-level exposure therapy experience (20+ cases), three reported intermediate experience (5–20 cases), and two were novices with fewer than five cases. Clinical settings spanned private practice, VA hospital, academic medical center, community mental health, and primary care. Documentation burden ranged from 0–2 hours per week to 8 or more, reflecting meaningful variation in administrative load across participants.

Table 1: Participant Demographics (N=12)

ID	Age	Gender	Race/Ethnicity	Professional Role	Degree Year	Clinical Setting	ET Experience	Docs (hrs/wk)
P1	35–44	Woman	White	Licensed Psychologist, Academic Professor	2017	Private Practice	Expert (20+)	0–2
P2	35–44	Woman	White	Clinical Psychologist	2024	Private Practice	Expert (20+)	0–2
P3	35–44	Woman	Asian	Licensed Psychologist	2015	Telehealth / Contract	Intermediate (5–20)	2–4
P4	35–44	Woman	White	Licensed Psychologist	2021	Other (Athletics)	Intermediate (5–20)	2–4
P5	25–34	Woman	Asian	Assistant Professor / Licensed Psychologist	2021	Private Practice	Intermediate (5–20)	0–2
P6	25–34	Non-binary	White	Licensed Psychological Associate; Doctoral Student	2021	Community Mental Health	Intermediate (5–20)	2–4
P7	45–54	Woman	Asian	Licensed Psychologist	2014	Private Practice	Expert (20+)	2–4
P8	45–54	Man	White	Licensed Psychologist	2002	Other (Primary Care)	Expert (20+)	2–4
P9	25–34	Woman	White	Clinical Psychology PhD Candidate	2024	VA Hospital / Private Practice	Novice (<5)	2–4
P10	35–44	Man	White	Licensed Psychologist	2012	Academic Medical Center	Expert (20+)	4–6
P11	45–54	Woman	White	Licensed Psychologist	2006	Private Practice	Expert (20+)	8+
P12	25–34	Woman	Hispanic/Latino	Student Clinician (Master's Level)	2025	Private Practice	Novice (<5)	6–8