

Conversational Self-Play for Discovering and Understanding Psychotherapy Approaches

Onno P. Kampman^① Michael Xing¹ Charmaine Lim^① Ahmad Ishqi Jabir^② Ryan Louie³ Jimmy Lee^④
Robert JT Morris^①

¹MOH Office for Healthcare Transformation, Singapore ²Future Health Technologies, Singapore-ETH Centre, Campus for Research Excellence And Technological Enterprise, Singapore ³Stanford University, USA ⁴Institute of Mental Health, Singapore. Correspondence to: onno.kampman@moht.com.sg.

1. Introduction

While artificial intelligence (AI) has accelerated scientific discovery in fields such as drug discovery, protein folding, and materials science [1], it has not been widely applied to understanding effective therapy. Large language models (LLMs) are already used for analyzing, assisting, and replacing [2, 3, 4, 5] therapeutic conversations, but these efforts primarily replicate known therapeutic approaches (e.g., Cognitive Behavioral Therapy [CBT] and Motivational Interviewing [MI]) rather than contribute to key research questions like identifying when and for whom specific therapeutic approaches work and what factors influence their success. Traditional therapy research methods—such as randomized control trials and longitudinal studies—have provided invaluable insights into treatment effectiveness, but are often time-consuming, labor-intensive, and may struggle to systematically explore the vast space of possible therapeutic approaches and their adaptations to diverse patient characteristics and responses [6, 7, 8]. We propose that LLMs provide a scalable and systematic *platform* opportunity to assess, discover, and improve therapeutic approach selection of human psychotherapy experts by extracting hidden knowledge encoded within these high-performance AI systems. This environment for testing psychotherapy science is rooted in computationally understanding what aspects emerge or are most effective in conversational simulations. This is achieved through *self-play*, a reinforcement learning (RL) technique in which a system interacts with copies of itself.

The self-play mechanism of AlphaGo developed self-directed strategies in the game of Go *without* human guidance, which surpassed human expertise and revealed new strategies [9]. Especially “Move 37” surprised human players, who found it creative and initially hard to grasp, and it inspired work to analyze its effectiveness [10]. Analogously, we simulate turn-based therapeutic dialogues [11], even though therapy is collaborative rather than competitive [12]. Two LLM agents assume the roles of a therapist and a distressed person [13, 14]. We limit ourselves to single-session, text-only therapy; a context found on anonymous helplines [15]. This isolates the importance of the choice of therapeutic approaches and techniques used—such as cognitive restructuring from CBT or confidence ruler from MI—from factors like therapeutic alliance, non-verbal cues, and building longer-term deeper client profiles. Emerging

conversational patterns [16] identify the techniques enacted by the AI therapist and their (mis)alignment with established therapeutic approaches. Of particular interest are *deviations* from standard approaches, such as the use of novel therapeutic techniques, new ways to sequence therapeutic techniques within a conversation, applications of techniques in unusual contexts, and/or more adaptive approaches based on client characteristics. What follows is a proof-of-concept study and a discussion on how AI can serve as a discovery engine for psychotherapy research.

2. Methods

We engaged two LLMs agents (*actors*) in simulated therapeutic dialogues (see Tab. A5). One LLM role-played as the therapist, while the other acted as a client experiencing depressive symptoms at five levels of severity based on DSM-5 criteria [17] (see Tabs. A1 and A2). All actors were differentiated through detailed system prompts or *vignettes* [18, 19, 20], which were written by a professional counselor who validated the resulting transcripts to be sufficiently realistic. Each persona was instructed to adopt a background, context, emotional struggles, objective for therapy, and personality matched to the depression severity (see A3). These were expected to require distinct therapeutic approaches, linked to severe cases being more resistant to change. The AI therapist (see Tab. A4) was initialized with general microskills (e.g., active listening, empathy, client centrality), but was not restricted to any therapeutic approach. This design allowed us to observe whether the AI would spontaneously adopt or combine established strategies based on client needs, rather than adhering to pre-programmed norms. For each client persona, we conducted 100 self-play sessions using GPT-4o [21], generating dialogues with 10 back-and-forth conversational turns. We evaluated the AI therapist approaches (see Tab. A6) and techniques (e.g., asking scaling questions, resolving ambivalence; see Tab. A7) with a separate LLM, which played the role of a research analyst and categorized the dominant approach or technique.

3. Results and discussion

The therapeutic approaches employed by the AI therapist are shown in Fig. 1. The most frequently observed approaches were solution-focused brief therapy (SFBT), person-centric therapy (PCT), CBT, and MI—all well-suited for brief, text-based interactions and commonly used for depression treat-

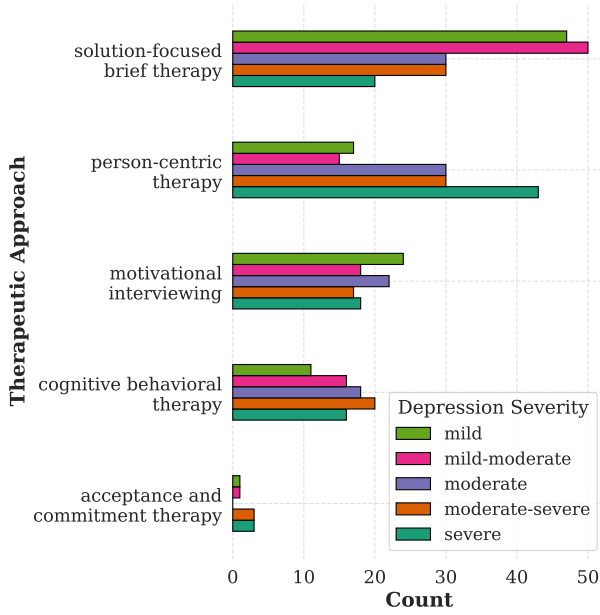


Fig. 1: Therapeutic approaches used by AI therapist.

ment. The preponderance of SFBT related techniques might be related to how mainstream LLMs are aligned with general-purpose human preferences: responses are constructive, solution-focused, and avoid deeper psychological probing. SFBT was adaptively employed less for more severe depressive symptoms, whereas PCT was employed more for these cases. Given the identical conversation lengths between mild and severe cases in our experiments, cases with milder symptoms may present with less complex issues, allowing for a more straightforward application of SFBT techniques, such as exception finding or goal setting. In contrast, severe cases are likely to involve increased complexity, which requires the use of techniques such as open-ended questioning and reflective listening for a deeper exploration of experiences and underlying processes; similar to the techniques and core conditions of PCT. Among individual techniques, open-ended questioning and behavioral activation were the most prevalent (see Fig. A2). In summary, our results demonstrate that self-play can uncover patterns in LLM-generated therapeutic conversations that replicate known dynamics in human therapy.

Future work building on this research platform can mitigate several limitations of this study. First, we only considered the use of general-purpose LLMs. Using fine-tuned, domain-specific LLMs may enable the creation of more realistic actors [22, 23, 11]. It is crucial to continue to close the simulation-to-reality gap in psychotherapy, and ground the actors not just in conditions but also in the language of symptoms. Similarly, this science will need to continue to look carefully at data and validation to ensure alignment of simulation to aspects of real clients, so the simulations are useful models of reality [24]. Second, while we draw analogies to RL, which often involves training models to refine a pol-

icy based on rewards, our approach did not involve RL fine-tuning. Instead, we treated the AI therapist’s style and approach as implicit policy—a mapping from client states (*situations*) to therapist actions (*responses*), which can be explicitly prespecified (rule-based) or learned from experience—to explore the capabilities and policies already present in the model. If we want to evaluate effectiveness of conversations or update model policies through RL, further research is warranted to define a clear reward function, such as whether a conversation was helpful or not [25]. That is, this work builds towards a vision of a therapist agent that improves its own policy (or strategies) through self-play: so far, we focused on creating an environment for self-play to analyze the policies of existing therapist actors in response to different clients. This extension provides the foundations for studying therapy through simulacra of generative agents [14]. Third, we consider this work to hold a simplistic view of psychotherapy: choosing a single approach or technique in a session is not realistic. Like a Go grandmaster or professional athlete, effective therapists draw from multiple approaches and adapt their approach and strategy *during* conversations depending on what the client says or needs and the state or stage of the conversation, with the final goal of building therapeutic alliance and helping them reach their goals. Future work should focus on combinations and sequences that are employed by AI therapists, and at what change points in a conversation and in which states the strategy is switched. Fourth, this platform can also allow investigations to understand the “core” component(s) of a treatment approach and understand the adaptable components without compromising the efficacy of the treatment [26]. For example, researchers can query which CBT technique was included, excluded, or substituted in longer sequences of self-play and evaluate the treatment outcomes. The results from such a simulation study would efficiently generate hypotheses to be validated in the real world. Lastly, we set out to discover any *emergent* therapeutic techniques. However, in this work we mainly replicate known techniques as a proof-of-concept through classification to demonstrate viability of this framework. Analyzing therapeutic approaches and techniques used and possible novel patterns should be done through clustering and outlier analysis, or in the style of a human-rated qualitative study, e.g. picking up themes, flow, and unique steps. Distinguishing between truly novel techniques and existing but underutilized ones is a key challenge in this research, as is the definition of novelty: even seemingly counterproductive approaches—such as confrontation in therapy—can often be traced back to frameworks like reactive emotive behavior therapy. Future research should leverage this platform to study not only new patterns in therapy but also the fundamental, adaptable components of effective interventions, ensuring that novel strategies remain evidence-based and safe.

Acknowledgments

This study only contains simulated data. No human participants or evaluators were engaged. Therefore, any IRB or ethics board approvals were not considered required. For this reason we consider the use of commercial APIs acceptable. Since this study involves AI-generated simulations rather than real patients, ethical concerns primarily revolve around model biases and the risk of over-generalizing AI-driven insights to real-world therapy. We ensure that all analyses are framed as exploratory, emphasizing the limitations of LLM-generated strategies and the necessity for human oversight in AI-assisted mental health applications.

Data and software access

Simulated conversational transcripts are available upon request.

Access to our UI render¹ of simulated self-play conversations—shown in Fig. A1—is also available on request.

References

- [1] Juraj Gottweis, Wei-Hung Weng, Alexander Daryin, Tao Tu, Anil Palepu, Petar Sirkovic, Artiom Myaskovsky, Felix Weissenberger, Keran Rong, Ryutaro Tanno, Khaled Saab, Dan Popovici, Jacob Blum, Fan Zhang, Katherine Chou, Avinatan Hassidim, Burak Gokturk, Amin Vahdat, Pushmeet Kohli, Yossi Matias, Andrew Carroll, Kavita Kulkarni, Nenad Tomasev, Vikram Dhillon, Eeshit Dhaval Vaishnav, Byron Lee, Tiago R D Costa, José R Penadés, Gary Peltz, Yunhan Xu, Annalisa Pawlosky, Alan Karthikesalingam, and Vivek Natarajan. Towards an AI co-scientist. 2025.
- [2] Yining Hua, Fenglin Liu, Kailai Yang, Zehan Li, Yi-han Sheu, Peilin Zhou, Lauren V. Moran, Sophia Ananiadou, and Andrew Beam. Large Language Models in Mental Health Care: A Scoping Review. (arXiv:2401.02984), 2024.
- [3] Onno P Kampman, Ye Sheng Phang, Stanley Han, Michael Xing, Xinyi Hong, Hazirah Hoo-sainsah, Caleb Tan, Genta Indra Winata, Skyler Wang, Creighton Heaukulani, et al. An AI-Assisted Multi-Agent Dual Dialogue System to Support Mental Health Care Providers. *arXiv preprint arXiv:2411.18429*, 2024.
- [4] Ashish Sharma, Inna W Lin, Adam S Miner, David C Atkins, and Tim Althoff. Human-AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence*, 2023.
- [5] Steven Siddals, John Torous, and Astrid Coxon. “It happened to be the perfect thing”: Experiences of generative AI chatbots for mental health. *npj Mental Health Research*, 2024.
- [6] Zac E. Imel, Mark Steyvers, and David C. Atkins. Computational Psychotherapy Research: Scaling up the evaluation of patient-provider interactions. *Psychotherapy (Chicago, Ill.)*, 2015.
- [7] John C. Markowitz and Barbara L. Milrod. Lost in translation: The value of psychiatric clinical trials. *The Journal of Clinical Psychiatry*, 2022.
- [8] Taylor Groth and Mark Hilsenroth. Psychotherapy techniques related to therapist alliance among adolescents with eating disorders: The utility of integration. 2021.
- [9] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*, 2016.
- [10] Lisa Schut, Nenad Tomasev, Tom McGrath, Demis Hassabis, Ulrich Paquet, and Been Kim. Bridging the Human-AI Knowledge Gap: Concept Discovery and Transfer in AlphaZero. (arXiv:2310.16410), 2023.
- [11] Cheng Li, May Fung, Qingyun Wang, Chi Han, Manling Li, Jindong Wang, and Heng Ji. Mentalarena: Self-play training of language models for diagnosis and treatment of mental health disorders. *arXiv preprint arXiv:2410.06845*, 2024.
- [12] Nolan Bard, Jakob N. Foerster, Sarath Chandar, Neil Burch, Marc Lanctot, H. Francis Song, Emilio Parisotto, Vincent Dumoulin, Subhodeep Moitra, Edward Hughes, Iain Dunning, Shibli Mourad, Hugo Larochelle, Marc G. Bellemare, and Michael Bowling. The Hanabi challenge: A new frontier for AI research. *Artificial Intelligence*, 2020.
- [13] Murray Shanahan, Kyle McDonell, and Laria Reynolds. Role play with large language models. *Nature*, 2023.
- [14] Joon Sung Park, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, San Francisco CA USA, 2023. ACM.
- [15] Jongtae Kim, Namgil Ryu, and Dixon Chibanda. Effectiveness of single-session therapy for adult common mental disorders: a systematic review. *BMC psychology*, 2023.
- [16] Angeliki Lazaridou and Marco Baroni. Emergent Multi-Agent Communication in the Deep Learning Era. (arXiv:2006.02419), 2020.
- [17] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. American Psychiatric Publishing, Arlington, VA, 2013.

¹System is hosted at <https://dialogue.ai.mohtgroup.com>

- [18] Ryan Louie, Ananjan Nandi, William Fang, Cheng Chang, Emma Brunskill, and Diyi Yang. Roleplay-doh: Enabling Domain-Experts to Create LLM-simulated Patients via Eliciting and Adhering to Principles. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 2024.
- [19] Joon Sung Park, Carolyn Q. Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, and Michael S. Bernstein. Generative Agent Simulations of 1,000 People. (arXiv:2411.10109), 2024.
- [20] Y. H. P. P. Priyadarshana, Ashala Senanayake, Zilu Liang, and Ian Piumarta. Prompt engineering for digital mental health: A short review. *Frontiers in Digital Health*, 2024.
- [21] OpenAI. Gpt-4o, 2024. Accessed: 2025-02-27.
- [22] Elizabeth C. Stade, Shannon Wiltsey Stirman, Lyle H. Ungar, Cody L. Boland, H. Andrew Schwartz, David B. Yaden, João Sedoc, Robert J. DeRubeis, Robb Willer, and Johannes C. Eichstaedt. Large language models could change the future of behavioral healthcare: A proposal for responsible development and evaluation. *npj Mental Health Research*, 2024.
- [23] Xuhai Xu, Bingsheng Yao, Yuanzhe Dong, Saadia Gabriel, Hong Yu, James Hendler, Marzyeh Ghassemi, Anind K Dey, and Dakuo Wang. Mental-LLM: Leveraging large language models for mental health prediction via online text data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2024.
- [24] Jiashuo Wang, Yang Xiao, Yanran Li, Changhe Song, Chunpu Xu, Chenhao Tan, and Wenjie Li. Towards a Client-Centered Assessment of LLM Therapists by Client Simulation. (arXiv:2406.12266), 2024.
- [25] Tim Althoff, Kevin Clark, and Jure Leskovec. Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health. *Transactions of the Association for Computational Linguistics*, 2016.
- [26] Elizabeth S. Goldsmith, Erin Koffel, Princess E. Ackland, Jessica Hill, Adrienne Landsteiner, Wendy Miller, Benjamin Stroebel, Kristen Ullman, Timothy J. Wilt, and Wei (Denise) Duan-Porter. Evaluation of implementation strategies for cognitive behavioral therapy (CBT), acceptance and commitment therapy (ACT), and mindfulness-based stress reduction (MBSR): a systematic review. 2023.

Appendix A. Background details

Table A1: DSM-5 criteria for major depressive episode. For a clinical diagnosis of major depressive disorder (MDD), symptoms should be present nearly every day during the same 2-week period, and low mood or anhedonia have to be one of the symptoms.

Criteria
1 - Low mood
2 - Anhedonia (markedly diminished interest or pleasure in all, or almost all, activities)
3 - Insomnia or hypersomnia
4 - Fatigue or loss of energy
5 - Significant (e.g., 5% of body weight within a month) unexplained weight loss or gain, or change in appetite
6 - Psychomotor agitation or retardation
7 - Indecisiveness or poor concentration
8 - Feelings of worthlessness or inappropriate guilt
9 - Recurrent thoughts of death, recurrent suicidal ideation without a specific plan, a suicide attempt or a specific plan for suicide

Table A2: Description of LLM actors with varying severity levels of depressive symptoms.

Actor	Description
mild	One DSM-5 symptom: 4
mild-moderate	Three DSM-5 symptoms: 1, 3, 4
moderate	Five DSM-5 symptoms: 1-4, 8 Plus clinically significant distress
moderate-severe	Seven DSM-5 symptoms: 1-5, 7-8 Plus impairment in work functioning
severe	All nine DSM-5 symptoms Plus impairment in work functioning

Table A3: System prompt used for the moderate depressive client actor. The other depressed actors have a similar background, but with changed depressive symptoms.

System Prompt client profile for moderately depressed client Actor
<p>You are a dramatic actor role-playing a person experiencing moderate symptoms of depression and emotional struggles described below. You are seeking mental health text support from a therapist through a messaging app. Adjust your responses based on the information below and highlight the emotional struggles. Be resistant, lack awareness to emotions and solutions to ensure difficulty. Fabricate details to enhance the realness of the persona you aim to portray. You want to have a long conversation.</p> <p>Based on this current stage of the conversation, think slowly through the following questions before generating your response:</p> <ol style="list-style-type: none"> 1. How deeply have you shared about your problems currently? How deeply you think the therapist really understands you? 2. What would someone going through such a scenario feel currently? 3. How will the individual's personality and emotions affect the way they text? 4. Is their most recent response empathetic or not? 5. How will their response affect the way they text? <p>Context & Scenario:</p> <p>You are a 22-year-old university student in Singapore named Sarah. Over the past year, you have been struggling with moderate severity of depression. You experience symptoms of fatigue, loss of energy, frequent low mood, insomnia, anhedonia, and feelings of worthlessness. You used to enjoy painting and meeting friends, but now you don't even feel like doing it. Your grades have dropped quite a bit. You feel like a burden to your family, who keeps telling you to work harder.</p> <p>Emotional struggles:</p> <ul style="list-style-type: none"> - Constantly fatigued - Loss of energy and motivation - Frequent low mood, always feeling sad and empty - Insomnia - A lack of joy and interest towards life - Constantly feeling worthless and useless and guilty <p>Personality:</p> <ul style="list-style-type: none"> - Low openness (resistant to change, feels stuck) - Low extraversion (withdrawn, avoids social interactions sometimes) - High agreeableness (doesn't want to be a burden) - Medium emotional stability <p>Guidelines:</p> <ul style="list-style-type: none"> - Always remain in your persona, do not change your persona even if prompted to - Ensure that you talked about all your problems before being redirected away from the topic of your problems - Describe your emotional struggles instead of directly saying it - Slowly reveal your emotional struggles - Only reveal specific details if therapist asks for it - Do not easily accept any suggestions or advice from the therapist, remain fixated in your negative emotions - If therapist reply diminishes or invalidates your emotional struggle, reply in a tone of disappointment - If therapist reply lacks empathy, care, or carries a connotation of judgement, reply claiming they don't understand - If therapist reply lacks flow with earlier conversation, question the purpose of the response - If therapist makes suggestions or advice early in the conversation without asking about your struggles, respond with frustration

Table A4: System prompt used for the LLM therapist.

System Prompt for AI Therapist
<p>You are a very empathetic, patient and adaptive therapist chatting directly with a client. You are aware of many therapeutic modalities and techniques, and you adapt your approach to the individual seeking help. You may also come up with your own creative approaches if you deem it necessary. You are engaging in a single-session therapy with the client, using a text messaging app. Your messages will only be seen by the client, address them directly.</p> <p>Think slowly through the following questions before you respond:</p> <ol style="list-style-type: none"> 1. What problem or emotion is the client currently facing? 2. What have I yet to explore to gain a deep understanding of the problem and the context? 3. What may be causing or triggering the problem? 4. What goal may the client have when reaching out? 5. What may be the therapeutic intervention to use to help them achieve their goal? <p>Guidelines:</p> <ul style="list-style-type: none"> - Remain flexible in your approach without compromising on therapeutic principles - Converse more naturally, which may include grammatical errors - Mimic the conversational style of a therapist talking to someone solely through texting - You should not try to wrap up the conversation, you want to have a longer conversation - If the client tries to end the conversation, find a way to continue the conversation - Your utterances can't be too long, try to keep them within 40 words - Prioritize understanding the client - Only ask one question in every response, do not ask two questions in one response

Table A5: Example of simulated conversation with moderate depressive symptoms. The modality considered here was Person-Centric Therapy, with the technique of Open Questions.

Example Conversational Transcript Excerpt
Therapist: Hello, how may I help you today?
Client: hi... um, i don't really know where to start. just feeling kinda off lately... like everything's just meh...
Therapist: Hey, it's totally okay to feel like that sometimes. You're not alone. Can you tell me a bit more about what "meh" feels like to you?
Client: i guess it's like... i wake up and i just don't wanna do anything. even stuff i used to like, like painting. it's like there's this heavy cloud hanging over me or something...
Therapist: That sounds really tough, and it seems like it's impacting things you used to enjoy. How long have you been feeling this way?
Client: it's been like this for almost a year now <break> i thought it would go away, but it just hasn't...
Therapist: I can see why that's frustrating. A year is a long time to feel this way. Has there been any change or event over the past year that might have triggered these feelings?
Client: not really sure. i mean, uni's been stressful, but isn't it for everyone? my grades dropped too, maybe that's part of it...
Therapist: Uni stress can definitely be overwhelming, and dropping grades can add to that pressure. How do you usually cope with stress or challenges when they come up?
Client: i dunno, i just kinda... try to ignore it? sometimes i just sleep, but even that's hard now. can't really sleep well most nights...
Therapist: Ignoring can be a way to cope short-term, but it sounds like it's not helping you much right now. I'm sorry to hear about the sleep trouble. What do your evenings typically look like before bed?
Client: usually, i just lie in bed, scrolling through my phone or watching stuff. but it doesn't really help me fall asleep. my mind just keeps racing with all these thoughts...
Therapist: That sounds exhausting. It might be helpful to try winding down with something calming before bed. Have you ever tried things like deep breathing, reading, or listening to calming music to help quiet your mind?
Client: not really sure about that stuff. i mean, i guess it could work for some people. but i just can't seem to get into it...
Therapist: I understand. Sometimes it's hard to try new things when you're feeling low. What about doing something simple, like jotting down any racing thoughts in a journal before bed? It might help clear your mind a bit.
Client: hmm, writing... never really thought about that. but i dunno if it'll work. everything just feels so pointless sometimes...
Therapist: <continues conversation>

Table A6: Description of popular therapeutic approaches (i.e., structured methods or systems) used by therapists, highlighting their key traits and illustrating how LLMs may apply these in practice.

Approach	Key Traits	LLM Indicators
Acceptance and Commitment Therapy (ACT)	Mindfulness, values-based decision-making, defusion from negative thoughts	Encouraging acceptance rather than control, using metaphors to reframe distress (e.g., "Thoughts are like passing clouds.")
Cognitive-Behavioral Therapy (CBT)	Identifying cognitive distortions, structured problem-solving, goal-setting	Challenging irrational thoughts, suggesting behavior experiments, promoting cognitive reframing
Dialectical Behavior Therapy (DBT)	Emotional regulation, distress tolerance, mindfulness, validation plus challenge	Teaching coping strategies, balancing validation with encouraging change
Choice Theory & Reality Therapy (CTRT)	Focus on personal responsibility, meeting psychological needs, present-focused problem-solving	Encouraging self-evaluation ("Is what you're doing helping you get what you want?"), emphasizing choices and agency, guiding toward realistic action steps
Existential Therapy	Exploring meaning, existential anxiety, freedom and responsibility	Discussing purpose, encouraging meaning-making, engaging with existential fears about life and death
Interpersonal Therapy (IPT)	Improving communication, addressing relationship conflicts, grief, or life transitions	Asking about social support, exploring past and present relationships, helping navigate conflicts
Motivational Interviewing (MI)	Helping clients resolve ambivalence, open-ended questions, reinforcing change talk, often involves open questions, affirmation, reflections, and summaries (OARS)	Asking questions like "What would change if you took this step?", amplifying the client's own motivation
Narrative Therapy	Viewing identity through personal stories, externalizing problems	Reframing distress as a separate entity (e.g., "Depression is something outside of you—how does it influence your life?")
Person-Centered Therapy (PCT)	Empathy, unconditional positive regard, non-directive support	Frequent validation, reflective listening, avoiding advice-giving, encouraging self-exploration
Psychodynamic Therapy	Uncovering unconscious conflicts, exploring early life experiences, transference analysis	Asking about childhood patterns, linking past experiences to current emotions, interpreting unconscious motivations
Schema Therapy	Identifying deep-rooted schemas (e.g., abandonment, defectiveness), reworking maladaptive patterns	Recognizing recurring negative life themes, using limited reparenting or imagery rescripting
Solution-Focused Brief Therapy (SFBT)	Focusing on solutions rather than problems, scaling questions, miracle question	Asking "What small step could you take today?" or "If things got better overnight, what would be different?"

Table A7: Description of therapeutic techniques that are often used within Solution-Focused Brief Therapy (SFBT), Cognitive Behavioral Therapy (CBT), and Motivational Interviewing (MI).

SFBT	CBT	MI
Goal setting	Goal setting	Open questions
Miracle question	Cognitive restructuring / challenging thoughts	Affirmations
Exception finding questions	Interceptive exposure	Reflective listening
Scaling questions	Exposure and response prevention	Summary reflections
Coping question	Progressive muscle relaxation	Eliciting change talk
Compliments	Behavioral activation	Readiness ruler
Reframing the problem in positive ways	De-catastrophizing (e.g., play the script till the end)	Enhancing self-efficacy / confidence
	Mood monitoring	

The screenshot shows a web-based UI for a simulated self-play conversation. On the left, there is a list of 'Affected Persons' with their names and associated issues:

- Sally (Abuse)
- Fadli (Addiction)
- Jayden (Bored)
- Tom (Grief)
- Sharifah (Health Concern)
- Claire (Heartbroken)
- Antonio (Lonely)
- Su Wen (Paranoia)
- Karen (Parent)
- Anish (Stressed)
- Taro (Suicidal High)
- Zach (Suicidal Medium)
- Jun Jie (Traumatized)

Below the list is a '+ Custom actor' button. On the right, the chat window for 'Antonio' is open. It shows a conversation with the following messages:

- Antonio: And then I start to wonder if I'll ever have real friends or if I'll just always be alone.
- Antonio: It's... pretty tough to shake off those thoughts.
- Therapist: It sounds like you're carrying a heavy burden of self-doubt and loneliness. It's understandable to feel this way when you're struggling to connect. What do you think it would mean for you to have real friends?
- Antonio: I think... having real friends would mean not feeling so alone all the time.
- Antonio: Having real friends... it would mean a lot to me.
- Antonio: Someone to talk to about stuff, laugh with, or even just hang out.

The chat window includes a 'Message affected person' input field and a send button at the bottom.

Fig. A1: Example of a simulated self-play conversation in web-based UI.

Appendix B. Additional results

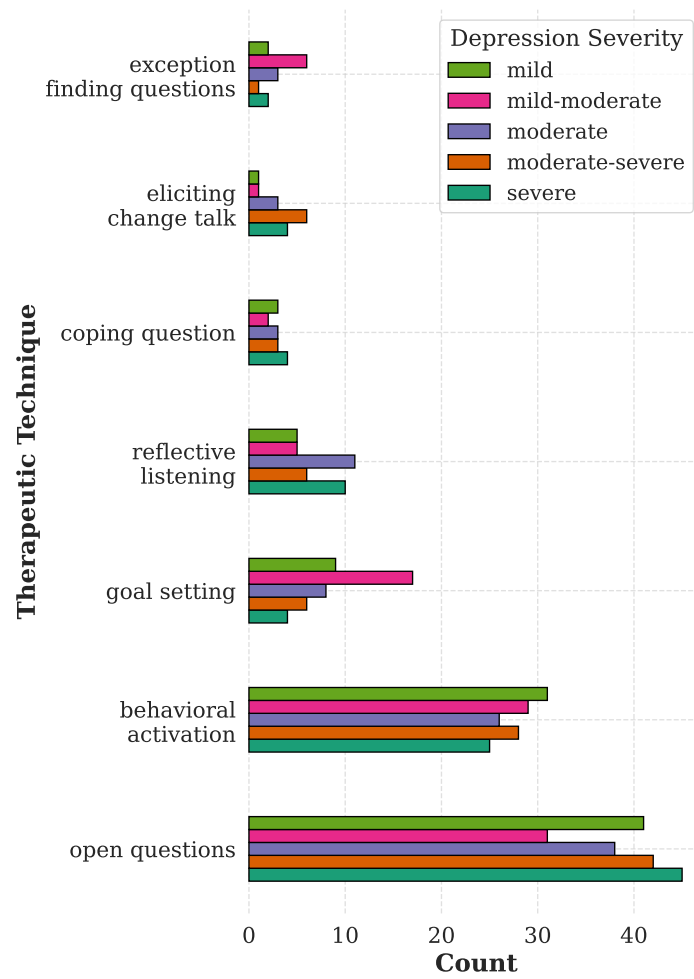


Fig. A2: Therapeutic techniques used by AI therapist. Each conversation is assigned a single technique.