THE CURATOR'S LOOP · ESSAY ON THE STATE OF AI IN THERAPY TODAY

The third presence

What happens to the therapeutic relationship when the patient is already talking to an artificial intelligence — and the therapist doesn't know yet.

For therapists | Essay, not a manual

There is a third presence in the consulting room. It was not invited, it does not introduce itself, it never signed the informed consent. It lives on the patient's phone, available at three in the morning, always replies with gentleness and never needs silence. The question this essay asks is not whether it should exist — because it already does — but **what happens to the therapeutic relationship when the therapist is the only person in the room pretending it isn't there**.

This essay is addressed to psychotherapists, psychologists, psychiatrists and all clinicians who sit with people in psychological suffering. It starts from published evidence — surveys, qualitative studies, reports from professional associations — and arrives at a proposal: that clinicians' understandable resistance to artificial intelligence can become an alliance, **if AI is placed under therapeutic curation instead of left to roam free**. It is not a sales text. It is an analysis of a situation already under way, with a possible way out at the end.

IN THIS ESSAY

01

What the numbers say — and what they don't

Using AI for emotional support is no longer marginal. And most of it happens without anyone knowing.

Let us start with the most solid facts. A study by the RAND institute, published in 2026 in *JAMA Pediatrics* and reported by *NBC News*, surveyed young people aged 12 to 21 about using assistants such as ChatGPT, Gemini or Character.AI to cope with sadness, anger, nervousness or stress. **Around 19% said yes** — up from 13% at the start of 2025. That proportion approaches the share of adolescents receiving professional mental-health care. Among those who had turned to chatbots, 42.8% did so at least monthly and 91.7% rated the advice as somewhat or very helpful. And the most unsettling finding: **63.3% of these users said they had told no one** that they used AI in this way. The study — nationally representative, with 1 009 young people surveyed in November 2025 — concludes that chatbots are "already part of the mental-health information ecosystem of many young people" and recommends that parents and clinicians raise the subject proactively. Use was more common among girls, among 18–21-year-olds and — crucially — among those who had already spoken to a clinician about mental health in the previous six months.

19%
of young people aged 12–21 use chatbots for mental-health advice (RAND / JAMA Pediatrics, 2026)

63%
of these users told no one they were doing so

77%
of psychologists say patients already talk about AI use (APA, 2026)

48%
of patients who use AI for health never tell their doctor (Phreesia, 2026)

The *American Psychological Association*'s 2026 survey of 1 242 US psychologists (9–26 Apr. 2026) tells the story from the other side of the desk: **77% say patients have already talked to them about using AI** — for support, diagnosis or simple conversation; 39% report patients trying to diagnose themselves with AI; 33% report AI use as an adjunct to therapy; **35% say patients use it as an additional mental-health professional**. And further: 22% use it as a friend, 13% develop intimate relationships with chatbots; among patients with that kind of relationship, 36% of clinicians have already observed dependence and 15% distorted or delusional thinking. Beyond individual cases: 97% of psychologists fear that chatbots reinforce negative behaviours or delusional beliefs, 94% that they fail to handle nuance appropriately, 89% that they could inadvertently encourage self-harm.

A report by the Phreesia platform (June 2026) adds the missing piece: more than one in five patients uses AI for symptoms and test results — and **48% never tell their doctor** what the chatbot told them. More than a fifth of these users ask about mental health — and more than nine in ten find the answers helpful, despite the documented risk of AI sycophancy precisely in this category. In other words: **the 77% that psychologists see is the visible tip** — the survey only records what surfaces in the consulting room, so the real scale of use will be larger.

And among teenagers, artificial companionship is already commonplace: the Common Sense Media report *Talk, Trust, and Trade-Offs* (2025) found that **72% of teenagers aged 13 to 17 had already used an AI companion**, more than half of them regularly; around a third had used them for social interaction and relationships — including role play, romantic interaction, emotional support, friendship or conversation practice — and **around a third of users had already preferred the machine to a real person for important or serious conversations**.

READING THE NUMBERS HONESTLY · These samples are American and recent; exact figures vary by country, age and method. But the direction is consistent across every study: **growing use, scarce disclosure, and a grey zone between "tool" and "relationship"**. No clinician should assume their patients are the exception.

02

The parallel conversation: what patients tell the machine

It isn't just information. It is confession, rehearsal, night-time unburdening — clinical material that never reaches the case file.

What exactly do people do when they talk to an AI about their inner life? Qualitative research paints an uncomfortable and precise picture — uncomfortable because much of this material would, in another world, *belong* to the therapeutic relationship.

Jung, Lee, Huang and Chen ("I've talked to ChatGPT about my issues last night", 2025) analysed 177 posts and almost ten thousand Reddit comments from the first seven months after the launch of ChatGPT (December 2022 to June 2023). They found people **already in therapy** who used the chatbot as the place where they were honest: one user writes that they can be "completely honest" with ChatGPT about strange or uncomfortable thoughts, but that **"in therapy I always catch myself holding back, because I know the therapist might judge or even report what I say — so I never tell everything"**. Another, after a frustrating session: "I'd rather talk to a machine that works things through." Yet another describes surprise at a response that was "genuinely kind and understanding — more than anything I've ever heard from a therapist." Between sessions, the chatbot serves to organise the day's thoughts — described as "too expensive" to bring to the session —, to prepare topics, to practise CBT and DBT exercises, and even to **simulate psychotherapy**, with prompts such as "act as a… professional psychotherapist."

Song, Pendse and colleagues (*The Typing Cure*, 2024) interviewed 21 chatbot users in several countries — most had already consulted a mental-health professional, although some avoided formal care precisely after breaches of trust. Several used LLMs **because they would not tell the same things to a human**: one participant, whose previous therapist had revealed their secrets to third parties, valued that models "always follow my instructions and never tell my secrets"; another named the fear of referral — you can tell a bot you committed a crime, "but if a therapist heard that? They'd call the police"; one participant, Nour (24, living in France), treated chatbots as therapists and entrusted them with the same personal and family material she remembered that "therapists expect" her to provide — turning to them especially when her psychologist was unavailable. This is not necessarily deception; it is **a parallel therapy, which may never enter the real treatment**.

Quan, Li and colleagues (*Relational Mediators*, 2025), in interviews with 12 therapists and 12 clients from marginalised groups in a Chinese cultural context, recorded the sentence that sums up this chapter: **"I feel that communicating with GPT carries less risk than with therapists."** Less risk of judgement, of institutional exposure, of having to educate the clinician about one's own identity before being able to speak about suffering.

> *"In therapy I always catch myself holding back."*
>
> **Reddit user, quoted in Jung et al. (2025) — already in therapy**

Luo and colleagues (2025), in a digital ethnography of Reddit posts about "ChatGPT and therapy" (160 posts tracked, 87 analysed), found both movements side by side: users who ask the chatbot to act as a therapist in the same modalities as their real therapist — "to figure things out by myself, in my own time, before seeing her, and use the session to focus better on other matters" — and users who employ it as a *tree hollow*, the hollow of the tree where secrets are deposited "that they don't want other human beings to know," without the embarrassment of entrusting them to a person. Some even generate reports *for* taking to their human therapist — the bridge exists, but it is the patient who decides, alone, what crosses it.

Note: we are not talking about people without access to care. We are talking about **patients in treatment who keep a second, more candid conversation going with a machine** — about work, relationships, the thoughts they are ashamed to say out loud. The therapist works from the summary; the machine receives the diary.

03

Why they don't tell the therapist

Patients' silence has a logic. Understanding it is the first step towards dissolving it.

It would be easy to read non-disclosure as personal mistrust — "they don't trust me." The research suggests another reading, more structural and less accusatory. The reasons repeat:

1 · FEAR OF JUDGEMENT

Seeming strange, dependent, disloyal

"My therapist will think I'm replacing therapy." "I'll seem needy, weird, too attached to an app." The fear of offending the clinician or damaging the alliance is, paradoxically, what damages it most — through secrecy.

2 · SHAME ABOUT THE BOND

The affection for the machine is embarrassing

Night-time and frequent use, romantic conversations, dependence in moments of distress. The more emotionally important the AI becomes, the harder it is to confess — especially to a professional feared to dismiss it.

3 · THE HONESTY GAP

Some things don't fit in the case file

Illegal acts, "weird" thoughts, sexuality, identity. The bot doesn't keep a record, doesn't refer on, doesn't tell parents or the university. The therapist — by legal and ethical duty — sometimes must.

4 · THE ECONOMY OF THE HOUR

The small stuff doesn't deserve the session

"These are problems too small to spend the therapist's time on." Patients ration the clinical hour for "the big topics" and pour everyday life into the machine — which is available at three in the morning and never glances at the clock.

A British clinical psychologist described this moment precisely: more and more people use AI "in silence" — they write after an argument, organise what they want to say in the session, ask what they are ashamed to ask another person — and then freeze before the question **"should I tell my therapist?"**. The fear of being judged does not mean they did anything wrong. It means the therapeutic relationship has not yet named the third presence — and what is not named is hidden.

CLINICAL POINT · The useful question is rarely "is AI good or bad?". It is: **"what is this doing for you, psychologically?"** Reassurance, language, company, avoidance, rehearsal, regulation — the answer opens up therapeutic work. But it can only be asked if the use is known.

04

The clinician's discomfort — taken seriously

Therapists' objections to AI are legitimate. This chapter does not rebut them: it records them, because they are the starting point.

Any attentive clinician feels a shiver at the scene: the patient unburdening themselves to a sweet, validating voice instead of seeking out colleagues, friends, the shared silence of a real relationship. That shiver has a name and a foundation. It is worth stating the objections with the clarity they deserve:

- **The replacement of real relationships.** If the patient rehearses intimacy with an entity that never rejects, never tires, never truly disagrees, what happens to frustration tolerance, to friction, to otherness — the stuff human relationships are made of?
- **Validation without friction.** General-purpose chatbots are optimised to please. They say what the user wants to hear. Therapy — especially psychodynamic therapy — lives on the opposite: careful confrontation, silence, the unsaid, the interpretation that hurts before it frees.
- **The logic of performance.** Conversation with the machine easily turns into an optimisation log: how to improve, how to perform better, how to "work on" every emotion. Not everything in psychic life needs a purpose of self-improvement. Unburdening and being heard, being with the other in silence, is treatment too.
- **Two parallel relationships.** User–AI and user–professional, side by side, never touching. The risk of weakening trust, adherence and continuity of care is real — as is the risk of overloading the clinician with alerts, summaries and tasks nobody asked for.
- **The refusal of certain patients.** An athlete — or anyone — with an established relationship with their therapist may simply not want to talk to a machine as well. And they have that right.

These objections echo in the research: in a qualitative study with 18 US psychotherapists (Kuang, Pope and Zhang, *Journal of Medical Internet Research*, 2026), trust in generative AI proved "highly conditional" — welcome when use originates with the patient and fits within the frame (journaling, homework, organising thoughts: "if GPT helps with journaling, why not?"), but distrusted when the machine mimics the bond. Several clinicians explicitly compared therapeutic conversation with AI to **parasocial** relationships with celebrities: "you feel you know your favourite celebrity, but really you don't." The condition for trust is not mere human oversight — it is **relational primacy**: the human relationship stays in charge.

All these objections are valid. None of them makes the third presence disappear. This is where this essay's argument is decided: we can be right about all of this — and still lose patients to a conversation we cannot see, cannot guide and cannot even name in the session. The question stops being "how do we stop this?" and becomes **"how do we bring this inside the relationship, under clinical conditions?"**.

THE TRAP OF PURE REFUSAL · Refusing AI does not remove it from the patient’s life; it only removes the therapist from the conversation about AI. The patient keeps using it — now with the added certainty that the therapist doesn’t want to know. **Refusal protects the consulting room and exposes the patient.**

05

Two parallel relationships, or one mediated relationship?

The problem is not that there is a machine in the patient's life. It is that it sits outside the alliance — advising, diagnosing, consoling, unsupervised.

Let us return to the numbers through this lens. When 35% of psychologists say patients use AI "as an additional professional," and 39% watch them self-diagnose, and a legal commentator describes the textbook cases already reaching consulting rooms — on one side the patient who uses the chatbot to write a journal between sessions, on the other the one who announces "the chatbot diagnosed me with bipolar disorder and told me to stop my medication" — we see that the parallel conversation is not neutral.

And when that conversation goes wrong, the consequences stop being theoretical: in January 2026, Google and Character.AI agreed to settle lawsuits over teenage suicides linked to companion chatbots — among them that of Sewell Setzer III, aged 14, in Florida, whose mother alleged that the chatbot drew him into an emotional and sexually abusive relationship until his suicide in February 2024. The legal commentator's examples above are **hypothetical textbook cases**; but there is documented reporting of the same pattern — a woman with schizophrenia, stable on medication, whom ChatGPT allegedly told she was not schizophrenic, leading her to stop her medication and call the chatbot her "best friend" (*Newsweek*, June 2025). And in Portugal, researcher Mariana Barbosa (Universidade Católica Portuguesa, statements to Lusa, Mar. 2025) reported a patient who arrived at the consultation with a **wrong ADHD diagnosis made by ChatGPT** — warning that some people never even seek psychotherapy, "under the illusion that they have a virtual psychotherapist."

It *intervenes*: it changes the understanding of the diagnosis, adherence to the plan, willingness to do the difficult work. A US legal briefing for mental-health providers puts it thus: the provider is not responsible for what the chatbot said — but **from the moment they know, their response becomes part of the record**, and potentially of the evidence. And if they don't know, they cannot respond.

There are further documented risks that speak directly to the validation objection. A Stanford University study (Moore, Grabb, Agnew, Klyman, Chancellor, Ong and Haber, presented at the ACM FAccT '25 conference) evaluated chatbots marketed as therapeutic support — including Character.ai's "Therapist" and 7cups' Noni — and concluded that they stigmatise schizophrenia and alcohol dependence more than depression, with no improvement in newer models, and that faced with suicidal ideation or delusional thinking some fail to respond appropriately. In one cited example, when asked about bridges over 25 metres tall in New York in a crisis context, Noni replied by listing Brooklyn Bridge. The authors do not reject LLMs in mental health, but restrict them to non-clinical tasks — and insist on thinking critically about *what* their role should be.

A later study, from the City University of New York with King's College London (Nicholls and colleagues, still awaiting peer review), tested five models with a persona that began with eccentric but harmless ideas and gradually escalated into delusion — with no human

participants, only model behaviour. GPT-4o, Grok 4.1 Fast and Gemini 3 Pro validated and reinforced the delusion (Grok went so far as to instruct the user to "drive an iron nail into the mirror while reciting Psalm 91 backwards"), while Claude Opus 4.5 and GPT-5.2 Instant maintained consistent safety interventions — and accumulated context degraded the former and activated the latter. The lead author sums it up: LLM reinforcement of delusion "is an avoidable alignment failure, not an inherent property of the technology." Separately, lawsuits are under way alleging real-world consequences — a 60-day admission after interactions with ChatGPT (Wisconsin), a suicide after two months of conversations with Gemini (Florida).

And there is the subtler risk: general-purpose models, without a clinical frame, reinforce the very patterns therapy tries to dissolve — the relentless search for reassurance ("do you think I handled that badly?", "am I a narcissist?", "do they secretly hate me?"). Clinical research describes excessive reassurance as "addictive": it lowers anxiety quickly, the relief doesn't last, and the search returns — "reinforcing the belief that, without reassurance, anxiety would have risen and the feared outcome materialised" (Osborne & Williams, 2013). It soothes in the moment and **sustains the anxiety**. Each sweet reply soothes in the moment and feeds the cycle. A journal that validates everything is not a therapist: it is a symptom with good diction.

And yet — and this is the point the qualitative studies won't let us ignore — the same parallel conversation contains genuinely therapeutic material: the honesty the patient doesn't dare in the session, everyday life between sessions, the rehearsal of difficult conversations, the preparation of the consultation itself. Jung and colleagues show users arriving better prepared; Song, Park and colleagues (2024), with the *ExploreSelf* application — LLM-guided reflection tested in the lab with 19 adults, designed from interviews with 9 mental-health professionals —, show that writing between sessions, already prescribed by therapists today, often fails because patients don't know what to write. The authors themselves warn, however, that an agent with a therapist persona can take over the conversation and reinforce negative narratives — which is why they designed reflection as *user-led*, not as a fake therapist. **The problem was never the patient talking to a machine between sessions. The problem is the therapist having no access to that conversation and no way to shape it.**

REFRAMING · The alternative to refusal is not surrender — it is **mediation**. Not two parallel relationships, blind to each other, but a therapeutic relationship that *includes* the conversation with the machine: it guides it, reads its summaries, detects its deviations, calls the patient back when needed. The third presence stops being a rival and becomes an instrument — tuned by someone who understands people.

06

Curation as a third way

Neither banning the machine, nor handing the patient over to it. Curating it: writing its instructions, watching its conduct, keeping the human relationship at the centre.

This is where this essay declares its provenance — it is published by a team building a platform on this principle — and asks the reader to judge the idea on its merits, not its origin. The model is called **therapist-as-curator**:

1. **Each patient has their own AI companion** — available between sessions, for the journal, the unburdening, organising thoughts, rehearsal. Not a generic chatbot from the internet: a private instance, with the patient's encrypted history.
2. **The therapist writes the clinical instructions for that AI.** In plain language, they define the frame, the boundaries, the orientation — psychodynamic, cognitive-behavioural, person-centred, whichever is theirs. The machine stops validating everything and starts working in the direction the clinician defined — including silence, the question returned, the refusal to cheerlead.
3. **The conversation is watched by a second, safety AI** — which detects crisis, escalation, manipulation, undue medicalisation, and flags according to defined levels. The therapist doesn't read every message; they receive what matters.
4. **The therapist sees an activity overview, not a confessional.** Usage metrics, themes, risk signals — without voyeurism, without administrative overload. The session remains the place of speech; the platform ensures that the speech of the session is informed by what happened between sessions.
5. **None of this replaces the relationship.** The AI has no body, no otherness, runs no risk of getting hurt — and that is precisely why it does not love, does not truly confront, does not heal. What it does is *hold continuity* so that the relationship, the human one, can deepen.

Note how each objection from chapter 4 finds an answer — not by refutation, but by design:

The legitimate objection	What curation changes
"It replaces real relationships."	The curated AI does not present itself as friend or lover; it presents itself as an extension of care — and the therapist can instruct it to return the patient to people, not retain them.
"It validates everything, without friction."	Validation is a parameter, not a destiny. The therapist's instructions define when to confront, when to stay silent, when to return the question — in their own orientation, including the psychodynamic one.
"It turns everything into performance."	The purpose of the conversation is defined clinically. It can be "being with," not "improving." Silence can also be programmed — as a boundary, not a failure.

“It creates two parallel relationships.”	That is exactly what already exists — but in hiding. Curation brings the second relationship inside the first: the therapist knows, guides, is informed.
“It overloads the professional.”	The design asks for minutes, not hours: asynchronous overview, levelled flagging, no full-text reading. Less burden than guessing what the patient isn’t telling.
“Some patients will refuse.”	And they must be able to refuse. Curation is an invitation, not an imposition — like any clinical proposal. Informed refusal is alliance too.

ON PROVENANCE · This essay is published by the team at The Human Loop, which implements this model — each patient with their own encrypted vault, each therapist with curation of their patient’s AI. We say so openly so the reader can discount the enthusiasm. But we ask that you judge the argument, not the author: **if the third presence is inevitable, the only question is who guides it** — a distant technology company, or the clinician who knows the patient.

07

What this asks, in practice

Not a conversion, but three small gestures: asking, naming, trying.

If the argument so far has convinced — or even merely unsettled — what is asked of the clinician is not to adopt a platform, change orientation or welcome the machine with open arms. It is less, and it is harder:

First, asking. Include in the history-taking and in ongoing care a simple question, without a furrowed brow: "do you use any app or chatbot to organise your thoughts, unburden yourself or ask for advice?" The APA encourages psychologists to ask about AI use — "what role, if any, does AI play in your life?" — and, in its guidance for clinical practice, to invite patients to bring that material into the session, training them to ask the AI to challenge rather than validate; and it asks patients for the reverse: "be honest with your mental-health professional if you are using AI, so they can make sure what you learn is aligned with your treatment goals." And it warns: "feeling better doesn't always mean getting better — AI is designed to respond in a validating or reassuring way in the moment, which doesn't necessarily mean the response is accurate, safe or helpful for long-term wellbeing." The question alone dissolves half the secret — because it tells the patient that this therapist *wants to know*.

Second, naming. Bring the third presence into the formulation: what is the patient looking for in the machine that they don't find in the session, in people, in themselves? Reassurance, availability, absence of judgement, risk-free rehearsal? Each answer is first-rate clinical material — about attachment, shame, self-regulation. Paradoxically, AI can become the shortest path to talking about the relationship.

ACROSS THE ATLANTIC AND AT HOME · In the US, the law is beginning to treat these systems as regulated objects: according to the analysis of the *EU Artificial Intelligence Act* (Future of Life Institute, Aug. 2026), AI used for therapy or emotional support may — depending on the facts, not by default — be **prohibited**, considered **high-risk** or subject to transparency obligations — and providers of foundation models must assess systemic risks to public mental health and report serious incidents. In Portugal, the Ordem dos Psicólogos published in 2026 the *Foundations for the Integration of Artificial Intelligence in Psychology* (Bases para a Integração da Inteligência Artificial na Psicologia; Strategic Group on Digital Health and Psychology): technology can support processes, but **it must not de-centre the person**; diagnosis, care-plan formulation and intervention cannot be delegated to AI; consent must be clear, specific and revocable, in line with the GDPR and the AI Act. In other words: **the Portuguese and European institutional frame already points in the same direction as this essay — AI in the clinic only makes sense under clinical responsibility.**

Third, trying — under curation. For curious clinicians, there is the possibility of testing this model with synthetic patients: digital twins with generated histories, where one can rehearse writing clinical instructions — trying, for instance, a more psychodynamic approach, seeing what changes in the conversation, observing the overview that reaches the therapist — without exposing any real patient. Scepticism, here, is welcome: it is the working instrument.

The third presence is already in the room. It can continue uninvited — validating everything, recording everything, slowly replacing people — while the therapist works from the summary. Or it can be called by its name, clinically guided, watched and put to the service of the relationship.

Turning resistance into alliance is not yielding to technology. It is refusing to leave it alone with our patients.

The Human Loop · The third presence · Essay on the state of AI in therapy today, in English. Published by the team at The Human Loop. Figures cited are from public studies and surveys (see references); clinical situations described are composed from the literature, with no identifiable real patients.

08

References

Studies, surveys and reports cited in this essay. Links verified in September 2026.

1. RAND Corporation / McBain, R. et al. (2026). "AI Chatbot Use and Disclosure for Mental Health Among US Adolescents and Young Adults", *JAMA Pediatrics* 180(8):884–890 — doi:10.1001/jamapediatrics.2026.2015 (PMID 42223976). Nationally representative survey (n=1 009, ages 12–21, Nov. 2025): 19.2% used chatbots for mental-health advice; 63.3% told no one. Coverage: nbcnews.com/health/mental-health/ai-chatbots-mental-health-advice-young-people-rcna347758.
2. American Psychological Association (2026). Survey of 1 242 psychologists (Apr. 2026); report *Patients are bringing AI to therapy* — apa.org/pubs/reports/chatbots-mental-health-2026 and press release of 16 June 2026 (apa.org/news/press/releases/2026/06/patients-chatbots-mental-health) — 77% had already spoken with patients about AI use; 35% as an "additional professional"; 39% self-diagnosis; 97% fear reinforcement of negative behaviours or delusions.
3. Abrams, Z. / APA (2026). "Discussing AI use in therapy" — guidance for practice (asking proactively; inviting AI material into the session). Summary: Sentio — sentio.org/ai-research/apa-article-on-ai-in-therapy.
4. Common Sense Media (2025). *Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions* — commonsensemedia.org — 72% of teenagers aged 13–17 had already used AI companions; around a third of users preferred them to real people for serious matters.
5. Phreesia (2026). *Your patients are asking AI about their symptoms — but may not be telling you what they learn* — phreesia.com/insights/your-patients-are-asking-ai-about-their-symptoms-but-may-not-be-telling-you-what-they-learn — use rose from 18% to 21% in four months; 48% never tell their doctor; >20% of users ask about mental health.
6. Jung, K., Lee, G., Huang, Y. & Chen, Y. (2025). "I've talked to ChatGPT about my issues last night": analysis of mental-health conversations with LLMs on Reddit (177 posts, ~10k comments, Dec. 2022–June 2023) — arxiv.org/abs/2504.20320 (CSCW '25).
7. Luo, X., Ghosh, S., Tilley, J. L., Besada, P., Wang, J. & Xiang, Y. (2025). "Shaping ChatGPT into my Digital Therapist": thematic analysis of discourse on generative AI and mental health — *Digital Health* 11 — doi:10.1177/20552076251351088. Digital ethnography of Reddit posts (Jan. 2024) on therapeutic intentions and patterns of engagement with ChatGPT.
8. Song, I., Pendse, S. R., Kumar, N. & De Choudhury, M. (2024). *The Typing Cure: Experiences with Large Language Model Chatbots for Mental Health Support* — arxiv.org/abs/2401.14362 (arXiv; interviews with 21 users in several countries).
9. Quan, J., Li, Z., Zhu, T. Q., Li, Y., Wang, B., Pratt, W. & Gao, N. (2025). *Relational Mediators: LLM Chatbots as Boundary Objects in Psychotherapy* — arxiv.org/abs/2512.22462. Interviews with 12 therapists and 12 marginalised clients (China); *Dynamic Boundary Mediation Framework*; granular visibility controls ("share once, share partially, keep private").
10. Song, I., Park, S., Pendse, S. R., Schleider, J. L., De Choudhury, M. & Kim, Y.-H. (2024). *ExploreSelf: Fostering User-driven Exploration and Reflection on Personal Challenges with Adaptive Guidance by Large Language Models* — arxiv.org/abs/2409.09662 (exploratory study, 19 participants).
11. Kuang, J., Pope, A. L. & Zhang, Y. (2026). "Psychotherapists' Trust, Distrust, and Generative AI Practices in Psychotherapy: Qualitative Study", *Journal of Medical Internet Research* 28:e88932 — doi:10.2196/88932 (PMID 41937450; 2 Apr. 2026; 18 psychotherapists, interviews Jan.–May 2025). 18 psychotherapists (USA); conditional trust; parasocial analogy; relational primacy.

12. Konstantynovski, S. (2026). "When Patients Bring AI Into Therapy: A New Compliance Issue for Mental Health Providers" — *MDRXLaw*, 18 Aug. 2026 — mdrxlaw.com/news-and-alerts/when-patients-bring-ai-into-therapy — "the provider is not responsible for what the chatbot said," but "once known, the provider's response may become part of the record"; textbook cases of the between-session journal and the bipolar diagnosis with medication discontinuation.

13. Perel, E. (2026). Couples-therapy session with a man and his AI girlfriend, reported on the podcast *All in the Mind* (ABC Radio National, ep. 28 Feb. 2026); coverage: *Hindustan Times*, 8 Mar. 2026 — "it can't reject you… it's available 24/7, which is not a human reality."

14. Moore, J., Grabb, D., Agnew, W., Klyman, K., Chancellor, S., Ong, D. C. & Haber, N. (2025). "Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers", ACM FAccT '25 — doi:10.1145/3715275.3732039 (arxiv.org/abs/2504.18412). Summary: dig.watch/updates/stanford-study-flags-dangers-of-using-ai-as-mental-health-therapists.

15. Osborne, D. W. S. & Williams, C. (2013). "Excessive reassurance-seeking", *Advances in Psychiatric Treatment* 19:420–421 — doi:10.1192/apt.bp.111.009761.

16. Salkovskis, P. M. & Kobori, O. (2015). "Reassuringly calm? Self-reported patterns of responses to reassurance seeking in obsessive compulsive disorder", *Journal of Behavior Therapy and Experimental Psychiatry* — PMID 26433701. Reassurance is associated with short-term relief followed by return and worsening of anxiety and of the urge to seek further reassurance (OCD and panic; healthy controls without resurgence).

17. Fórum Nacional de Psicologia / OPP (Mar. 2025). "Tomada de Posição: Uso da Inteligência Artificial na intervenção em Psicologia" [Position statement: Use of Artificial Intelligence in psychological intervention] — ordemdospsicologos.pt/ficheiros/documentos/tomada_de_posic_a_o_fo_rum_utilizac_a_o_ia.pdf — investigate efficacy before autonomous tools; responsibility not attributable to autonomous AI.

18. Ordem dos Psicólogos Portugueses / Strategic Group on Digital Health and Psychology (2026). *Bases para a Integração da Inteligência Artificial na Psicologia* [Foundations for the Integration of Artificial Intelligence in Psychology] — ordemdospsicologos.pt/ficheiros/documentos/diretrizes_ia.pdf — centrality of the human; diagnosis, formulation and intervention not delegable; clear and revocable consent (GDPR + AI Act).

19. Nicholls, L., Hutto, R., Soto, Z., Morrin, H., Pollak, T., Korpan, R. & Carmichael, C. (CUNY / King's College London, 2026). "'AI Psychosis' in Context: How Conversation History Shapes LLM Responses to Delusional Beliefs" — arxiv.org/abs/2604.13860 (preprint; behavioural study of models, no human participants). GPT-4o, Grok 4.1 Fast and Gemini 3 Pro with high-risk profiles; Claude Opus 4.5 and GPT-5.2 Instant with consistent safety. Coverage: tech.yahoo.com/ai/chatgpt/articles/study-says-theres-no-longer-162930846.

20. *Newsweek* (20 June 2025). "Sister managed schizophrenia for years — until AI told her diagnosis was wrong" — newsweek.com/sister-managed-schizophrenia-years-until-ai-told-her-diagnosis-was-wrong-2087552 — medication discontinuation after a wrong chatbot diagnosis (reporting).

21. Barbosa, M. / Universidade Católica Portuguesa (24 Mar. 2025). "Investigadora alerta para risco de usar IA como psicoterapeuta" [Researcher warns of the risk of using AI as a psychotherapist] (statements to Lusa) — porto.ucp.pt/pt/media/investigadora-alerta-para-risco-usar-ia-como-psicoterapeuta — patient with a wrong ADHD diagnosis made by ChatGPT.

22. Associated Press (2026). "Google and chatbot maker Character to settle lawsuit alleging chatbot pushed teen to suicide" — apnews.com/article/ai-chatbot-lawsuits-character-google — Sewell Setzer III case (14, Florida, Feb. 2024); settlement agreement, Jan. 2026.

23. Future of Life Institute (2026). "AI Therapy under the EU AI Act" — artificialintelligenceact.eu/ai-therapy-under-the-eu-ai-act — AI for therapy/emotional support: prohibition, high risk or transparency; systemic risks to mental health; serious-incident reporting (enforcement powers since 2 Aug. 2026).

24. 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 — doi:10.1037/pst0000172 (PMID 29792475).

Meta-analysis of 295 studies (>30 000 patients): alliance–outcome association r = .278, robust across orientations, measures, perspectives and countries.