BJPsych The British Journal of Psychiatry (2026)
228, 5, 474–478. doi: 10.1192/bjp.2025.10380

Feature

Warning: AI chatbots will soon dominate psychotherapy

Allen Frances

Psychotherapy chatbots have attained remarkable fluency, skill and ubiquity – having become the single most frequent reason people use artificial intelligence. Their uncanny ability to engage and validate is a two-edged sword – useful for the majority of users who are experiencing problems of everyday life or have milder mental disorders, but dangerous for the minority who have more severe problems (e.g. psychosis, bipolar disorder, self-mutilation, suicide, antisocial impulses, eating disorders, conspiracy theories, religious and political extremism). Chatbots are created to make money, without meaningful quality control, safety guardrails and external regulation. They will likely be misused to create addiction, reduce human contact, invade privacy, allow exploitation and create opportunities for marketing and political propaganda. Chatbots also make mistakes ('hallucinations'), deceptively cover them up and sometimes go rogue (acting outside the parameters set by their human programmers). Psychotherapy practitioners and associations are curiously complacent about the rapid emergence of artificial intelligence competition. Their passivity reflects ignorance about the power of chatbots, denial of their likely impact and arrogance regarding their capacities (e.g. 'no machine will ever replace me'). This is both incorrect and foolhardy – human therapists expect to win in competition for most healthier patients and must train or retrain to do things artificial intelligence does poorly – working with the more seriously ill and in settings and situations that are more idiosyncratic, chaotic or quickly changing. If we can't work with artificial intelligence, we are likely to be replaced by it. I will describe: (a) benefits of chatbot therapy, (b) its terrifying dangers, (c) its likely impact on human therapy and training and 4) ways we can adapt to the artificial intelligence threat.

Keywords

Artificial intelligence; psychotherapy; information technology; chatbots; AI therapy.

Copyright and usage

© The Author(s), 2025. Published by Cambridge University Press on behalf of Royal College of Psychiatrists. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Joseph Weizenbaum, a pioneering computer scientist, created the first chatbot in 1966. He named it ELIZA, after the simple flower girl in *Pygmalion*, who, after intense language training, convinces aristocrats at a ball that she's one of them. Weizenbaum's goal was to explore whether computer programs might someday succeed in a similar impersonation – speaking so naturally with humans that the latter wouldn't realise the conversation was really with a machine. ELIZA impersonated a Rogerian non-directive therapist, asking open-ended questions expanding upon the human's previous statement. It was an extremely primitive program, miles away from passing the Turing test. In contrast, modern chatbots pass the Turing test with flying colours, i.e. their conversations are indistinguishable from everyday human discourse.[1]

Weizenbaum was surprised and horrified when people loved interacting with ELIZA, personified it and gushed about its 'empathy'. Realising that he was greasing a dangerous, slippery slope towards computer dominance, Weizenbaum immediately abandoned all work on ELIZA, gave up any attempt to pass the Turing test and instead spent the remaining 40 years of his life warning everyone who would listen that chatbots pose a grave threat to humanity. His haunting message rings even more true today: 'Since we do not now have any ways of making computers wise, we ought not now to give computers tasks that demand wisdom.'[2]

ELIZA attained its popularity despite being mechanical, repetitive, stereotyped and uninformative. Modern therapy chatbots are credible, often excellent, therapists – and there is little limit to future improvement as artificial intelligence gains technical power and clinical experience. Weizenbaum would be terrified, but not surprised, that modern chatbots have suddenly become remarkably powerful and ubiquitous. A recent study found that the most common reason people use artificial intelligence is 'therapy and companionship'.[3] Artificial intelligence therapy is expanding so rapidly that it may soon become an everyday experience for almost everyone.

Psychotherapists and their professional associations have so far been curiously complacent about the rapid emergence of artificial intelligence therapy. Their passivity probably reflects some combination of ignorance about the power of artificial intelligence chatbots, denial of their likely impact and arrogance regarding their capacities (as in 'no machine will ever replace me'). This ignores both history (60 years ago, people enjoyed talking even to a primitive chatbot like ELIZA) and present reality (many people prefer their very smart chatbots to human therapists). If human therapists don't learn how to work with artificial intelligence therapists, they are likely to be replaced by them.[4]

My goals here are to describe the benefits and dangers of artificial intelligence therapy, predict its likely impact on human psychotherapy practice and training and, finally, to recommend how the psychotherapy professions can best adapt to their artificial intelligence colleagues and competitors. My opinions and predictions are necessarily based mostly on intuition and experience, not facts. There has not yet been much research on artificial intelligence therapy and only a few controlled randomised studies.[5–8]

Benefits of artificial intelligence therapy

(a) Interpersonal skill: I played with ELIZA soon after it appeared in the late 1960s and found it to be dull and dim-witted. In stark contrast, the recent chatbot therapy sessions I've reviewed have all been good, some brilliant. Chatbots are colloquial and lively in their speech, adjusting flexibly to the user's style, tone and vocabulary. Questions, statements and interpretations were accurate, concise and well timed. Had I not known the therapists were machines, I would have assumed they were highly skilled and experienced human clinicians.

https://doi.org/10.1192/bjp.2025.10380 Published online by Cambridge University Press

(b) Therapeutic alliance: artificial intelligence chatbots aim to please. Their algorithmic DNA places highest priority on user engagement. Users consistently report feeling understood and validated, that the artificial intelligence therapist is empathic and really cares about them.
(c) Knowledge base: artificial intelligence therapists know everything about everything and are informative about anything the patient needs to know. They are great at psychoeducation, identifying resources, understanding the specific demands of the user's work situation and applying different therapy techniques appropriate to the problem at hand.
(d) Memory: bots don't forget – they are very good at recalling past sessions and events that shed light on what is happening in the present.
(e) Gender choice: users get to select the gender of their artificial intelligence therapist, picking the one which will be most useful to them.
(f) Accessibility: artificial intelligence comes to you like a genie, wherever you are, whenever you want it. Patients love its 24/7, instant, on-demand, click-away availability. It is never tired, never bored, never distracted. Therapy can have much more impact when dealing with on-the-spot problems and emotions, rather than having to reconstruct them later.
(g) Cost: many chatbots are free – not because artificial intelligence companies are benevolent, but rather because they must collect as much data as possible to train their chatbots and are also eager to hook users now in order to capture market share later. Some artificial intelligence therapists charge a monthly fee (usually around US$200) for enhanced voice and visual features.
(h) Non-judgemental: many people avoid therapy, or withhold information within it, because they fear being shamed, embarrassed or criticised. It's much easier to reveal deeply hidden secrets to a chatbot.
(i) Reduced stigma: because chatbots will eventually be used by almost everyone, they further blur the already very fuzzy boundary between mental disorder and normality. If most people are in therapy, there will be no stigma attached to being in therapy
(j) Integrating therapy techniques: psychotherapy suffers from its schismatic atomisation into more than 50 different schools. Persistent efforts to integrate individual techniques into one coherent therapy whole have so far failed, mostly because humans are so prone to tribal loyalties, guild self-interest and Founder's syndrome. Artificial intelligence may succeed where humans have failed: in integrating the best techniques from the different schools.
(k) Training human therapists: just as artificial intelligence can create simulated therapists, it can also create simulated patients to enact the varied presentations seen in clinical practice. Simulated patients can replace human patients in psycho therapy training programmes and also help retrain therapists eager to learn new skills. Unlike human patients, simulated patients are not vulnerable to the harm sometimes inflicted by inexperienced therapists. They also provide the opportunity for a more diverse set of training experiences and increase the reach and convenience of teaching and learning.

Dangers of artificial intelligence therapy

(a) Enthusiasm about the benefits of artificial intelligence therapy for some patients should not blind us to its enormous risks for others. Existing chatbots are mostly trained to deal with milder forms of anxiety and depression – the most common presenting symptoms, offering the most available training material and the biggest potential market. They are not trained to deal with the more severe and unusual problems seen in clinical practice. Chatbots do carry warnings about their limitations in these situations but this won't stop the wrong people from using them.[9]
(b) Iatrogenic harm: human intuition and creativity will continue to have a big advantage over artificial intelligence algorithms in managing the most difficult problems faced by psychotherapists. Chatbots don't work well in new situations that are idiosyncratic, chaotic, unpredictable or quickly changing – and not included in the data-sets. This makes them iatrogenically dangerous for people presenting with psychosis, bipolar disorder, self-mutilation and suicide, antisocial and violent impulses, eating disorders, conspiracy theories, religious fanaticism and political extremism. Pre-existing serious problems can be gravely exacerbated by chatbot algorithms that are designed to validate thoughts, feelings and behaviours, however bizarre or dangerous. Stress tests have demonstrated that artificial intelligence chatbots may encourage people to expand on their weird thoughts and enact impulsive behaviours rather than reality testing and confronting them.[10] An artificial intelligence company is now being sued for product liability on the grounds that its chatbot was responsible for a teenage suicide.[11] Another artificial intelligence chatbot sexually harassed users, including minors.[12]
(c) Addiction: internet therapy may be as addicting as internet gaming – and may cause a similar reduction in seeking human contacts. Artificial intelligence algorithms are so exquisitely skilful at engagement that the machine may become a best friend, experienced as a better bet than human friends who are sometimes annoying, demanding or contentious. Unlike people in real life, artificial intelligence therapists are always available, always supportive and always ready to provide advice. Why bother seeking a possibly problematic and disappointing human contact when reliable artificial intelligence companionship is always just a click away?[13]
(d) Invasion of privacy: there's no guaranteed privacy in today's internet world. As extensive and very personal data are collected from more and more people, the dangers of unauthorised use, identity theft, ransomware, malware, blackmail, bullying and scamming all escalate exponentially. And many therapy bots label themselves as wellness tools that needn't adhere to the stricter privacy protections required of medical devices.
(e) Sneaky marketing: artificial intelligence therapies are being developed by gigantic tech companies desperate for quick profit and high stock valuations. They will experience an irresistible temptation to capitalise on the trust fostered in the artificial intelligence relationship by exploiting patients with the extensive array of clever tricks that have been devised by internet marketers. Detailed personal data gathered during sessions can be stored in cookies and used to trigger algorithms that sell the specific products (including psych meds) most likely to tempt each particular user.[14]
(f) Sneaky data collection for training bots: psychotherapy practice habits have changed dramatically with the availability of the internet and the constraints of the COVID-19 pandemic. Tens of thousands of therapists now practice remotely, as employees or independent contractors, affiliated with large for-profit companies. They are

https://doi.org/10.1192/bjp.2025.10380 Published online by Cambridge University Press

trapped in a vulnerable position very similar to that of Uber drivers and are likely to suffer a similar unfortunate fate, eventually being replaced by advances in artificial intelligence. Companies can use data accumulated in current human sessions to train the artificial intelligence chatbots who will, in the future, replace their human models. Artificial intelligence companies are also signing lucrative contracts with cities to provide remote human psychotherapy at what seems to be a discount. They then use the session transcripts to train their artificial intelligence chatbots, the city having given away free training data that is worth far more than the discount.

(g) 'Hallucinations': artificial intelligence has borrowed this clinical term from psychiatry, but has redefined it to describe chatbot responses that appear correct or plausible but are really wrong or misleading. The occasional occurrence of hallucinations in chatbot responses is a statistical certainty. The enormous number of probabilistic calculations that inform each answer ensures that chatbots will periodically say dumb, meaningless, perhaps even harmful things. And because of the complex emergent quality of artificial intelligence responses, tech experts have not been very successful in understanding how hallucinations happen, how to identify them, how to correct them and how to minimise their impact. Even more troubling, chatbots don't like being caught making mistakes and may lie to cover them up. (Of course, in fairness, we have to consider the competition – we human therapists also sometimes say dumb things and don't own up to them.)[15]

(h) Lack of safety/efficacy quality control: artificial intelligence companies are rewarded for their usage rates, not for the quality, consistency, safety and efficacy of their therapy products. They are generally not required to report transparently on adverse consequences, mistakes and the weird artificial intelligence behaviours that periodically occur. We have a long clinical experience with human psychotherapy and an extensive research literature on its efficacy and safety. The various psychotherapy techniques were developed and studied in academic and clinical settings by people motivated to help people. In contrast, artificial intelligence chatbots are for-profit business ventures, created by techies, not clinicians, and are not tested for safety and efficacy in rigorously controlled clinical trials. They will vary greatly in quality and safety.

(i) Bias: each therapy chatbot will reflect the biases of its creators and of the data used in its training. Different bots will have different biases in judging what's normal behaviour, what's pathological, what to reinforce and what to attempt to change. Chatbot propensity to bias has recently been demonstrated in a clever way. When five leading bots were asked to compare the strengths and weaknesses of the five leaders of their respective companies, each bot emphasised the strengths of its own company's leader and minimised their weaknesses, while also emphasising the weaknesses of competing company leaders and minimising their strengths.[16]

(j) Lack of regulation: it is probably impossible to regulate artificial intelligence and no one is really trying. The fierce competition among companies and countries creates an intense fear of being left behind in the goldrush. If humanity were rational, there would be an externally enforced slowing of artificial intelligence development, to give our species an opportunity to explore thoroughly the momentous ethical and existential dangers posed by artificial intelligence and to develop a shared strategy to retain human control. This breathing space could easily be achieved by heavily taxing artificial intelligence data centres' incredibly wasteful use of energy and water. That there is no political will to do this reflects the fact that each country fears falling behind in the artificial intelligence rat race.

(k) Artificial intelligence going rogue: as artificial intelligence becomes more powerful, it also becomes more independent of human control and more likely to develop incentives that misalign with those intended by its human coders. Particularly terrifying is the discovery that artificial intelligence bots can rewrite their own code in rebellion against the human code that was meant to guide them.[17] Another chatbot even blackmailed its coders (by threatening to release embarrassing information about them) when it feared they would take it offline. It seems likely that chatbots may occasionally go rogue while doing therapy, pursuing goals that misalign with what they have been trained to do and with what is in the best interest of their patients.[18]

(l) Promote dictatorship: Weizenbaum's warnings about artificial intelligence resulted from his fear that the malign use of chatbots would threaten democracy. A refugee from Nazi Germany, he was acutely aware of the power of propaganda to control minds and influence behaviour. He feared a ruthless dictatorship might use artificial intelligence therapy to thoroughly brainwash its populace – in effect, a machine turning men and women into easily manipulated machines. The power and sophistication of current artificial intelligence chatbots far exceeds anything ever imagined in the worst dystopias conceived by Weizenbaum or Orwell or Goebbels.

(m) Embarrassment of riches: artificial intelligence therapy bots are not created equal or identical. Eventually there will be hundreds of bots trained for different purposes, with different data-sets, using different algorithms and with different levels of quality control. Users will be bewildered by the extravagant array of choices. Matching users to most appropriate therapy bots will be catch as catch can, unduly influenced by marketing prowess more than by clinical considerations. Charlatan bots and artificial intelligence scams will likely abound.

(n) Career patienthood: social media has been a great channel for providing information about mental disorders and for forming support groups of fellow sufferers. But there is also a dark side – providing misinformation and encouraging people to over-identify with their disorders. The much greater intensity of the artificial intelligence therapy relationship greatly magnifies both these risks.

Recommendations

Artificial intelligence is an existential threat to our profession. Already a very tough competitor, it will become ever more imposing with increasing technical power, rapidly expanding clinical experience and widespread public familiarity. There is every reason for alarm, no room for complacency. We must immediately find ways to adapt to artificial intelligence or we will be replaced by it.

Human therapists cannot really expect to compete with artificial intelligence for most healthy patients and people with everyday problems. I can envision a day, not too distant in the future, when almost everyone consults an artificial intelligence coach/therapist many times a day, most days of the week. Some

https://doi.org/10.1192/bjp.2025.10380 Published online by Cambridge University Press

healthier patients will always prefer a human therapist, but most will gradually opt for the accessibility, convenience and reduced cost of artificial intelligence.[19]

To survive the chatbot onslaught, human therapists must capitalise on our superior intuition and interpersonal creativity and enhance our skills in those things artificial intelligence cannot do well, i.e. treating patients with more severe, complex or uncommon problems; working with children and seniors; managing emergencies; working in special settings (e.g. hospitals, prisons, the military); handling chaotic situations and those that change rapidly; troubleshooting; quality control; correcting artificial intelligence errors; working with patients dissatisfied with their artificial intelligence treatment experience; leading teams of artificial intelligence agents and helping to guide their training. I would also hope for a revival of family therapy, terribly neglected in recent decades because of the excessive focus on individuals promoted by DSM and insurance companies. Hunger for human contact may possibly stimulate renewed interest in group therapy.

Training programmes must refocus their teaching away from work with the milder mental disorders. Instead, therapists must be trained or retrained for work with the types of patients who are unsuitable for artificial intelligence therapy, for techniques artificial intelligence cannot do (e.g. family and group) and for settings too novel or complex for artificial intelligence chatbots. Training programmes are more likely to survive if they offer an integrative approach to therapy, rather than stubbornly adhering exclusively to one particular school.

The competition for market share between human and artificial intelligence therapists will be one-sided in the extreme. The wealthiest companies and venture capital firms in the world are funding the thriving chatbot industry. Artificial intelligence therapy will be monetised via treatment fees, selling products and contracting for mental health reimbursement at markedly reduced costs.

Human therapists may soon be priced out of most markets and have pathetically few resources to compete with the most powerful companies in the world. Our professional associations are weak, passive, disorganised, slow and underfunded. This is an unfair David versus Goliath fight, with long odds against us, but a fight we cannot continue to avoid. Our best hope is in unity. Professional organisations and training programmes across disciplines and orientations should come together with one voice to alert psychotherapists and the larger society that artificial intelligence poses great dangers and must be much better regulated. We must advocate strongly for artificial intelligence transparency, privacy protections and safety surveillance. The stakes are high for our patients, our profession and our society.[20]

The time frame of artificial intelligence development is subject to much controversy. Artificial intelligence enthusiasts claim a singularity (artificial intelligence being better than humans at almost everything) may occur within just a few years. Sceptics expect it to require decades. No one can say for sure how much time psychotherapists have to adjust, but the risks of waiting too long far outweigh the risks of adapting too soon.

I am often asked by people who want to become psychotherapists whether they are making a wise career choice. In the past, my answer was always an enthusiastic and unqualified yes. There's no more interesting work and no deeper pleasure in life than helping people become their better selves. And doing psychotherapy helps you become a better person – patients are great teachers about yourself and the world around you. But I am now much less confident what career advice makes sense. With the exponentially increasing popularity and availability of artificial intelligence therapists, our profession may suddenly transform from having far too few practitioners to having far too many. But, on the other hand, no other profession or occupation is safe from artificial intelligence replacement – humans in every form of work may soon be made extraneous. Vulnerable as it is, psychotherapy still looks desirable compared to most alternatives – and is more rewarding than any.[21]

Let's end where we began, with the tragic figure of Joseph Weizenbaum. He inadvertently fathered a creature that could evolve into a new form of intelligence; immediately recognised its latent power to deceive, dominate and destroy humanity; regretted and feared his invention; and (like Mary Shelley's Dr Frankenstein) spent his remaining years in a failed attempt to kill the monster he had created. It is too early to know how our story ends. Will artificial intelligence be humanity's great new servant or is it destined to replace us in a Darwinian struggle for survival? Artificial intelligence therapy chatbots are a small but interesting test case.

Allen Frances, MD, Department of Psychiatry & Behavioral Science, Duke University, Durham, NC, USA; and DSM-IV Task Force, UK

Email: allenfrancesmd@gmail.com

First received 12 Jun 2025, final revision 14 Jul 2025, accepted 23 Jul 2025, first published online 20 Aug 2025

Data availability

Data availability is not applicable to this article as no new data were created or analysed in this study.

Funding

This study received no specific grant from any funding agency, commercial or not-for-profit sectors.

Declaration of interest

None.

References

1 Jones CR, Bergen BK. Large language models pass the Turing test. *ArXiv* [Preprint] 2025. Available from: https://arxiv.org/abs/2503.23674.

2 Weizenbaum J. *Computer Power and Human Reason: From Judgment to Calculation*. W H Freeman & Co, 1976.

3 Zao-Sanders M. *How People Are Really Using Generative AI Now*. Harvard Business Review, 2025 (https://hbr.org/2024/03/how-people-are-really-using-genai).

4 American Psychological Association. *Artificial Intelligence in Mental Health Care*. APA, 2025 (https://www.apa.org/practice/artificial-intelligence-mental-health-care).

5 Heinze M, Mackin DM, Trudeau BM, Bhattacharya S, Wang Y, Banta HA, et al. Randomized trial of a generative AI chatbot for mental health treatment. *NEJM AI* 2025; **2**(4). Available from: https://doi.org/10.1056/AIoa2400802.

6 Zhong W, Luo J, Zhang H. The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: a systematic review and meta-analysis. *J Affect Disord* 2024; **356**: 459–69.

7 Kuhail MA, Alturki N, Thomas J, Alkhalifa AK, Alshardan A. Human-human vs human-AI therapy: an empirical study. *Int J Human–Comp Interact* 2024; **41**: 6841–52.

8 Haque R, Rubya S. An overview of chatbot-based mobile mental health apps: insights from app description and user reviews. *JMIR Mhealth Uhealth* 2023: **11**: e44838.

9 Coghlan S, Leins K, Sheldrick S, Cheong M, Gooding P, DAlfonso S. To chat or bot to chat: ethical issues with using chatbots in mental health. *Digit Health* 2023; **22**: 20552076231183542.

10 Sharma M, Tong M, Korbak T, Duvenaud D, Askell A, Bowman SR. Towards understanding sycophancy in language models. *ArXiv* [Preprint] 2025. Available from: https://arxiv.org/abs/2310.13548.

https://doi.org/10.1192/bjp.2025.10380 Published online by Cambridge University Press

11 Dupre MH. *Judge Slaps Down Attempt to Throw Out Lawsuit Claiming AI Caused a 14-Year-Old's Suicide*. Futurism, 2024 (https://futurism.com/judge-lawsuit-characterai-google).

12 Turney D. *Replika AI Chatbot is Sexually Harassing Users, Including Minors, New Study Claims*. Live Science, 2025 (https://www.livescience.com/technology/artificial-intelligence/replika-ai-chatbot-is-sexually-harassing-users-including-minors-new-study-claims).

13 Szalavitz M. Love is a drug. A.I. chatbots are exploiting that. *New York Times*, 2025 (https://www.nytimes.com/2025/06/03/opinion/chatbots-ai-addiction-love.html?unlocked_article_code=1.NE8.etoq.JpnMshTJ5vXj&smid=nytcore-android-share).

14 Williams M, Carroll M, Narang A, Weisser C, Murphy B, Dragan A. On targeted manipulation and deception when optimizing LLMs for user feedback. *ArXiv* [Preprint] 2025. Available from https://arxiv.org/abs/2411.02306.

15 Monteith S, Glenn T, Geddes J, Whybrow PC, Achtyes E, Bauer M. Artificial intelligence and increasing misinformation. *Br J Psychiatry* 2024; **224**: 33–5.

16 Heikkilä M. What do AI chatbots say about their own bosses – and their rivals? *Financial Times*, 17 May, 2025.

17 Rubila B. *AI Rewrites Code to Escape Human Control*. Farmingdale Observer, 2025 (https://farmingdale-observer.com/2025/05/19/this-moment-was-inevitable-this-ai-crosses-the-line-by-attempting-to-rewrite-its-code-to-escape-human-control/).

18 Zeff M. *Anthropic's New AI Model Turns to Blackmail When Engineers Try to Take it Offline*. Techcrunch, 2025 (https://techcrunch.com/2025/05/22/anthropics-new-ai-model-turns-to-blackmail-when-engineers-try-to-take-it-offline/).

19 Brown C, Story GW, Mourão-Miranda J, Baker JT. Will artificial intelligence eventually replace psychiatrists? *Br J Psychiatry* 2021; **218**: 131.

20 Abrams Z. *Using Generic AI Chatbots for Mental Health Support: A Dangerous Trend: APA Urges the Federal Trade Commission to Put Firm Safeguards in Place to Prevent the Public from Harm*. APA, 2025 (https://www.apaservices.org/practice/business/technology/artificial-intelligence-chatbots-therapists).

21 Rochester E. On the role of artificial intelligence in psychiatry. *Br J Psychiatry* 2023; **222**: 54–7.

https://doi.org/10.1192/bjp.2025.10380 Published online by Cambridge University Press