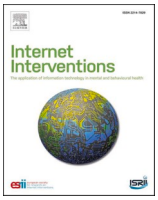
Internet Interventions 32 (2023) 100621

Contents lists available at ScienceDirect

Internet Interventions

journal homepage: www.elsevier.com/locate/invent



A new era in Internet interventions: The advent of Chat-GPT and AI-assisted therapist guidance

ARTICLE INFO

Keywords
Internet-based treatments
Artificial intelligence guided interventions
Therapeutic alliance
Digital self-help programs
Conversational agents
Client preferences

There are a wealth of different internet-based treatments available that are commonly classified into two categories: live, real time video therapy – such as telepsychiatry using Zoom – or digital self-help programs with or without human guidance that is often asynchronous (Andersson et al., 2019). Combinations of these two approaches also exist along with the blending of internet-based treatments with face-to-face therapy (Erbe et al., 2017). Understanding the significance of differences in approaches is complicated by inconsistent and confusing terminology used to describe the approaches (Smoktunowicz et al., 2020). Typically, guided internet-based treatments tend to be more effective than pure self-help (Baumeister et al., 2014). According to a recent meta-analysis, therapist-guided internet-based cognitive behavior therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders (Hedman-Lagerlöf et al., in press). Moreover, without support, the effects of internet-based treatment tend to be smaller, and dropout rates increase (Koelen et al., 2022). While clearly beneficial, human guidance cannot easily be scaled to cover the treatment needs (Fairburn and Patel, 2017). Consequently, there has been a growing interest in improving the effectiveness of internet-based interventions with lower degrees of guidance. While some advancements have been made using automated support (e.g., Titov et al., 2013), and support on demand (Hadjistavropoulos et al., 2019), there is still room for improving lower-intensity internet-based treatments. There are, in fact, very few pure internet-based self-help studies available, as most of these studies incorporate human contact in the form of diagnostic or screening interviews before treatment begins, or they provide access to human backup support in the event of an emergency or technological challenge during treatment.

With developments in artificial intelligence (AI), study of the integration of AI and internet-based treatment represents an important direction in optimizing internet-based treatments. One possible solution could be to integrate AI, in the form of a conversational agent, into the platforms (Ly et al., 2017; Morris et al., 2018) either as a tutor that becomes a virtual psychoeducational coach available to the client at all hours of the day, or taking on the role of a psychotherapist, giving constructive feedback and posing clinically informed inquiries such as Socratic questions that could directly help the client when analyzing negative automatic thoughts or maladaptive core beliefs. In this context, it is important to acknowledge that automated interactions with a computer have existed for a long time, but were only widely implemented more recently (Epstein and Klinkenberg, 2001). Arguably, AI can be seen as a way to improve automated treatments like the early Eliza that was invented already in the 1960's. With good implementation, the new AI-applications have the potential to significantly improve internet-based treatments and save clinicians time, but it also raises questions about the potential negative implications of using AI and large language models, such as GPT, in this field.

One important limitation is that a machine does not yet possess human-like empathy or emotions, and hence has difficulty understanding the nuances of human language. A conversational agent lacks the emotional intelligence and personal experience of a human being, but that can be hidden by expressing empathy-like utterances. In fact, a recent randomized controlled trial showed that a human–AI collaboration outperformed humans with 19.6 % more empathic conversations in text-based peer-to-peer mental health support (Sharma et al., 2023). Furthermore, GPT-4 has been reported to have an advanced understanding of the theory of mind (Bubeck et al., 2023). However, a conversational agent mimicking empathy and responding appropriately may not be enough. When a human communicates, one may say one thing with words, but the body language might convey a different meaning or emotion. Interpreting irony from a client is also a potential pitfall, which could lead to inappropriate therapeutic suggestions (such as promoting safety-seeking behaviors; Rozental et al., 2018). The use of AI guided interventions with more difficult and severe cases requires careful consideration and likely safeguards such as availability of clinicians who are capable of managing deterioration and other potential negative events.

Research has shown that the therapeutic alliance, which refers to the collaborative relationship between a human clinician and a human

https://doi.org/10.1016/j.invent.2023.100621
Received 3 April 2023; Accepted 3 April 2023
Available online 11 April 2023

client, is a robust predictor of treatment outcomes (Wampold and Flückiger, 2023). However, while AI and clients may agree on treatment tasks and goals (such as interoceptive exposure to reduce avoidance and anxiety), there are limitations in developing a therapeutic bond (Miloff et al., 2020).

In addition to this, the role of client preference is important, and the knowledge that support is not being provided by a human could be perceived negatively. Nevertheless, there may be some clients where AI guidance is perceived positively and preferred, perhaps similar to how many individuals prefer to self-manage symptoms rather than seek therapeutic support (Andrade et al., 2014).

Assuming AI guidance is offered, it is necessary that clients are informed when they are supported by a computer rather than a human. In a recent experiment on a mental health support platform involving ChatGPT-3, it was found that peer support boosted by AI-generated suggestions received higher ratings than support written solely by humans (Biron, 2023). Nevertheless, Twitter users expressed concern that users were not adequately informed about the possibility of the author not being a real person (Hsu, 2023). The lack of transparency surrounding AI-generated messages could lead to trust issues for patients. It is crucial for the sender to be perceived as a real human being with genuine experiences. For instance, a study on therapist behaviors in internet-based cognitive-behavioral therapy for depressive symptoms found that including self-disclosure in written messages was significantly correlated with improvement at post-treatment (Holländare et al., 2016). While this study was not an experiment directly testing the effects of self-disclosure, it highlights the possible importance of openly sharing personal experiences as a means of enhancing the therapeutic alliance and improving depression treatment outcomes. Would humans appreciate communicating with an AI offering examples of 'lived experience'? Currently, we predict the majority of clients are likely to choose a human therapist over AI-generated messages if given the option. A crucial aspect of psychotherapy remains the therapist's capacity to empathize and establish a personal connection with the patient. Nevertheless, that may change in the future, as AI improves. In an era dominated by in-person treatments as the gold standard, positive remarks about internet-based therapies without face-to-face interactions, such as "I didn't have to look her in the eyes" (Lindqvist et al., 2022), may appear paradoxical. Nevertheless, such statements could signal a shift - the potential for individuals to confide in AI without the pressures of social desirability and stigma. As we venture into an era of increasing reliance on AI, it is crucial to carefully study client preferences, the effects, and consequences of this shift, and, hopefully, determine what works, what requires improvement, and what should be discarded.

Declaration of competing interest

The authors, Per Carlbring, Heather Hadjistavropoulos, Annet Kleiboer, and Gerhard Andersson, hereby declare that they have no known conflicts of interest or financial ties related to the research reported in the manuscript titled "A New Era in Internet Interventions: The Advent of Chat-GPT and AI-Assisted Therapist Guidance."

References

Andersson, G., Titov, N., Dear, B.F., Rozental, A., Carlbring, P., 2019. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 18 (1), 20–28. https://doi.org/10.1002/wps.20610.

Andrade, L.H., Alonso, J., Mneimneh, Z., Wells, J.E., Al-Hamzawi, A., Borges, G., Bromet, E., Bruffaerts, R., de Girolamo, G., de Graaf, R., Florescu, S., Gureje, O., Hinkov, H.R., Hu, C., Huang, Y., Hwang, I., Jin, R., Karam, E.G., Kovess-Masfety, V., Kessler, R.C., 2014. Barriers to mental health treatment: results from the WHO world mental health surveys. Psychol. Med. 44 (6), 1303–1317. https://doi.org/10.1017/S0033291713001943.

Baumeister, H., Reichler, L., Munzinger, M., Lin, J., 2014. The impact of guidance on internet-based mental health interventions—a systematic review. Internet Interv. 1 (4), 205–215. https://doi.org/10.1016/j.invent.2014.08.003.

Biron, B., 2023. Online mental health company uses ChatGPT to help respond to users in experiment—raising ethical concerns around healthcare and AI technology. Bus. Insid. https://www.businessinsider.com/company-using-chatgpt-mental-health-support-ethical-issues-2023-1.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y.T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M.T., Zhang, Y., 2023. Sparks of Artificial General Intelligence: Early Experiments With GPT-4 (arXiv: 2303.12712). arXiv. doi:10.48550/arXiv.2303.12712.

Epstein, J., Klinkenberg, W.D., 2001. From Eliza to internet: a brief history of computerized assessment. Comput. Hum. Behav. 17 (3), 295–314. https://doi.org/10.1016/S0747-5632(01)00004-8.

Erbe, D., Eichert, H.-C., Riper, H., Ebert, D.D., 2017. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J. Med. Internet Res. 19 (9), e6588 https://doi.org/10.2196/jmir.6588.

Fairburn, C.G., Patel, V., 2017. The impact of digital technology on psychological treatments and their dissemination. Behav. Res. Ther. 88, 19–25. https://doi.org/10.1016/j.brat.2016.08.012.

Hadjistavropoulos, H.D., Schneider, L.H., Mehta, S., Karin, E., Dear, B.F., Titov, N., 2019. Preference trial of internet-delivered cognitive behaviour therapy comparing standard weekly versus optional weekly therapist support. J. Anxiety Disord. 63, 51–60. https://doi.org/10.1016/j.janxdis.2019.02.002.

Hedman-Lagerlöf, E., Carlbring, P., Svärdman, F., Riper, H., Cuijpers, P., & Andersson. G. (in press). Therapist-supported Internet-based cognitive behaviour therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. World Psychiatry.

Holländare, F., Gustafsson, S.A., Berglind, M., Grape, F., Carlbring, P., Andersson, G., Hadjistavropoulos, H., Tillfors, M., 2016. Therapist behaviours in internet-based cognitive behaviour therapy (ICBT) for depressive symptoms. Internet Interv. 3, 1–7. https://doi.org/10.1016/j.invent.2015.11.002.

Hsu, J., 2023. Mental health service criticised for experiment with AI chatbot. New Scientist. https://www.newscientist.com/article/2354077-mental-health-service-criticised-for-experiment-with-ai-chatbot/.

Koelen, J.A., Vonk, A., Klein, A., de Koning, L., Vonk, P., de Vet, S., Wiers, R., 2022. Man vs. Machine: a meta-analysis on the added value of human support in text-based internet treatments ("e-therapy") for mental disorders. Clin. Psychol. Rev. 96, 102179 https://doi.org/10.1016/j.cpr.2022.102179.

Lindqvist, K., Mechler, J., Midgley, N., Carlbring, P., Carstorp, K., Neikter, H.K., Strid, F., Von Below, C., Philips, B., 2022. "I didn't have to look her in the eyes"—Participants' experiences of the therapeutic relationship in internet-based psychodynamic therapy for adolescent depression. Psychother. Res. 1–15. https://doi.org/10.1080/10503307.2022.2150583.

Ly, K.H., Ly, A.-M., Andersson, G., 2017. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv. 10, 39–46. https://doi.org/10.1016/j.invent.2017.10.002.

Miloff, A., Carlbring, P., Hamilton, W., Andersson, G., Reuterskiöld, L., Lindner, P., 2020. Measuring Alliance toward embodied virtual therapists in the era of automated treatments with the virtual therapist Alliance scale (VTAS): development and psychometric evaluation. J. Med. Internet Res. 22 (3), e16660 https://doi.org/10.2196/16660.

Morris, R.R., Kouddous, K., Kshirsagar, R., Schueller, S.M., 2018. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. J. Med. Internet Res. 20 (6), e10148 https://doi.org/10.2196/10148.

Rozental, A., Castonguay, L., Dimidjian, S., Lambert, M., Shafran, R., Andersson, G., Carlbring, P., 2018. Negative effects in psychotherapy: commentary and recommendations for future research and clinical practice. BJPsych Open 4 (4), 307–312. https://doi.org/10.1192/bjo.2018.42.

Sharma, A., Lin, I.W., Miner, A.S., Atkins, D.C., Althoff, T., 2023. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. Nat. Mach. Intell. 5 (1), 1 https://doi.org/10.1038/s42256-022-00593-2.

Smoktunowicz, E., Barak, A., Andersson, G., Banos, R.M., Berger, T., Botella, C., Dear, B. F., Donker, T., Ebert, D.D., Hadjistavropoulos, H., Hodgins, D.C., Kaldo, V., Mohr, D. C., Nordgreen, T., Powers, M.B., Riper, H., Ritterband, L.M., Rozental, A., Schueller, S.M., Carlbring, P., 2020. Consensus statement on the problem of terminology in psychological interventions using the internet or digital components. Internet Interv. 21, 100331 https://doi.org/10.1016/j.invent.2020.100331.

Titov, N., Dear, B.F., Johnston, L., Lorian, C., Zou, J., Wootton, B., Spence, J., McEvoy, P. M., Rapee, R.M., 2013. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: randomised controlled trial. PLOS ONE 8 (7), e62873. https://doi.org/10.1371/journal.pone.0062873.

Wampold, B.E., Flückiger, C., 2023. The alliance in mental health care: conceptualization, evidence and clinical applications. World Psychiatry 22 (1), 25–41. https://doi.org/10.1002/wps.21035.

Per Carlbring[a,*], Heather Hadjistavropoulos[b], Annet Kleiboer[c], Gerhard Andersson[d]

[a] *Stockholm University, Department of Psychology, Stockholm, Sweden*

[b] *University of Regina, Regina, Department of Psychology, Saskatchewan, Canada*

[c] *VU Amsterdam, Department of Clinical Neuro- and Developmental Psychology, Amsterdam, Netherlands*

[d] *Linkoping University, Department of Behavioural Sciences and Learning, SE-581 83 Linköping, Sweden*

[*] Corresponding author.
E-mail address: per.carlbring@psychology.su.se (P. Carlbring).