

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/348063596>

# The potential use of artificial intelligence in the therapy of borderline personality disorder

Article in *Journal of Evaluation in Clinical Practice* · December 2020

DOI: 10.1111/jep.13530

---

CITATIONS

2

READS

304

1 author:



Judit Szalai

Eötvös Loránd University

6 PUBLICATIONS 23 CITATIONS

SEE PROFILE

# The potential use of artificial intelligence in the therapy of borderline personality disorder

Judit Szalai PhD 

Institute of Philosophy, Eötvös Loránd University, Budapest, Hungary

## Correspondence

Judit Szalai, Eötvös Loránd University, Múzeum krt. 4/1 1088 Budapest, Hungary.  
Email: szalai.judit@btk.elte.hu

## Funding information

Hungarian Ministry of Innovation and Technology, Grant/Award Number: 120375

## Abstract

This paper explores the possibility of AI-based addendum therapy for borderline personality disorder, its potential advantages and limitations. Identity disturbance in this condition is strongly connected to self-narratives, which manifest excessive incoherence, causal gaps, dysfunctional beliefs, and diminished self-attributions of agency. Different types of therapy aim at boosting self-knowledge through self-narratives in BPD. The suggestion of this paper is that human-to-human therapy could be complemented by AI assistance holding out the promise of making patients' self-narratives more coherent through improving the accuracy of their self-assessments, reflection on their emotions, and understanding their relationships with others. Theoretical and pragmatic arguments are presented in favour of this idea, and certain technical solutions are suggested to implement it.

## KEYWORDS

AI, borderline personality disorder, identity disturbance, self-narrative

## 1 | INTRODUCTION

Borderline personality disorder is a condition that constitutes one of the greatest practical challenges to psychotherapists. Apart from the diversity of its manifestations, what makes BPD especially hard to manage therapeutically is patients' difficulties in interpersonal functioning, due to which they tend to resist therapy, mistrust therapists' motivations, and sabotage treatment. Dropout rates are extremely high.<sup>1</sup> Since pharmacological treatment and hospitalization are inadequate by themselves in BPD, therapy should be sufficiently tweaked to accommodate patients' motivations in the long-term management of this condition.

In addition to being marred by motivational difficulties, BPD treatment is also costly, requiring high levels of service utilization.<sup>2,3</sup> Patients often have limited access to the extensive care needed for substantial improvement due to shortage of health care workers and long waiting lists<sup>1</sup>; also, they may get 'routed out of care through a variety of direct and indirect means',<sup>5</sup> (p. 82.) being labelled 'difficult patients'.

The idea to be proposed in this paper is that AI-supported sessions could constitute an addendum to human-to-human therapy,

potentially significantly improving prospects of recovery. AI assistance could be integrated into mainstream therapeutic efforts such as dialectical-behavioural therapy<sup>5</sup> and mentalization therapy,<sup>6</sup> as well as less established ones such as narrative therapy.<sup>7</sup> The suggestion is that AI using natural language processing could help patients re-author their self-narratives into more coherent and meaningful sequences, in which they view themselves more like agents and less subject to external control, as well as assign more consistent roles to others. The therapy assistant could promote mentalization, detect the inconsistencies between causal attributions and lack of self-attributions of agency in situations in which the subject is the apparent agent, and prompt patients forward by targeted questions towards insights concerning causal connections and self-agency.

## 2 | SELF-NARRATIVES IN BORDERLINE PERSONALITY DISORDER AND THEIR THERAPEUTIC USE

Borderline personality disorder is characterized by disturbed identity and self-image; impaired interpersonal functioning and unstable

[Correction added after first online publication on 19 March 2021: The references 27-32 and the their corresponding in-text citation have been updated in this current version.]

relationships; feelings of emptiness and abandonment; inappropriate and extreme emotional reactions (most typically, fits of rage) and impulsivity (often substance abuse or erratic sexual behaviour). Several accounts of the character and aetiology of BPD have been put forward. Linehan suggests<sup>8</sup> that BPD is best understood as a disorder of emotional dysregulation: the child's genetically given emotional vulnerability and early environmental input characterized by invalidating reactions are responsible for the emerging dysfunctional affective, cognitive, and behavioural patterns. Bateman and Fonagy's mentalization-based model focuses on the hyper responsiveness of the attachment system due to disruptions in early attachment relationships, which result in deficient mentalization abilities. Zanarini identifies two key features: chronic, intense pain or suffering associated with dysphoric affects and cognitions the way in which patients 'both hide this pain and [...] express it'.<sup>9(p. 507)</sup> There also seems to be good reason to give BPD a cognitive account, based on the dysfunctional, rigid beliefs<sup>10,11</sup> and cognitive biases such as dichotomous (black-and-white) thinking and a tendency to interpret neutral interpersonal interactions in a negative way.<sup>12</sup>

Wherever we place the emphasis in accounting for BPD, patients' interpersonal problems and identity disturbance stand out as underlying and, through feedback loops,<sup>13</sup> aggravating their condition and playing a major role in therapeutic difficulties as well. The two issues are strongly related: a fragmented experience of identity corresponds to incoherent self-narratives,<sup>2</sup> ones that manifest poor mentalizing and tend to assign inconsistent roles to others. In the following paragraphs, I will look at the characteristics of BPD patients' self-narratives and therapeutic efforts intended to help patients rework those narratives.

A diagnostic criterion of borderline personality disorder in DSM-5 is identity disturbance,<sup>3</sup> described as 'markedly and persistently unstable self-image or sense of self'.<sup>14(p.663)</sup> This unstable self-image manifests itself in a less coherent self-concept and self-narratives. What are borderline self-narratives like? Interestingly, what used to be believed about borderline patients, viz. that they lack self-knowledge in the sense of not having sufficient insight into their own (individual) mental states, is less characteristic than it is commonly held to be.<sup>16</sup> What seems pronouncedly deficient is the *integration* of mental states in narratives. Patients would tell stories in which a person close to them appears in a very positive and then a very negative light. In other cases, subjects have unclear and contradictory views of their preferences.

Apart from exhibiting inconsistencies, narratives of persons with borderline personality disorder also tend to be deficient in terms of causal coherence. When narrating interpersonal experiences, patients may fail to make the connection between their own and other people's verbal and non-verbal behaviour, especially in affectively intense situations. They may fail to see what triggered their frustrated or depressive states, or may not be able to support their negative judgement of an encounter by details.<sup>17 (p.10)</sup>

Certain deeply ingrained beliefs, such as that of 'embedded badness', may also influence narrative coherence:

[...] a deeply embedded and often unconscious self-perception of inherent badness, that is, evil, defective, worthless, lazy, or ugly. This sense of badness is often not immediately apparent and difficult to measure in research studies since it can be repressed and denied, even to the point that patients can appear grandiose with an inflated self-appraisal for much of the time. The badness can also be projected onto others, such that BPD patients can become mistrustful, avoidant, or denigrating of others as a way to protect against feelings of shame.<sup>18(p. 7)</sup>

Due to their cognitive rigidity and epistemic mistrust, BPD patients are easily stuck with the kind of narratives characterized above. (This is sometimes labelled a state of 'epistemic petrification'.<sup>19</sup>) This can also cause therapeutic difficulties. The patient can slip into what has been labelled 'psychic equivalence mode' and 'only experience his or her own point of view, and it is necessary for the therapist to recognize that alternative viewpoints are unacceptable; any attempt to provide one will simply lead to argument or complaints from the patient that the therapist does not understand'.<sup>6(p. 192.)</sup>

Autobiographical narratives are used by a wide range of therapies. In dialectical behavioural therapy, 'diary cards' are used as a prop to the narration of between-session events, to organize the session in terms of thematic priorities (such as events experienced as threatening; self-harming or therapy-interfering behaviours; quality of life issues). The aim is to improve patients' emotion regulation, stress tolerance, and interpersonal skills based on their narratives.<sup>5</sup>

In mentalization-based therapy, too, the first step in a session is attending to the patient's narrative.<sup>21 (p.ix)</sup> This can constitute the basis of empathic validation. The goal is to help the patient construct a mentalization-based narrative, one in which her own and others' affects and intentions are correctly identified. In addition to narratives about current events and mental life, reconstruction of childhood experiences 'generates a coherent self-narrative assuming a historical continuity of self which may itself be of therapeutic value'.<sup>20(p. 107)</sup>

In metacognitive interpersonal therapy, the therapist aims to promote awareness of emotions and their triggers, focusing on the emotions involved in narrative episodes, looking for non-verbal markers to complement self-narratives, asking clarificatory question to heighten awareness of emotions.<sup>21</sup> Other goals are to empower agency and help distinguish between reality and fantasy. Reframing narratives and drawing attention to parallels between different narrative episodes have a role in both.

Narrative therapy in BPD<sup>7</sup> aims at making patients re-author their narratives, produce a more coherent, meaningful and less negative understanding of their experiences, and learn to distinguish themselves from their condition (externalization). Patients are helped to find connections between their different mental states and follow those in their sequence. Therapists also assist patients in focusing on the important aspects of their narratives, eliminate inconsistencies, clarify the roles of characters and their reciprocal emotional positions.



As these examples illustrate, self-narratives provide the material for various kinds of BPD therapy. These narratives are not only symptomatic of the condition; the incoherent, fragmented ways in which patients view themselves and relate their life events is part of the identity disturbance therapy aims to alleviate. Therapy, it will be suggested below, could benefit from AI-assisted work on self-narratives as complementing face-to-face sessions.

### 3 | THE POTENTIAL USE OF ARTIFICIAL INTELLIGENCE IN BPD THERAPY

Perhaps the greatest challenge for the patient in borderline psychotherapy is dealing with fellow humans. Interpersonal dysfunction is central to BPD<sup>22</sup>; there is also increased sensitivity to interpersonal stressors, mistrust of others' intentions, including those of the therapist, and a perception of therapy aimed at changing thought and behavioural patterns as invalidating. The therapeutic relationship may itself activate attachment-related neural patterns that hinder metacognitive functioning, undermining therapy.<sup>1,23</sup> While human therapists cannot be left out of the loop (for reasons that will be presented in Section 4), an obvious suggestion seems to be the enhancement of therapy by an element in which human-to-human communication does not play a role.

Digitally assisted psychotherapy services are an expanding industry.<sup>4</sup> Guidelines were suggested for such technologies already in 2010, according to which ' [d]ependent on the specific nature of an intervention (eg, the disorder experienced, the intervention methods applied) systems can be developed which aim to support, as well as monitor, specific therapeutic outcomes (e.g. reductions in anxiety levels, improvements in moods, improved coping strategies)'.<sup>24</sup> Today, creating chatbots is a relatively routine task for IT experts, which—with the help of transfer learning—can, with the right specifications, be accomplished within a short time. Chatbots have 'personas', mimic human interaction, use creative images for non-verbal cues.<sup>5</sup> The suggested BPD therapy assistant would be a natural language processing-based AI device that asks targeted questions, offers conversational prompts, and processes the affective tone and content coherence of answers as well as clues of a sense of agency and control.

Technology is now sufficiently advanced to recognize and track not only the emotional tone (positive and negative affective content) of a narrative; fine-grained emotions are also identifiable on the basis of large annotated data-sets.<sup>25</sup> Such tools have been developed with the explicit intention of language-based emotion recognition and enhancing 'human-robot interaction' in, for example, customer feedback systems and chatbots for various purposes. The text is broken down into smaller units (like sentences or paragraphs), in order to categorize emotion expressions appearing in them. Other than obvious classes such as fear, disgust, and joy, a number of more subtle categories (eg, 'caring', 'realization', 'confusion') are also used.

Word choices can be monitored for clues of agency and internal control (whether the patient attributes events, especially with positive

outcomes, to their own agency). The relative use of the active and passive voices can give an indication as to the subjects' sense of acting or being acted upon, which could also help reveal patterns of blaming and self-blaming.

The so-called 'frame semantics' could be of use. Here, 'mental spaces' for speakers' realities are identified within which the referents of words are constructed, as an aid to revealing 'the functional-cognitive foundations of meaning construction'.<sup>26</sup> For instance, the utterance 'I feel emptiness' can be interpreted as the expression of a sense of social vacuum, lack of goals, lack of meaningfulness in one's activities, and so on. The right interpretation may be opaque to the subject themselves; exploring the cognitive frame within which the utterance was made by the therapy assistant may better situate the patient's issue. Digital therapists could analyse a wide range of frame situations and identify the role of concepts or objects/persons referred to.

Apart from identifying affective and semantic patterns, 'digital therapists' could help shape the subject's narrative by leading conversations towards and revealing inconsistencies between self-attributions of mental states, as well as between contradictory judgements concerning others to the patient. Thus, causal reasoning about events of affective significance can be improved; patients can be helped to reduce semantic inconsistencies, negative and antagonistic framing, to identify more important and secondary aspects of emotion-laden issues, to externalize problematic attitudes and behaviour, and to detect and work on problematic assumptions, such as that of embedded badness.

Progress could be monitored in all the above areas: in terms of the coherence and emotional arc of narratives, word choices regarding relationship with others, agency and control. In addition, emotion recognition could be used to monitor patients' progress regarding mood shifts independent of particular narratives.

While some of the above would be special contribution of digital assistants to BPD therapy, ordinary, human-to-human techniques could also be employed. Cognitive therapy techniques like the 'continuum method' could be adapted to AI, prompting the patient to identify more possibilities on the continuum between two ('black' and 'white') extremes, in an effort to attenuate the black-and-white character of their thinking. Externalization, aimed at inducing patients to come to regard problematic attitudes and behaviour as external to themselves, rather than belonging to the 'core' of their personalities, could be boosted by targeted questions. Here, labelling and framing are essential and could be improved in dialogues. Similarly, the method of 'deconstruction' could be used to reduce overgeneralization and to pin down problems more concretely. This again can be achieved by asking targeted question to help patients zero in on the central aspects of their problems.

Since the technology under consideration does not exist yet, for concrete illustration of how BPD addendum therapy with AI assistance could work, I will use conversations from the BPD literature that could without significant change take place between a BPD patient and an AI-based therapy assistant. (Note that these dialogues do not make use of all the capacities of the human clinician necessary for

therapy [see below]; thus, they could not constitute the whole of borderline therapy.) The first one concerns causal misattribution (attributing the wrong intentions to persons, beyond what their behaviour warrants) and prompting reflection on affective reaction:

Therapist: So tell me about the difficult situation when you met with your ex-boyfriend [prompting narrative of emotionally significant event]...it seems to me that it was a bit frustrating for you to meet with him, is that right to put it that way?

Gina: He met with me just to see that I was still in pain...

Therapist: What was it like for you to meet with him? [prompting reflection on affective reaction to events described in narrative]

Gina: Awful...I mean...I spent two hours in hell...he kept telling me about his life, and that he is going out and has a lot of new friends...so annoying...I hate him and he doesn't respect me, just want to bug me...

Therapist: That doesn't sound nice...did you feel anything in particular in that situation?

Gina: I [...] told you I HATE HIM...what is it you don't get? (talks very loud, and seems agitated)

Therapist: Wow...it seems to me that you are very much upset about what happened... sorry, it wasn't my intention to annoy you...

Gina: You all say you're sorry, but it is a lie (talks really loud and very fast)...John (ex-boyfriend), says he is sorry that we could not stay together...bullshit...he is not sorry about anything...Lisa ... says she wants to help me and tries to understand me all the time... she is not trying to understand anything or help anyone...

Therapist: Hold on, hold on for a second Gina, this goes really fast, and I can't quite figure it all out [implicitly pointing out faulty causal reasoning]...can we please pause for a second, and look at what happened here...

Gina: I don't want to pause anything, I know what you are up to, you want to blame me, tell me it is my own fault, that I have to work with myself...no way, you obviously don't want to help me, that is clear...Lisa doesn't like me, I know for a fact, and Carl (head of the institution) ignores me on purpose...<sup>27(p. 177.)</sup>

In the second dialogue, the therapist tries to help the patient identify their mixed emotional states (in the AI case, movement detection would be used):

Patient: Why do people say I am angry all the time? I am not.

Clinician: Well, I thought that you were as you are standing up, you raise your voice, and you are gesticulating with your arms... You come across as angry to me.... If you are not angry, can you say how you do feel

[suggesting identifying feeling state] so that I can understand it better?

Patient: Not angry.

Clinician: I accept that. [validation] So, can you describe what you do feel at the moment? Start by saying what your body feels like, if that helps...

Patient: I feel tense, which is why I am standing up. Actually, I am frustrated as I don't think people ever really listen to me.

Clinician: What would tell you that we were listening?

Patient: I don't really think that you all see me as important.

Clinician: That is really helpful. So when I see you as angry and agitated, I really need to consider if you are feeling unimportant, which I can easily miss. Is there something at the moment that we are doing that makes you feel unimportant?<sup>28(p. 254-5.)</sup>

#### 4 | PROS AND CAVEATS OF AI USE IN BPD THERAPY

What are the advantages are of digitally enhanced therapy assistance?

Two such potential advantages have already been mentioned: offering a therapeutic tool to those without sufficient access to therapy (eg, for financial reasons or because the therapeutic process is finished by the clinician, which patients sometimes experience as abandonment)<sup>29</sup> and eliminating certain potential interpersonal complications (while retaining some 'useful' complications in the human therapist-patient relationship).

A third potential merit of AI therapy assistants would be that more information revealed to non-human devices could be expected than to human therapists. Lucas and co-authors have found that, in clinical interviews, patients are willing to disclose information to 'virtual humans' perceived as supportive and safe.<sup>30</sup> Apparently, even the awareness of revealed information being later shared with psychiatrists does not keep patients from telling more than in a human-to-human interview. Borderline patients, who tend to experience problems with disclosing information about themselves to others, could also be anticipated to find it easier in such circumstances.

The absence of interpersonal complications presupposes that the patient does not perceive the AI device as a person or agent with its own intentions and agenda. Thus, the digital assistant should be regarded as a domain specific therapeutic tool<sup>6</sup> rather than assume a proper agentic role—which, apart from raising questions of autonomy, could in some cases undermine and jeopardize the success of the therapeutic process due to mistrust analogous to that of a human therapist.

Thus, the use of digital therapy assistants should be limited; particularly, they should not be thought of as replacing human therapists, for the following reasons. First, the expertise and oversight of a clinician is necessary for the design, implementation and assessment of the therapeutic procedure. While certain tasks, such as the development and strengthening of some of the patient's skills, can be partially delegated, the whole process has to be guided by a human therapist.



Again, the therapeutic relationship can be the basis for social re-learning. In borderline therapy, trust between patient and therapist is an important factor. In a clinical survey, patients reported the conviction that trust is essential for a fruitful therapeutic alliance.<sup>32</sup> A trusting relationship can also help them develop the self-trust the process of recovery can build on.

Apart from basic trust, other features of the therapeutic relationship can also promote social re-learning. Ruptures in the therapeutic allegiance may be turned to advantage by leading to insights, improving communication and even strengthening the therapeutic bond.<sup>7</sup>

Therapists' specifically human capacities may also contribute to the success of the therapeutic effort. One such capacity is empathy. In mentalization-based therapy, for instance, empathic validation 'seeks to engender in the patient a sense that the clinician has understood his/her internal state, that the clinician really "gets" the patient and the issue he/ she is talking about'.<sup>28 (p. ix)</sup>

Therapy may be advanced by the therapist's affective capacities in other ways as well. In positive countertransference, in which the therapist projects their own feelings onto the patient, the therapist's reactions can help understand the patients' mental processes. The therapist's actual emotions, rather apparent expressions of emotions, that play a role, thus are highly unlikely to be reproduced by AI in the near future.

A last point is that in AI-based conversations, the subject has more control in terms of starting and stopping. While the patient's assuming a more active role in her therapy may be a positive development, being able to end conversations when they come to sensitive points may also make it harder to change thought habits.

Thus, AI-assisted therapy cannot replace traditional forms of therapy; rather, it can only support the therapist's efforts and help consolidate the results. Human-to-human and AI-enhanced therapy could in several respects complement each other.

## 5 | CONCLUSION

As we have seen, digitalization of BPD psychotherapy holds out the promise of constituting an add-on to human-to-human therapy in a way that would combine the advantages of trust and useful ruptures in the therapeutic relationship, the competence and affective capacities of the human therapist, and some distinctive positive features of AI assistance. Other than providing a chance of reducing those interpersonal problems which are not conducive to the success of the therapy, these features include availability to patients without sufficient access to therapy and an opportunity for more openness.

## ACKNOWLEDGEMENTS

I would like to thank András Simonyi for help with thinking about technological solutions concerning a BPD therapy assistant, as well as Tim Thorton and Csaba Pléh for their comments on an earlier version of some of the ideas presented here.

This research was financed by project no. 120375 of the National Research, Development and Innovation Office of Hungary and the excellence project *Problem-Solving Systems* of ELTE, Faculty of Humanities.

## CONFLICT OF INTEREST

The author declares no conflict of interest.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

## ETHICS STATEMENT

No ethical approval is required for this study.

## ORCID

Judit Szalai  <https://orcid.org/0000-0001-7861-4908>

## ENDNOTES

<sup>1</sup> According to the WHO Mental Health Atlas 2017, the global median of psychiatrists per 100 000 is 1.<sup>4 (p.34)</sup>

<sup>2</sup> Narratives can be relevant to the understanding of different psychopathologies at least in two ways. The first is that characteristic narratives can reveal the presence of different psychopathologies. Persons' conflicted relationship to others, extremely low self-esteem, perception of themselves as worthless or rejected will appear in their self-narratives. Lack of coherence in one's self-narrative can also be an indicator of psychopathology. The second possibility is that self-narratives are a causal factor in, or even partly constitute, mental pathologies. BPD could also be understood in this second way: an integral part of BPD is identity disturbance, and the identity in point is—at least in part—narrative. I take it that narratives are relevant to the understanding and explanation of BPD in both ways: narrative disturbance is both constitutive and symptomatic of BPD.

<sup>3</sup> In Section 3 of DSM-5, the Alternative Model for Personality Disorders characterizes PDs in terms of levels personality functioning or organization and pathological personality traits. The elements of personality functioning—at least moderate-level impairment of which is required for the diagnosis of a PD—are self-functioning and interpersonal functioning. The former is focused on identity (stable self-esteem, accurate self-assessment, clear boundaries, etc.) and self-direction (pursuit of meaningful and consistent goals, etc.).<sup>14 (p.762)</sup> The 11th edition of the *International Classification of Diseases* (ICD-11) also bases PD diagnoses on functional impairment in aspects of self (identity, self-worth, accuracy of self-assessment, self-direction) and interpersonal functioning.<sup>15</sup>

<sup>4</sup> The most well-known self-help chatbots for mental conditions are Woebot and Wysa, which offer mental health support to persons with anxiety and depression.

<sup>5</sup> There is no reason to present the AI therapy assistant device in the form of a humanoid robot. In fact, doing so might be distractive and potentially trigger interaction problems on the part of the patient.

<sup>6</sup> Raising awareness about this is in line with the European Commission's Higher-Level Expert Group Guidelines: 'Human beings should always know if they are directly interacting with another human being or a machine, and it is the responsibility of AI practitioners that this is reliably achieved. AI practitioners should therefore ensure that humans are made aware of—or able to request and validate the fact that—they interact with an AI system (for instance, by issuing clear and transparent disclaimers). It should be borne in mind that the confusion between humans and machines could have multiple consequences such as attachment, influence, or reduction of the value of being human.<sup>31(p. 34)</sup>

<sup>7</sup> In other cases, the rupture may disrupt the process. When this happens, digitally enhanced therapy may serve to save commitment to therapy on the part of patients who are withdrawing from human-to-human therapeutic communication.

## REFERENCES

1. Martino F, Menchetti M, Pozzi E, Berardi D. Predictors of dropout among personality disorders in a specialist outpatients psychosocial treatment: a preliminary study. *Psychiatry Clin Neurosci*. 2012;66(3):180-186. <https://doi.org/10.1111/j.1440-1819.2012.02329.x>.
2. Sansone RA, Farukhi S, Wiederman MW. Utilization of primary care physicians in borderline personality. *Gen Hosp Psychiatry*. 2011;33:343-346. <https://doi.org/10.1016/j.genhosppsych.2011.04.006>.
3. Soeteman DI, Hakkaart-van Roijen L, Verheul R, Busschbach JJ. The economic burden of personality disorders in mental health care. *J Clin Psychiatry*. 2008;69(2):259-265.
4. World Health Organization. *Mental Health Atlas 2017*. Geneva, Switzerland: World Health Organization; 2018 [http://www.who.int/mental\\_health/evidence/atlas/mental\\_health\\_atlas\\_2017/en/](http://www.who.int/mental_health/evidence/atlas/mental_health_atlas_2017/en/).
5. Linehan M. Dialectical behavior therapy for borderline personality disorder. In: Barlow DH, ed. *Clinical Handbook of Personality Disorders*. New York, NY; London, England: The Guilford Press; 2008:365-420.
6. Bateman AW, Fonagy P. Mentalization-based treatment for BPD. In: Hoffmann PD, Steiner-Grossmann P, eds. *Borderline Personality Disorder: Meeting the Challenges to Successful Treatment*. Binghampton, NY: The Haworth Press; 2008:187-202.
7. Tadros E, Cappetto M, Kaur L. Treating symptoms of borderline personality disorder through narrative therapy and naltrexone. *Am J Fam Ther*. 2019;47(2):87-101. <https://doi.org/10.1080/01926187.2019.1590787>.
8. Linehan MM. *Cognitive-Behavioral Treatment of Borderline Personality Disorder*. New York, NY: Guilford Press; 1993.
9. Zanarini MC. Reasons for change in borderline personality disorder (and other axis II disorders). *Psychiatr Clin N Am*. 2008;31(3):505-515. <https://doi.org/10.1016/j.psc.2008.03.006>.
10. Newman CF. Cognitive therapy for borderline personality disorder. In: VandeCreek L, Knapp S, Jackson T, eds. *Innovations in Clinical Practice: A Sourcebook*, Vol. 116. Sarasota, FL: Professional Resource Press; 1998:17-38.
11. Wenzel A, Chapman JE, Newman CF, Beck AT, Brown GK. Hypothesized mechanisms of change in cognitive therapy for borderline personality disorder. *J Clin Psychol*. 2006;62(4):503-516. <https://doi.org/10.1002/jclp.20244>.
12. Butler AC, Brown GK, Beck AT, Grisham JR. Assessment of dysfunctional beliefs in borderline personality disorder. *Behav Res Ther*. 2002;40(10):1231-1240. [https://doi.org/10.1016/s0005-7967\(02\)00031-1](https://doi.org/10.1016/s0005-7967(02)00031-1).
13. Winsper C. The aetiology of borderline personality disorder (BPD): contemporary theories and putative mechanisms. *Curr Opin Psychol*. 2018;21:105-110. <https://doi.org/10.1016/j.copsyc.2017.10.005>.
14. American Psychiatric Association (APA). *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC: Author; 2013 (DSM-5).
15. World Health Organization. *International Classification of Diseases for Mortality and Morbidity Statistics (11th Revision)*. Geneva: World Health Organization; 2018. <https://icd.who.int/browse11/l-m/en>
16. Semerari A, Dimaggio G, Carcione A, Conti L, Nicolo G. General treatment principles. In: Dimaggio G, Semerari A, Carcione A, Nicolo G, Procacci M, eds. *Psychotherapy of Personality Disorders: Metacognition, States of Mind, and Interpersonal Cycles*. London, England; New York, NY: Routledge; 2007:29-46.
17. Korn CW, la Rosee L, Heekeren HR, Roepke S. Social feedback processing in borderline personality disorder. *Psychol Med*. 2015;46(3):1-13. <https://doi.org/10.1017/S003329171500207X>.
18. Gregory RJ. *Remediation for Treatment-Resistant Borderline Personality Disorder: Manual of Dynamic Deconstructive Psychotherapy*, 2016. (manuscript) [https://www.upstate.edu/psych/pdf/education/psychotherapy/ddp\\_manual.pdf](https://www.upstate.edu/psych/pdf/education/psychotherapy/ddp_manual.pdf). Accessed May 13, 2020.
19. Fonagy P, Luyten P, Allison E. Epistemic petrification and the restoration of epistemic trust: a new conceptualization of borderline personality disorder and its psychosocial treatment. *J Pers Disord*. 2015;29(5):575-609. <https://doi.org/10.1521/pedi.2015.29.5.575>.
20. Bateman AW, Fonagy P. *Psychotherapy for Borderline Personality Disorder: Mentalization-Based Treatment*. Oxford, England; New York, NY: Oxford University Press; 2004.
21. DiMaggio G, Montano A, Popolo R, Salvatore G. *Metacognitive Interpersonal Therapy for Personality Disorders*. London, England: Routledge; 2015.
22. Lazarus SA, Cheavens JS, Festa F, Rosenthal MZ. Interpersonal functioning in borderline personality disorder: a systematic review of behavioral and laboratory-based assessments. *Clin Psychol Rev*. 2014;34:193-205. <https://doi.org/10.1016/j.cpr.2014.01.007>.
23. Prunetti E, Framba R, Barone L, Fiore D, Sera F, Liotti G. Attachment disorganization and borderline patients' metacognitive responses to therapists' expressed understanding of their states of mind: a pilot study. *Psychother Res*. 2008;18(1):28-36. <https://doi.org/10.1080/10503300701320645>.
24. Doherty G, Coyle D, Matthews M. Design and evaluation guidelines for mental health technologies. *Interact Comput*. 2010;22:243-252.
25. Demszky D et al. GoEmotions: a dataset of fine-grained emotions. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online. Association for Computational Linguistics; 2020:4040-4054.
26. Oakley T, Pascual E. Blending theory and its application in semantics and discourse studies. In: Dancygier B, ed. *Cambridge Handbook of Cognitive Linguistics*. Cambridge, England: Cambridge University Press; 2017.
27. Bo S, Sharp C, Fonagy P, Kongerslev M. Hypermentalizing, attachment, and epistemic trust in adolescent BPD: Clinical illustrations. *Personal Disord. Personality Disorders: Theory Res. Treat*. 2017;8(2):172-182. <https://doi.org/10.1037/per0000161>.
28. Bateman A, Fonagy P. *Mentalization-based treatment for personality disorders*. Oxford: Oxford University Press, 2016.
29. Juul S, Simonsen S, Bateman A. The capacity to end: termination of mentalization-based therapy for borderline personality disorder. *J Contemp Psychother*. 2020. Advance online publication. <https://doi.org/10.1007/s10879-020-09456-6>
30. Lucas GM, Gratch J, King A, Morency LP. It's only a computer: virtual humans increase willingness to disclose. *Comput Hum. Behav*. 2014;37:94-100. <https://doi.org/10.1016/j.chb.2014.04.043>.
31. [European Commission Higher-Level Group, HLEG ]. Ethics Guidelines for Trustworthy AI, 2019. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>. Accessed 30 April 2020.
32. Langley GC, Klopper HJ. Trust as a foundation for the therapeutic intervention for patients with borderline personality disorder. *Psychiatr Ment Health Nurs*. 2005;12:23-32.

**How to cite this article:** Szalai J. The potential use of artificial intelligence in the therapy of borderline personality disorder. *J Eval Clin Pract*. 2021;27:491-496. <https://doi.org/10.1111/jep.13530>