Artificial intelligence and psychotherapy: A counterpoint

Derek Richards1,2

1 Amwell Science, Dublin, Ireland
2 School of Psychology, Trinity College Dublin, Dublin, Ireland

Correspondence
Derek Richards, Amwell Science, One Stephen Street Upper, Dublin D08 DR9P, Ireland.
Email: derek.richards@amwell.com

Abstract
Psychotherapy practice is a human endeavour. Research on the specific and non-specific factors of treatment has helped crystallise its relevance and clinical impact. The challenges currently faced by the field revolve around ensuring access to evidence-based treatments and enhancing their effectiveness. Digitally delivered formats of empirically supported treatments increase access while supporting the relevance of the treatment-specific ingredients and the necessity for human guidance. Excitement surrounds the potential integration of novel artificial intelligence (AI) machine learning methods to advance psychotherapy effectiveness. The abundance of data in digitally delivered formats positions them well to harness the capabilities of AI. Recent work provides proof of concept in areas including detection and diagnosis, predicting outcomes, treatment adherence, remission and relapse. A potential risk emerges when applying machine learning methods, in which an overreliance on AI inferences may overshadow the human aspect of psychotherapy. The contrast is simple: we may over-invest in the rationality and relevance of our AI inferences, blindly obeying the algorithmic counsel that may lead to unintended consequences, such as oversimplifying human complexity. This would amount to changing psychotherapy from a human-centric to a techno-centric endeavour, something we should steadily avoid. This perspective highlights the importance of balancing enthusiasm for AI advancements with a cautious approach. The discussion outlines the risks associated with overdependence on AI and provides reasons to avoid a scenario in which psychotherapy loses its human essence. In conclusion, the perspective suggests avenues for future research to prevent such a transformation and maintain the human-centric nature of psychotherapy.

KEYWORDS
client group, decision making, practice topics, psychotherapy, psychotherapy and counselling approaches, research methods

1 INTRODUCTION
Empirically supported psychotherapy treatments have demonstrated efficacy and are recommended in clinical guidelines. Consequently, they are heavily used in mental health care, are often a preferred treatment choice for patients and are generally recommended as first-line treatment options. An abundance of research and clinical trials has been synthesised to demonstrate the effectiveness of...
psychotherapy treatments for a broad range of mental and behav-
ioral health disorders across diverse settings (primary care, com-
munity and specialist services) and populations (children and young
people, adults and older adults; Nathan & Gorman, 2015). In addi-
tion, research in psychotherapy has illustrated the relative contribu-
tions of specific and non-specific factors to treatment outcomes
(Norcross & Lambert, 2019). It is clear that the technical aspects of
the treatments, for instance, core treatment techniques and strate-
gies, alongside dimensional factors, such as the therapeutic rela-
tionship, therapist competence and adherence to treatment protocols,
contribute to the clinical effectiveness of the treatments. Therefore,
psychotherapy has always been a human endeavour involving the
interplay of treatment-specific ingredients delivered as part of a
professional relationship to produce a clinical change in clients’
symptoms and functioning. While there is general agreement that
empirically supported treatments work, a caveat is that it is still an
empirical question regarding the relative contribution of specific
and non-specific factors across theoretically divergent treatments
(Huibers & Cuijpers, 2015).

The field of psychotherapy faces two important challenges. First,
there is a growing prevalence of mental health conditions, putting
strain on service delivery because of a shortage of trained profes-
sionals. This scarcity makes it challenging for individuals to access
evidence-based treatment options, exacerbated by the recent pan-
demic (Santomauro et al., 2021). Second, only some benefit from
psychotherapy treatment; about 50% respond to treatment and
about 30% move to remission of symptoms. Despite significant
advancements, diagnostic clarity and prognosis prediction remain
elusive. Consequently, establishing reliable clinical outcomes for
individuals necessitates ongoing psychotherapeutic and pharma-
ceutical interactions to comprehensively understand the patient’s
long-term diagnosis, prognosis and the most effective treatment
regimen. These challenges underscore the importance of addressing
access to treatment and improving treatment effectiveness.

2 | DIGITAL PSYCHOTHERAPY

Digital psychotherapy is based on using available technologies to
deliver services. They include teletherapy and online support groups
that follow traditional therapeutic delivery models and new modes,
such as guided digital interventions and self-guided apps. A primary
driving force for developing digitally delivered psychotherapy inter-
ventions has been the need to extend access to a greater number of
people in an attempt to offset the growing burden of mental health,
the demands on treatment services and the shortage of trained pro-
fessionals. Other benefits include offering greater access choices,
reducing geographical and ability barriers and potentially reaching
people who would not usually access traditional psychotherapy
services.

The meta-synthesis of research trials demonstrates the
effectiveness of delivering evidence-based treatment proto-
cols via various technologies for mental and behavioural health
disorders across diverse settings (primary care, community,
specialist services) and populations (children and young people,
adults and older adults; Andrews et al., 2018; Eilert et al., 2021,
2022; Lee et al., 2023; Moshe et al., 2021; Palacios, Adegoke,
et al., 2022; Palacios, Enrique, et al., 2022; Riadi et al., 2022;
Wright et al., 2019). In particular, the research on novel modes of
treatment delivery (guided digital interventions) has grown to the
point in which several supported digitally enabled interventions
for treating common mental health disorders, including depression
and anxiety disorders, are recommended in clinical guidelines (e.g.,
National Institute for Health and Care Excellence, 2023a, 2023b).
Examine specific and non-specific factors reveals their ongoing
significance in digitally delivered formats. Research highlights the
importance of human relationships in achieving treatment suc-
cess. However, the exact role of human relationships may vary
from traditional psychotherapy, posing an important empirical
question yet to be fully understood. Expert consensus has rec-
ommended that digital interventions be made widely available to
people with common mental health disorders, that they should be
integrated into mental healthcare services and that evidence stan-
dards frameworks should be developed to support their adoption
(Mohr et al., 2021).

Over the last 10–15 years, several countries have established
digital mental health clinics that offer virtual treatments to patients

<table>
<thead>
<tr>
<th>Implications for practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>• As we advance the intersection of artificial intelligence (AI) and psychotherapy research, we must acknowledge and address potential risks conscientiously through evidence-based practices.</td>
</tr>
<tr>
<td>• Researchers in AI and psychotherapy should be aware of emerging legal and regulatory frameworks regarding algorithms.</td>
</tr>
<tr>
<td>• Researchers should explore psychotherapeutic databases with limited data using theory-informed perspectives to generate meaningful research questions and hypotheses.</td>
</tr>
<tr>
<td>• Investigating the use of privacy-enhancing technologies and the potential of synthetic data sets to address current limitations in research may prove valuable.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Implications for policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Implement policy recommendations addressing the risks associated with applying machine learning methods, in which an overreliance on AI inferences may overshadow the human aspect of psychotherapy.</td>
</tr>
<tr>
<td>• An agenda for continuous training and development for professionals in the field is advised to ensure a balanced integration of technology and human expertise.</td>
</tr>
</tbody>
</table>
(Titov et al., 2018). One example is the UK ‘Talking Therapies’, which delivers various interventions (e.g., guided digital interventions, bibliotherapy, group cognitive behavioural therapy (CBT), and online and in-person face-to-face counselling). The evidence is clear that adding technology-based delivery modes advances the access agenda. For instance, in 2021/2022, 1.24 million people accessed the services and 648,617 digital therapy sessions were delivered (NHS England, 2023).

3 | THE PROMISE OF AI

Digital treatments have brought something unprecedented to psychotherapy: multi-dimensional data sets. Such data sets promise to propel us beyond the conventional inferential paradigm prevalent in psychotherapeutic research, characterised by techniques such as p-values and effect sizes. The approach has faced criticism for its limited replicability and where reported results may not be clinically meaningful (Dwyer et al., 2018; Ioannidis, 2016). The promise is that utilising multi-dimensional data (including biological, demographic, clinical and other data types) alongside novel developments of artificial intelligence (AI) and machine learning methods could yield significant insight into decisions about diagnosing, prognosis and treating people with mental illness (Dwyer et al., 2018), thereby enhancing treatment effectiveness. Clinically, this could be very significant because patients could be assessed more accurately, stratified and assigned their most appropriate treatment or combined treatment (Dwyer et al., 2018). In a world with waiting lists and staff shortages, maximising therapist time and treatment efficiencies to achieve greater numbers with positive clinical outcomes would be welcome. Moving the needle even in small percentage points could yield significant individual, economic and societal gains.

Studies including machine learning models have analysed session transcripts and used pre-, post- and sometimes continuous patient-reported outcomes, including baseline patient characteristics, neuroimaging, biomarkers and sensor data (Aafjes-van Doorn et al., 2021). Several of these studies used machine learning to identify patients at risk of relapse (Lorimer et al., 2021) or to predict symptom improvement and remission after treatment (Angstman et al., 2017; Chekroud et al., 2016; Hilbert et al., 2020, 2021; Pearson et al., 2019; Prasad et al., 2023), to forecast failure to respond to treatment (Kautzky et al., 2018; Perlis, 2013) or to predict unfavourable course of disease progression in life (Kessler et al., 2016; Maarsingh et al., 2018). In this nascent field of machine learning, AI methods and psychotherapy, these studies and their corresponding reviews demonstrate a clear proof of concept for further work and investigation (Aafjes-van Doorn et al., 2021; Chekroud et al., 2021; Dwyer et al., 2018; Shatte et al., 2019). At the same time, researchers are beginning to become familiar with the prospects of the potential limitations of these approaches and the illusions they may serve (Chekroud et al., 2024).

4 | AI AND PSYCHOTHERAPY: A COUNTERPOINT

In recent decades, there has been a growing trend of incorporating data collection across various aspects of life, facilitating the development and application of new AI methods. This trend is evident across domains in which multi-dimensional data sets are prevalent, and AI methods are commonly used. Researchers from various disciplines increasingly highlight a pervasive assumption in relying on these methods. There is a risk of overemphasising the reliability of algorithmic inferences when unaccounting for recommendations. The study by Chekroud et al. (2024) illustrates the limits of the generalisability of an algorithm to predict outcomes from schizophrenia treatment, debunking the implicit assumption of algorithmic rationality. The literature is replete with examples of algorithms that harm vulnerable and marginalised groups, even when they ‘work’ (Broussard, 2018; Eubanks, 2019; O’Neil, 2016). For instance, the case of the sentencing algorithm that discriminated against Black people and resulted in historical racial injustices being propagated. The system predicted that Black people posed a higher risk of recidivism (and the reverse for white people), leading to higher and longer sentences (Eubanks, 2019). In addition, our current ‘multi-dimensional’ databases and algorithms are rife with biases and heuristics that rely on arbitrary classifications, messy data and many other concealed uncertainties (Hong, 2020). The like the criminal sentencing algorithm example, it relied on biased data in the first instance: historically discriminatory court records.

Psychotherapy data, too, are biased; for example, CBT has been predominately developed and used with White and well-educated heterosexuals. Any algorithm based on such historical data risks ignoring large sways of people, including neuro-diverse, racial and ethnic minorities, culturally diverse, LGBTQ+, and diverse socio-economic groups, to name a few. Furthermore, history, background, lived experiences and context matter in psychotherapy. The data sources for the study on Schizophrenia were chosen for their ‘comparability and consistency’, crucial dimensions in developing inferences, but not necessarily indicative of completeness. Consequently, many environmental factors important to mental health treatment outcomes were not accounted for. This is a serious limitation for all current data sources in psychotherapy research and AI, which highlights another important consideration: Data show only a limited perspective of the real world (Wong, 2023). The application of machine learning automates homogeneity and consequently relegates the humans whose lives the data represent to the background (Crawford, 2021).

5 | WHY WE SHOULD AVOID THIS

In the age of technology, a prevalent perspective is that every problem is a technology problem that only technology can solve.
under the surface, is rife with injustice and inequity, potentially flat -
if we fail in this task, we will build AI that is superficially smart but, 
clinical measures involved choices between, for example, a 50 per 
lished outcome measures (e.g., PHQ-9, GAD-7; Prasad et al., 
structs are open to interpretation. To give a simple example, recently, 
ing models is crucial to prevent unintended harm or exacerbation 
ues, interpersonal dynamics and cultural contexts, among other 
algorithms excel at analysing large data sets and creating infer -
Human experiences, emotions and psychological states are inher -
the role of the human aspect of psychotherapeutic interventions. 
ences, emotions and psychological states are inher -
reducing or using reliable change indices on estab -
reduction or using reliable change indices on estab -
be blindly followed.

Overcoming challenges in psychotherapy, such as accessing 
transform diagnosis, prognosis and treatment outcomes. However, 
multi-dimensional databases introduce biases and arbitrary classifi -
potential risk emerges when applying machine learning methods, in which 
overreliance on AI inferences may overshadow the human aspect 

takes me to changing psychotherapy from a human-centred to a techno-centred endeav -
thing on definitions, understanding the 
clinical measures used and temporal considerations for achieving ac -
overreliance on AI inferences may overshadow the human aspect 
overreliance on AI inferences may overshadow the human aspect 

6 | RESEARCH IN AI & PSYCHOTHERAPY

As we advance the intersection of AI and psychotherapy research, we 
must acknowledge and address potential risks conscientiously through 
evidence-based practices. The principle of machine learning models’ 
explainability (or interpretability) is a primary consideration. This in -
volves sensibly explaining the model’s workings and outputs to another 
human, including any inherent biases. In our example above, when 
predicting clinical outcomes, clarity on definitions, understanding the 
clinical measures used and temporal considerations for achieving ac -
overreliance on AI inferences may overshadow the human aspect 
overreliance on AI inferences may overshadow the human aspect 

7 | CONCLUSION

Overcoming challenges in psychotherapy, such as accessing 
evidence-based treatments and enhancing their effectiveness, is 
crucial. Digital formats have proven effective, offering increased ac -
cessibility and leveraging the potential of AI. Recent advancements 
in AI and psychotherapy demonstrate a proof of concept that could 
form transform diagnosis, prognosis and treatment outcomes. However, 
multi-dimensional databases introduce biases and arbitrary classifi -
potential risk emerges when applying machine learning methods, in which 
am an overreliance on AI inferences may overshadow the human aspect 

blindly following algorithmic counsel may lead to 
ungrounded investment (Kahneman, 

The overreliance on inferences could inadvertently diminish 
the role of the human aspect of psychotherapeutic interventions. 

Psychotherapy research introduces complexity, in which con -
structs are open to interpretation. To give a simple example, recently, 
I contributed to developing an algorithm predicting psychotherapy outcomes (Prasad et al., 2023). Deciding parameters for meaningful 
clinical outcomes involved choices between, for example, a 50 per 
Decisions were made, and the algorithm ‘works’, but its discrimi -
natory effectiveness remains uncertain due to trade-offs between 
sensitivity and specificity, embedding false positives and negatives. 
Focusing on a specific, well-defined issue (reliable improvement) 
with limited dimensionality (PHQ-9 and GAD-7 scores over three 
measures) makes the algorithm more straightforward to explain. In 
practical deployment, the risk emerges that its determinations may 
be blindly followed.

RICHARDS
acknowledges the limitations of algorithmic decision-making. A collaborative and cautious integration of AI into psychotherapy, guided by ethical considerations and a deep understanding of the complexities involved, is essential to harness the benefits of technology while prioritising the well-being of individuals seeking mental health treatments.

Additionally, the paper offers some considerations for researchers in AI and psychotherapy, emphasising awareness of emerging legal frameworks. It encourages exploring limited data with theory-informed perspectives to generate meaningful research questions and hypotheses. Data aggregation and the use of privacy-enhancing technologies are recommended, as well as recognising the potential of synthetic data sets to address current limitations. Deploying AI in psychotherapy requires using state-of-the-art technologies to prevent overreliance on algorithmic counsel and minimise errors. Finally, continuous training and development for professionals in the field are advised to ensure a balanced integration of technology and human expertise.

ACKNOWLEDGEMENTS

I thank Daniel Duffy for his helpful feedback on a previous draft. Open access funding provided by IReL.

CONFLICT OF INTEREST STATEMENT

Derek Richards is an employee of Amwell and holds shares in the company. He was previously the Chief Science Officer at SilverCloud Health, a subsidiary of Amwell, from August 2021.

REFERENCES


Ioannidis, J. P. A. (2016). Why most clinical research is not useful. PLoS Medicine, 13(6), e1002049. https://doi.org/10.1371/journal.pmed.1002049


Wong, W. H. (2023). We, the data: Human rights in the digital age. MIT Press.


**AUTHOR BIOGRAPHY**

Derek Richards leads the strategic research for SilverCloud by Amwell. Since 2005, he has been extensively involved in clinical research, development and implementation of technology-delivered interventions for mental health. Over the last 15 years, Derek has built a team of world-class scientists whose published work is cited widely and whose impact has spearheaded continued innovation in digital mental health care. Since SilverCloud’s acquisition by Amwell in 2021, Derek has led the broader research initiatives to deliver impact, value and outcomes research across Amwell’s diverse telemedicine product solutions.

---

**How to cite this article:** Richards, D. (2024). Artificial intelligence and psychotherapy: A counterpoint. Counselling and Psychotherapy Research, 00, 1–6. https://doi.org/10.1002/capr.12758