Administration and Policy in Mental Health and Mental Health Services Research
https://doi.org/10.1007/s10488-023-01309-3

ORIGINAL ARTICLE

Leveraging Novel Technologies and Artificial Intelligence to Advance Practice-Oriented Research

Dana Atzil-Slonim[1] · **Juan Martin Gomez Penedo**[2] · **Wolfgang Lutz**[3]

Accepted: 29 September 2023

Abstract

Mental health services are experiencing notable transformations as innovative technologies and artificial intelligence (AI) are increasingly utilized in a growing number of studies and services.

These cutting-edge technologies carry the promise of substantial improvements in the field of mental health. Nevertheless, questions emerge about the alignment of novel technologies and AI systems with human needs, especially in the context of vulnerable populations receiving mental healthcare. The practice-oriented research (POR) model is pivotal in seamlessly integrating these emerging technologies into clinical research and practice. It underscores the importance of tight collaboration between clinicians and researchers, all driven by the central goal of ensuring and elevating client well-being. This paper focuses on how novel technologies can enhance the POR model and highlights its pivotal role in integrating these technologies into clinical research and practice. We discuss two key phases: pre-treatment, and during treatment. For each phase, we describe the challenges, present the major technological innovations, describe recent studies exemplifying technology use, and suggest future directions. Ethical concerns and the importance of aligning humans and technology are also considered, in addition to implications for practice and training.

Keywords Artificial intelligence · Practice-oriented research · Machine learning · Psychotherapy research

Imagine the following scenario, which is likely to occur in the near future: *An individual seeks mental health help and enrolls in a program that incorporates an artificial intelligence (AI) system to augment the care delivered by mental health professionals. This AI system was developed and is continually supervised in close partnership with mental health practitioners and researchers to ensure ethical and responsible use for the benefit of the individual and to enhance the standard of care. Drawing from a wide range of pre-treatment data and identifying the client's closest cluster of similar individuals, the system suggests a therapist and treatment plan that best suits the individual's unique needs. Initially, it recommends self-help guidelines and a series of online therapy sessions. If the client's progress plateaus or significant events occur, there is the option to transition to in-person sessions. By monitoring the client's progress, the system provides practical recommendations and gives both the client and the clinician the autonomy to choose which ones to follow. Simultaneously, the system learns from their feedback and adjusts its recommendations accordingly, thus refining its insights over time. Real-time understandings about the therapeutic interaction are derived from analyzing a variety of data collected during each session. The AI uses this information to suggest adaptive interventions tailored to the client's specific needs at that moment. During supervision, the therapist and supervisor assess session recordings and AI feedback by focusing on repetitive patterns and content, and on significant events that occurred during the session. They work together to enhance the client's coping strategies in future sessions, which in turn contribute to the therapist's skills and the treatment's effectiveness. Once the client achieves greater well-being, therapy concludes, but the client has the option to extend the monitoring of their*

✉ Dana Atzil-Slonim
dana.slonim@gmail.com

[1] Department of Psychology, Bar-Ilan University, Ramat-Gan, Israel

[2] Department of Psychology, University of Buenos Aires, Buenos Aires, Argentina

[3] Department of Psychology, University of Trier, Trier, Germany

Published online: 25 October 2023

progress for a while longer, to ensure an ongoing positive trajectory.

In recent years the ways in which mental health services are delivered and investigated have undergone radical transformations, with more and more services and studies utilizing innovative technologies and AI. These transformations provide unprecedented opportunities to advance practice-oriented research (POR). The main goals of POR are to understand whether, how, for whom, and by whom, mental health services in clinical routine can be effective, and to define strategies to improve practice. Over the past few decades, POR has produced a significant amount of clinically relevant information on the effectiveness and delivery of mental health services for various populations, and has provided a wide range of insights into ways to improve practice (e.g., Castonguay et al., 2021; Lutz et al., 2021). Because the POR model is based on the synergy between research and practice, recent technological advances in these two domains can enrich each other and benefit the mental health field as a whole. Nevertheless, it is imperative to approach these advances cautiously to ensure responsible implementation to maximize the advantages for clients.

Major advances in psychotherapy research have been facilitated by the rise of machine learning (ML), a subset of AI that involves training computer systems to automatically improve their performance on a specific task by learning from data. ML are capable of different types of learning, which are usually categorized into supervised, unsupervised, and reinforcement learning (Alpaydin, 2020; Delgadillo & Atzil-Slonim, 2022). Supervised learning utilizes datasets labeled by humans to make predictions about a certain outcome (e.g., each therapist's sentence in a transcribed dataset is labeled by humans as containing a specific intervention type and these data are used to make prediction about clients' outcome). Unsupervised learning does not require any pre-labeled datasets; rather, the algorithm detects patterns in the data, clusters them in terms of their distinguishing characteristics and examines whether they are predictive of a certain outcome (e.g., the algorithm might identify certain recurring themes or topics across sessions that are predictive of positive change). Reinforcement learning is a dynamic learning process where the model improves its decision-making abilities by taking actions and receiving feedback in the form of rewards on these actions (Géron, 2022) (e.g., an AI system provides written feedback for therapists at the end of each session about their interventions during the session. The therapists provide feedback by rating the AI's feedback as helpful or non-helpful. Over time, the system improves its suggestions based on this feedback). These approaches are particularly well-suited for processing complex mental health data and allow computer systems to learn and refine their predictions based on experience with data from other clients, which then serves to maximize prediction accuracy about new clients. There are many ML modeling techniques (for a review, see Aafjes-van Doorn et al., 2021; Delgadillo & Atzil-Slonim, 2022). In recent years, deep learning models, specifically transformer-based language models, have emerged as the dominant ML method (Devlin et al., 2018). These models are pre-trained on huge datasets of unlabeled text by randomly masking some of the words and training the model to predict them (unsupervised learning). This allows the model to learn the underlying structure of the language and the context in which words appear. After pre-training, the model can be fine-tuned for a specific task on a smaller labeled dataset (such as emotion recognition in natural psychotherapy text) by updating its parameters to optimize a task-specific objective (supervised learning). The versatility and capability of transformer-based language models have resulted in their widespread application across diverse research fields, including mental health (Delgadillo & Atzil-Slonim, 2022). Recent developments in computing power and deep learning techniques have led to significant advances in generative AI (a type of AI that is capable of creating new content) and large language models (LLMs), such as ChatGPT and GPT-3/4 (Bommasani et al., 2021). LLMs excel at comprehending and generating human language, grasping intricate patterns, and contextual nuances. They can be fine-tuned for specific purposes (with techniques such as reinforcement learning), that enhance their performance and accuracy by optimizing task-specific parameters (Stade et al., 2023).

Progress in technology has also led to vast transformations in the ways mental health services can be delivered. Internet-based tools, mobile applications, and AI chatbots are increasingly being utilized for mental health assessments and psychotherapeutic interventions. They are available in various formats including video, audio, text, and gaming (Hermes et al., 2019). While some are designed for direct client use, others are facilitated by mental health professionals (Stade et al.,2023). The recent advancements in LLMs and generative AI provide a wide range of abilities that may augment and support mental health services, including generating intervention suggestions and feedback for clients, therapists, and supervisors (Stade et al., 2023). These cutting-edge resources have contributed to the broader dissemination of mental healthcare, by enabling greater access for those in need and hold great potential to advance the effectiveness of treatments. However, as AI systems continue to advance in their power and capabilities, there are concerns that they may pursue objectives that do not fully align with human needs, which can potentially result in unintended consequences (Bommasani et al., 2021). The AI alignment problem is an active area of research in AI safety (Gabriel, 2020), but is particularly critical when it comes to mental

healthcare, where AI systems may interact with vulnerable populations, and where a lack of alignment between AI and humans could have extremely deleterious consequences.

In addition, despite significant progress in recent years, AI systems are still far from being capable of autonomously treating humans (Stade et al., 2023). While certain tasks can be performed well by AI systems (e.g., providing automated feedback to therapists; Flemotomos et al., 2021), other capabilities require further development before they can be effectively implemented in real-world settings (e.g., providing real-time intervention suggestions to therapists). Certain abilities inherently rely on human qualities and may not be replicable by AI systems in the foreseeable future (e.g., being attuned to clients' implicit needs at any given moment). For these reasons, the POR model plays a critical role in integrating these technologies into clinical practice. The POR stresses the importance of investigating the implementation of novel methods in routine practice while actively involving clinicians in their design, development, and monitoring. By acknowledging the criticality of a sense of shared ownership between clinicians and researchers, the POR model allows for a deep alignment of these technologies with the everyday nuances of clinical environments. Given the communication barriers that often arise in interdisciplinary collaborations, the POR model incorporates strategies to enhance productive collaboration among clinicians, psychotherapy researchers and AI researchers. By acknowledging the perspectives and needs of each stakeholder and encouraging ongoing feedback, these strategies not only enhance the adaptability and evolution of emerging technologies but also allow researchers and clinicians to formulate research goals that directly address key issues in the delivery of mental health services. This tight collaboration between clinicians and researchers is especially critical given the nascent integration of AI technology into clinical routine, because it can ensure that ethical considerations and the well-being of clients remain front and center.

This paper highlight areas where technological advances in practice and research can significantly enhance POR and the ways in which the POR model itself can contribute to the integration of these technologies in clinical research and practice. We discuss two general phases of care within clinical settings where novel technology and POR can mutually reinforce each other and promote mental health: (1) before treatment, and (2) during treatment. For each phase, we cover some of the main challenges facing the current mental healthcare field, present the major innovations in technology that can substantially advance the field, describe recent studies that illustrate the use of these technologies, and suggest directions for the future. We also consider ethical concerns and the importance of human-technology alignment. Finally, we discuss implications for practice and training.

Before Treatment: How Can Novel Technologies Enhance Prevention, Early Detection, Accessibility, Diagnosis, Prognosis, and Treatment Selection?

The Current Situation

Mental health issues affect millions of people worldwide (World Health Organization, 2022). Currently, most individuals requiring mental health services do not receive any form of treatment (Kazdin, 2021), with greater barriers to access for ethnic minorities (Alegría et al., 2008), low-income populations (Esponda et al., 2020), and rural residents (Hodgkinson et al., 2017). In many cases, mental health issues can be effectively treated or at times prevented through early detection and intervention. However, lack of information about prevention strategies, access to care, and awareness of available mental health services, the cost of mental health services, the shortage of trained mental health professionals delivering in-person sessions, and the stigma around seeking help for mental health issues mean that many individuals do not receive the support they require until their mental health concerns have escalated (Kazdin, 2021).

When these people contact mental health services, they are often sub-optimally diagnosed and frequently mis-assigned to treatment. Most clinicians and psychotherapy researchers administer semi-structured interviews for diagnostic purposes based on classifications defined in the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013) or the International Classification of Diseases (World Health Organization, 2020). However, it is widely acknowledged that categorical diagnoses are error-prone, mainly descriptive rather than explanatory, and too crude to capture the complex heterogeneity and multifactorial nature of mental health problems (e.g., Fried & Nesse, 2015). This heterogeneity is clearly documented in the different clinical presentations, clinical courses, and treatment responses of clients within the same diagnostic group. Research has also shown that diagnostic categories are not mutually exclusive (Bickman et al., 2012). Conventional diagnostic techniques are heavily dependent on clinician's proficiency and expertise in evaluating clients' verbal and non-verbal cues. This know-how, however, may not be readily available or scalable.

Treatment selection decisions are often based on these diagnoses, clinical judgment, intuition, and practical issues, such as the availability of a specific treatment within a service. Nevertheless, evidence suggests that these diagnosis-derived treatment recommendations are prone to errors and reliability issues (e.g., Deacon, 2013).

Numerous attempts have been made to determine the most effective treatment for individual clients. Traditionally, these studies have relied on single or a small set of client variables (e.g., diagnosis, symptom level, interpersonal problems) to determine whether these can predict positive outcomes for a given type of treatment (e.g., cognitive behavioral therapy versus interpersonal psychotherapy; Gomez Penedo et al., 2019). One of the major shortcomings of these studies is their emphasis on the average treatment effect, or group differences, and not on the clients who receive and therapists who deliver treatment. Clients may differ in their responses to different therapists and to the same procedure delivered in different ways by different therapists (Coyne et al., 2022; Huibers et al., 2021). Similarly, therapists may have relative strengths and weaknesses in treating certain types of mental health problems (Boswell et al., 2022). By only comparing averages, the heterogeneous treatment effects in subgroups and individual differences in treatment response are cancelled out despite their importance. (e.g., Lutz et al., 2022a).

The use of classical statistical approaches in research to confirm or refute specific hypotheses is another setback (Bickman, 2020). The traditional inferential approach encounters issues with replication, clinical relevance, accurate application to individuals, and p-value testing (Dwyer et al., 2018). In addition, most psychotherapy studies rely on self-report questionnaires to assess clients' characteristics and symptoms. However, recorded intake interviews contain very rich untapped data, such as the clients' verbal and non-verbal behavior, that can be used to improve diagnosis and treatment selection. These limitations point to the need for alternatives to the traditional ways of doing research that can recognize the complexity of mental health problems, the heterogenicity of clients, therapists, treatment settings, and contexts and the variety of ways in which therapeutic interventions can help or hinder.

The State-of-the-Art and Primary Future Directions

To expand access to treatment, a broad range of technologies are being harnessed for the delivery of therapy (Kazdin, 2021). Several studies have shown that technology-based interventions can be effective for many mental health problems. This provides an opportunity to reduce the gap between the need and access to treatments (Andersson et al., 2019; Lim et al., 2022) since technology-based interventions can reduce barriers such as travel time, scheduling, stigma, and costs (Warmerdam et al., 2010).

Technology-based interventions also generate novel data that psychotherapy researchers can use to monitor clients' mental states, inform decision-making, and improve treatment outcomes. For example, in a study on crisis interventions via text messages, 3.2 million text messages were analyzed and the results showed that specific conversation strategies were associated with better session outcomes (Althoff et al., 2016).

Recent developments in computational methods and the availability of large and multimodal datasets provide opportunities to increase precision in prevention, early detection, diagnosis, prognosis, and treatment selection. Mental health disorders have a marked, observable influence on the expression of affect and interpersonal communication. Clinicians often (subjectively) use clients' verbal and non-verbal behavior for diagnostic purposes. Technological advances such as signal processing, natural language processing (NLP), and computer vision techniques have shown significant potential in improving diagnostic precision by employing computerized methods to capture and model key behavioral signals (for reviews, see Cohn et al., 2018; Graham et al., 2019). The widespread use of smartphones and social media has made it easier to collect vast amounts of data about clients' day-to-day lives, interpersonal relationships, activities, and behavior, which can be valuable for diagnosis and treatment selection. In recent years, researchers have become increasingly interested in using computerized measures to automatically identify clients' mental states in various data modalities such as electronic health records (EHR), voice, text, facial expressions, bio-markers, neuroimaging, physiology, motor activity, questionnaires, and interviews (Bhadra & Kumar, 2022; Cohn et al., 2018). Applying ML techniques to analyze such high-dimensional datasets facilitates the development of risk models that can determine an individual's predisposition or risk of mental illness and enhance diagnosis, prognosis, and treatment selection (Shatte et al., 2019). For example, EHRs include data routinely collected and preserved for each individual over the course of their clinical care. This information is especially valuable for constructing predictive models in psychotherapy research, which can be seamlessly incorporated into care delivery in clinical settings (Chekroud et al., 2021). Several recent studies have demonstrated the usefulness of using large datasets of EHR and ML techniques to detect mental health problems, such as depression (e.g., Choi et al., 2019; Shen et al., 2020) and psychosis (Raket et al., 2020). A recent study used unsupervised ML with large EHR data from the UK Biobank to accurately predict future depression one year or more before it occurred in adults with no previous psychiatric history (Bilu et al., 2023).

Various data modalities can also be valuable for early detection, prevention, diagnosis, and treatment selection. For instance, voice serves as a key medium for expressing and communicating emotions. Analyzing clients' recorded speech patterns can provide a direct, more objective method for evaluating mental states (Juslin & Scherer, 2005). ML techniques can be used to successfully identify mental states

in speech data (for reviews, see Cummins et al., 2015; He et al., 2022). For example, Ma et al. (2016) used deep learning models to recognize the severity of depression in auditory data.

Facial expressions can also help capture non-verbal mental state indicators. Computer vision methods have been employed by researchers to automatically examine facial expressions and identify mood disorders (see Girard et al., 2015; Nasser et al., 2020 for review). For example, Harati et al. (2020) reported that an unsupervised ML approach for analysing muted video data could differentiate between high and low depression severity levels. Other studies have focused on imaging biomarker data and utilized support vector machine models (a supervised ML approach which is often applied for classification tasks) to detect various psychiatric disorders, yielding satisfactory performance (Orrù et al., 2012).

Textual data can also be invaluable for diagnosis, prognosis, and treatment selection, since clients' use of words and language can reflect their inner thoughts and emotions, and reveal crucial information about their mental states. Numerous studies have employed NLP techniques to automatically identify psychiatric conditions from textual data (for reviews, see Castillo-Sánchez et al., 2020; Le Glaz et al., 2021). NLP is a sub-field of AI that enables algorithms to read, understand, and derive meaning from human languages. For example, Haque et al. (2020) demonstrated high accuracy in detecting suicidal ideation from social media data using a transformer-based deep learning approach.

Other technology-based measures, such as ecological momentary assessment and passive data collection from clients via mobile devices or wearables allow for intensive and continuous evaluations, and have shown potential in enhancing diagnosis (for review, see Yim et al., 2020). Recent findings suggest that combining multimodal information is superior to using information from individual modalities in isolation (Santos & Gurevych, 2018).

ML methods and the analysis of rich and large datasets also provide opportunities for predicting prognosis in mental health. One of the first uses of ML methods in mental health research involved training a K-Nearest Neighbors model to predict the outcomes of psychological treatments (Lutz et al., 2005). This model finds cases with highly similar characteristics in a dataset and predicts the outcome for each individual based on the data from their closest neighbors. Since this pioneering research, there has been a surge in models designed to predict treatment outcomes using data collected before treatment. For instance, Gómez Penedo et al. (2021) analyzed the effects of cross-lagged problem-coping experiences on outcomes using dynamic structural equation modeling during the first 10 sessions of therapy. They then used different ML algorithms to predict these outcomes based on clients' initial characteristics. The algorithm that performed the best was a Random Forest algorithm that explained 14.7% of process-outcome association in a training sample. When predicting the same effect on randomly selected validation samples, the results of the algorithm remained stable, explaining 15.4% of the effects on outcome. Sajjadian et al. (2021) analyzed 54 studies that built models to predict responses to antidepressant treatments in clients with major depressive disorder. These studies employed a variety of ML techniques, including random forest, extreme gradient boosting, least absolute shrinkage and selection operator (LASSO) regularization, elastic net, naïve Bayes, support vector machine, and others. Although some of the studies had small sample sizes and limited evidence of external cross-validation, the review found promising prediction accuracy indicators (internal accuracy 0.71–0.86; external accuracy 0.70–0.79).

Others have demonstrated the usefulness of ML techniques in improving precision in treatment selection. For example, the Personalized Advantage Index model developed by DeRubeis et al. (2014) predicts the most effective treatment for specific clients based on their pre-treatment characteristics. Studies have used this model to determine which therapy would be the most beneficial for a given client. The findings show that clients who received the "optimal" treatment tended to have better treatment outcomes (e.g., Deisenhofer et al., 2018; Schwartz et al., 2021). Some studies have developed models to help mental health providers choose low- or high-intensity treatments for clients, depending on their expected prognosis. For instance, Lorenzo-Luaces et al. (2017) created prognostic indices using a LASSO-style bootstrap variable selection procedure to predict recovery from depression. Their findings indicated that clients with a high prognostic index had similar recovery rates regardless of treatment, whereas clients with the poorest prognosis had significantly higher recovery rates in the high intensity condition. Research has begun to use pre-treatment data to identify subgroups of clients who respond differently to available treatments (Delgadillo & Lutz, 2020), those who are at risk of early dropout (Bennemann et al., 2022), or those who may benefit the most from therapist strength-based matching (Boswell et al., 2022).

These findings align with the emerging focus on precision mental health. They are consistent with the widely held belief among professionals and researchers that psychological treatments should be personalized (Cohen et al., 2021). The incorporation of advanced technologies for long-term individual monitoring, coupled with broader treatment delivery methods, and the emergence of AI and ML can enhance and personalize mental healthcare. This could make mental health support more accessible and better targeted to meet the needs of individuals requiring assistance.

The POR model could be instrumental in guiding the development and integration of these technologies into clinical practice. The ongoing interaction between clinicians and researchers enables feedback from real clinical situations, so that technology-assisted diagnosis, prognosis, and treatment selection can be rapidly incorporated in ways that may lead to more precise and effective treatment. When clinicians play an active role in shaping and fine-tuning these AI tools, their trust in and acceptance of these technologies is likely to increase, leading to a more seamless integration into everyday clinical practice.

During Treatment: How Can Novel Technologies Advance Treatment Course Guidance?

The Current Situation

Over the past four decades, a multitude of evidence-based treatment approaches to mental health disorders have been developed. Unfortunately, up to half of all clients do not experience significant benefits from the treatment they receive (Cuijpers et al., 2022). Individual responses to these interventions vary considerably, with some clients showing significant improvement, while others seeing little to no improvement, or even experiencing deterioration (Cuijpers et al., 2022). In addition, many clients drop out of treatment prematurely, which makes it difficult to ascertain the extent of their progress and whether their status would have improved if they had continued with treatment (Lutz et al., 2018). Among those who do benefit from treatment, a significant proportion do not achieve remission, and many experience relapse within a year (Cuijpers et al., 2022). These findings highlight the need for personalized approaches to mental health treatment that can consider individual differences in treatment response and lead to better outcomes.

Feedback in psychotherapy aims to track clients' progress over the course of their treatment and provide this information to the therapist to enhance client outcomes. Incorporating feedback into routine clinical practice and training has been one of the most successful developments in psychotherapy research in the last 20 years (Lutz et al., 2022a). Given that many therapists do not gather any data and are often unsure of their clients' progress, providing consistent feedback presents an accessible and cost-effective strategy to enhance outcomes (Barkham, 2023). Research has shown that giving therapists regular updates on their clients' progress and using clinical support tools can improve treatment effectiveness and decrease dropout rates (e.g., Bar-Kalifa et al., 2016; de Jong et al., 2021). However, most clinical support tools only provide information about clients' symptoms rather than suggestions as to which interventions to use with whom and when. Therapists would benefit from feedback that helps them select the most appropriate interventions for specific clients at particular times, elucidate the session processes and outcomes, identify their strengths and weaknesses, and provide guidance for future sessions (Lutz et al., 2022a).

Despite decades of research on the processes and mechanisms underlying therapeutic change, much remains to be known about how and why treatment works (Crits-Christoph & Gibbons, 2021; Constantino et al., 2021). Data on mechanisms of change are also mostly analyzed at the group level; however, different interventions can have varying impacts on clients depending on the timing and context of implementation.

The technology used to analyze the client-therapist interactions, the active component of psychotherapy, has not undergone significant change in decades, thereby restricting the scale and specificity of process-outcome research (Imel et al., 2015). Similar to diagnostic and treatment selection studies, most research relies on self-report measures (Crits-Christoph & Gibbons, 2021). Although standardized subjective measures are fundamental to psychotherapy research, they have significant limitations, such as the extent of participants' self-awareness, their willingness to complete questionnaires, and the limited choice of responses (Kazdin, 2008). To examine what occurs within psychotherapy sessions, researchers have developed various observer coding systems (e.g., Pascual-Leone & Greenberg, 2007) that can provide insights into the moment-to-moment interactions between therapists and clients that contribute to therapeutic change. However, these studies typically only involve a few therapeutic components, a relatively small sample of clients, and limited time points, since human observational coding is highly labor-intensive and expensive to implement. The use of small samples limits progress in studying more complex processes in psychotherapy, such as how sequences of moment-to-moment positive effects can accumulate into larger change within and across treatment sessions. To determine the most effective components of psychotherapy for specific clients at specific times, the data need to be quantified and gathered from a sufficiently large sample to draw meaningful conclusions.

The State-of-the-Art and Primary Future Directions

Research on the processes and mechanisms that underly therapeutic change can be substantially enriched using AI and personalized approaches. When utilizing session-level information, ML models can potentially predict the most helpful interventions for specific clients and the ideal sequence in which these interventions should be

administered. This crucial information could then be communicated back to therapists to guide their interventions, fostering a more personalized therapy approach (Delgadillo & Atzil-Slonim, 2022). For example, the Trier Treatment Navigator, a pioneering project outlined by Lutz et al. (2019), alerts therapists when clients' symptoms are not improving as anticipated, and recommends alternate clinical interventions based on extensive session-level data from that client and other clients with similar characteristics. These data are also beneficial for supervision and training by helping therapists recognize their own strengths and areas for improvement, and achieving a better understanding of their clients' needs.

Parallel advances have been observed in the domain of chatbots for psychotherapy. Current chatbots for psychotherapy largely depend on predefined response options (Lim et al., 2022). Nonetheless, progress in LLMs and generative AI have the potential to enable personalized feedback for clients, therapists, and supervisors (Stade et al., 2023). Ongoing research is already exploring these possibilities. For instance, a recent study employed LLMs to create an AI in the loop agent, which offers feedback to counselors to enhance their empathetic responses in text-based conversations (Sharma et al., 2023).

The use of computerized methods extends into the analyses of varied sources of within-session data, such as the words clients use, their tone of voice, facial expressions, and physiological states (Schwartz et al., 2023). By focusing on granular elements within psychotherapy sessions and employing automated measures, the ability to upscale research and refine the specificity of understanding intervention effectiveness for individual clients is greatly augmented. This approach ultimately contributes to a more profound and comprehensive understanding of psychotherapy processes and outcomes (Imel et al., 2015). For instance, several studies have focused on the vocal channel to identify subtle yet clinically relevant changes in affective states in psychotherapy (e.g., Soma et al., 2020). In a study that examined clients' and therapists' intra- and interpersonal vocal affect dynamics using a measure that combined several acoustic features, positive associations were observed between these dynamics and treatment outcome both within sessions and over treatment (Paz et al., 2021).

Other studies have focused on text analysis approaches given that the dialogue between the client and the therapist can reveal important information about their interaction (Tausczik & Pennebaker, 2010; Warikoo et al., 2022). For instance, Atzil-Slonim et al. (2021) employed topic modeling to autonomously identify themes explored in therapy sessions, by utilizing ML methodologies to investigate which topics could predict client functionality and potential ruptures in the therapeutic alliance. Other studies have used deep learning techniques to automatically categorize therapists' utterances in internet-facilitated cognitive behavioral therapy (Ewbank et al., 2020), as well as both client and therapist utterances in motivational interviewing (Cao et al., 2019). These findings suggest that ML presents a promising avenue for large-scale annotation of the therapeutic dialogue.

Physiological and biological measures can also capture emotional regulation processes which are central to many psychological conditions and are thus a primary target of numerous therapeutic interventions. Capturing these regulatory dynamics within therapy necessitates the analysis of the ways that emotions fluctuate within and across individuals and dyads at the session level. For instance, Bar-Kalifa et al. (2019) monitored the electrodermal activity of clients and therapists during sessions to investigate the role of physiological synchrony during emotion-focused techniques versus cognitive-behavioral techniques. They found that increased synchrony during emotion-focused segments (but not cognitive-behavioral ones) was linked to a stronger therapeutic alliance. Several biomarkers (including those related to hormones, immune response, and inflammation) have been explored as potential indicators of progress in psychotherapy (Cristea et al., 2019). For instance, Atzil-Slonim et al. (2022) found a correlation between an increased oxytocin response to the therapeutic interaction and a reduction in depressive symptoms throughout the course of treatment.

These technological advances in the ways individuals and treatments are monitored, and in data analysis methods may also play a vital role in post-treatment follow-up and the long-term monitoring of clients who have completed psychotherapy treatment. For example, ML algorithms could be trained to predict the risk of relapse based on multi-modal data that can be captured by digital tools. This could provide early warning signs and prompt intervention if needed. These tools could be used to continuously monitor clients' well-being and provide self-health support or encourage the client to get professional help, depending on their specific needs.

The POR model is essential in shaping technologically based treatment course guidance. Undoubtedly, successful collaboration between clinicians and AI technologies in mental health care requires a comprehensive understanding of each party's strengths and weaknesses and how they can mutually enhance one another to improve psychotherapy effectiveness. When clinicians participate in the design and development of technology-based treatment guidance, it markedly increases their likelihood of utilizing the feedback more effectively. This collaborative approach also allows for the further development and improvement of the capabilities of these systems in ways that may lead to better

understandings of the mechanisms that underlie therapeutic change and to increasing the effectiveness of treatments.

Summary, Limitations, and Implication for Practice and Training

Significant technological advances are propelling POR forward. The application of technologically-based services can enhance the accessibility and breadth of mental health services, thereby bridging the gap between populations requiring these services and those who can actually avail them. This is especially crucial in the context of addressing long-standing inequities in mental healthcare access and effectiveness, especially for racial or ethnic minorities.

The implementation of AI and ML methods, coupled with multimodal datasets and individualized strategies, is facilitating a shift from oversimplified diagnostic evaluations to comprehensive, multi-faceted profiles. By correlating these profiles with outcomes, identifying clinically relevant subgroups for prognosis and determining which clients will respond differently to various interventions and providers becomes feasible.

By pinpointing the specific characteristics and needs of clients, as well as therapists' specific competencies that predict improved treatment results, these factors can be incorporated into algorithms that yield practical treatment recommendations. These algorithms can be embedded in pre-treatment support tools, enabling more precise client-therapist matching and selection of the most beneficial treatment for each individual. This can contribute to remedying current trial-and-error in treatment and therapist selection, and lead to more prompt delivery of effective care, improved mental health outcomes, and a decrease in associated delays and costs.

The speed and accuracy of computer technology also allow for a robust and reliable analysis of large amounts of within-session non-verbal (e.g., movement, voice, etc.) and verbal (e.g., speech) data, as well as the automatization of complex classification tasks (e.g., which intervention was applied at a certain moment in time). Implementing these technologies in psychotherapy data analysis could significantly reduce the reliance on human coders and allow for a more cost-effective analysis of larger datasets, potentially yielding more reliable insights into the processes driving therapeutic progress.

These insights can be incorporated into session-by-session support tools and training and supervision programs to enhance therapists' skills. For instance, under the guidance of their supervisors, therapists can use feedback from the AI system to better understand their clients' productive and less productive processes, as well as their own strengths and areas for improvement. This, in turn, can help select the most appropriate interventions for a specific therapist to use with a specific client at a specific time. These technologies could also be further developed for post-treatment follow-ups and long-term tracking of clients who have completed psychotherapy. Future studies would benefit from testing whether ML algorithms can predict relapse risks using diverse data gathered via digital tools designed to enable early detection and intervention. These tools could check clients' well-being, offer self-care support, or recommend professional help based on individual needs.

Although cutting-edge technologies hold significant promise for POR, they face specific challenges and constraints. One primary concern is ethics. Studies have indicated that inherent biases (such as gender and race) in the original datasets lead to prejudiced AI decisions (Obermeyer et al., 2019). Developing unbiased applications may help reduce this form of discrimination in AI (e.g., Yeung, 2018). Gathering extensive multi-modal datasets which contain sensitive and personal data necessitates the guarantee of stringent data protection standards.

As AI systems become increasingly powerful, there is a risk that they may unintentionally target goals misaligned with human needs, leading to unexpected outcomes. This AI alignment problem is particularly crucial in mental healthcare where AI interacts with vulnerable groups. Misalignment could result in inappropriate advice, misinterpretations of severe distress, or failure to refer to professionals when necessary. There is also a concern about "overtrust" in AI, where individuals might depend excessively on AI and believe it to be infallible. To avoid these risks, it is essential to align AI systems with human values and needs. This requires integrating safety and ethical considerations into AI design and ensuring ongoing monitoring and evaluation in real-world scenarios.

There are also implementation challenges. Mental health professionals and services frequently hesitate before embracing new technologies. Although these pioneering technologies are projected to be cost-effective in the long run, they are currently expensive and may not be accessible to researchers, practitioners, and clients in various regions of the world. Moreover, AI models do not always generate sufficiently accurate predictions (Chekroud et al., 2021). These models are also often perceived as "black boxes" that do not provide rationales for their predictions (Chekroud et al., 2021). Although data scientists are developing methods for better "explainability" (e.g., Lundberg & Lee, 2017), it will take time for clinicians to comprehend why models make certain predictions, accord greater trust to these models, and consider incorporating them into their clinical decision-making procedures. Additionally, AI advances might be viewed by therapists as a possible threat, that can prompt

fears as to the potential replacement of their profession by these technologies.

These considerations underscore the essential role of POR in fostering vigorous collaborations between mental health professionals and researchers in assimilating these technologies into clinical routines. POR can contribute to showing that AI is intended to enhance and assist the efforts of mental health professionals, rather than supplant them. For broader implementation, AI-based clinical tools should be developed in close cooperation with clinicians and researchers and be guided by clients' preferences and reality considerations. AI technologies can support mental health practitioners and researchers, which may lead to improvements in the quality, accessibility, consistency, and scalability of therapeutic treatments and clinical research. By generating cutting-edge technologies guided by clinical knowledge and research findings, and studying the functionality of these systems as they are incorporated into daily practices, the chances of clients obtaining significant benefits from these technologies can be amplified. However, it is crucial to keep in mind that therapists must be adequately trained to utilize these novel technologies. AI systems should not replace therapists in making clinical decisions. Rather, it is the responsibility of well-trained therapists to expertly weave AI suggestions into their clinical thinking and practice.

A considerable amount of work is needed to scientifically validate the clinical benefits of these technologies before they are adopted in routine practice. While this article underscored the importance of the POR model in assimilating novel technologies into clinical routines by focusing on studies that have investigated psychotherapy in routine clinical settings, it is imperative that the accuracy and beneficial value for client outcomes of these technologies be rigorously assessed through prospective randomized controlled studies. Recent research, including studies by Delgadillo et al. (2021), Fletcher et al. (2021); Lutz et al. (2022b), has highlighted the effectiveness of data-driven and technology-supported methods in improving treatment outcomes. It is crucial to devote more efforts to ensuring that these technological advances are ethical, cost-effective, reliable, explainable, and appealing to both mental health providers and users.

The exploration of using novel technologies and AI in psychotherapy is in its early stages, facing numerous challenges. The role of POR in this context is crucial as it can guide the integration of these emerging technologies into clinical research and practice. As we move forward, the impact of merging these technologies into research, practice, and training will become more apparent, potentially bringing about the significant advancements we aspire to see in mental healthcare services.

References

Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., & Aafjes, M. (2021). A scoping review of machine learning in psychotherapy research. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *31*(1), 92–116. https://doi.org/10.1080/10503307.2020.1808729

Alegría, M., Chatterji, P., Wells, K., Cao, Z., Chen, C. N., Takeuchi, D., Jackson, J., & Meng, X. L. (2008). Disparity in depression treatment among racial and ethnic minority populations in the United States. *Psychiatric Services*, *59*(11), 1264–1272. https://doi.org/10.1176/ps.2008.59.11.1264

Alpaydin, E. (2020). *Introduction to machine learning*. MIT Press.

Althoff, T., Clark, K., & Leskovec, J. (2016). Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics*, *4*, 463–476. https://doi.org/10.1162/tacl_a_00111

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. American Psychiatric Association.

Andersson, G., Carlbring, P., Titov, N., & Lindefors, N. (2019). Internet interventions for adults with anxiety and mood disorders: A narrative umbrella review of recent meta-analyses. *Canadian Journal of Psychiatry Revue Canadienne de Psychiatrie*, *64*(7), 465–470. https://doi.org/10.1177/0706743719839381

Atzil-Slonim, D., Juravski, D., Bar-Kalifa, E., Gilboa-Schechtman, E., Tuval-Mashiach, R., Shapira, N., & Goldberg, Y. (2021). Using topic models to identify clients' functioning levels and alliance ruptures in psychotherapy. *Psychotherapy*, *58*(2), 324–339. https://doi.org/10.1037/pst0000362

Atzil-Slonim, D., Stolowicz-Melman, D., Bar-Kalifa, E., Gilboa-Schechtman, E., Paz, A., Wolff, M., Rotter, I., Zagoory, O., & Feldman, R. (2022). Oxytocin reactivity to the therapeutic encounter as a biomarker of change in the treatment of depression. *Journal of Counseling Psychology*, *69*(5), 755–760. https://doi.org/10.1037/cou0000617

Bar-Kalifa, E., Atzil-Slonim, D., Rafaeli, E., Peri, T., Rubel, J., & Lutz, W. (2016). Therapist–client agreement in assessments of clients' functioning. *Journal of Consulting and Clinical Psychology*, *84*(12), 1127–1134. https://doi.org/10.1037/ccp0000157

Bar-Kalifa, E., Prinz, J. N., Atzil-Slonim, D., Rubel, J. A., Lutz, W., & Rafaeli, E. (2019). Physiological synchrony and therapeutic alliance in an imagery-based treatment. *Journal of Counseling Psychology*, *66*(4), 508–517. https://doi.org/10.1037/cou0000358

Barkham, M. (2023). Smaller effects matter in the psychological therapies: 25 years on from Wampold et al. (1997). *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *33*(4), 530–532. https://doi.org/10.1080/10503307.2022.2141589

Bennemann, B., Schwartz, B., Giesemann, J., & Lutz, W. (2022). Predicting patients who will drop out of out-patient psychotherapy using machine learning algorithms. *British Journal of Psychiatry: The Journal of Mental Science*, *220*(4), 1–10. https://doi.org/10.1192/bjp.2022.17

Bhadra, S., & Kumar, C. J. (2022). An insight into diagnosis of depression using machine learning techniques: A systematic review. *Current Medical Research and Opinion*, *38*(5), 749–771. https://doi.org/10.1080/03007995.2022.2038487

Bickman, L. (2020). Improving mental health services: A 50-year journey from randomized experiments to artificial intelligence and precision mental health. *Administration and Policy in Mental Health*, *47*(5), 795–843. https://doi.org/10.1007/s10488-020-01065-8

Bickman, L., Wighton, L. G., Lambert, E. W., Karver, M. S., & Steding, L. (2012). Problems in using diagnosis in child and adolescent mental health services research. *Journal of Methods and*

Measurement in the Social Sciences, 3(1), 1, Retrieved from https://doi.org/10.2458/jmm.v3i1.16110

Bilu, Y., Kalkstein, N., Gilboa-Schechtman, E., Akiva, P., Zalsman, G., Itzhaky, L., & Atzil-Slonim, D. (2023). Predicting future onset of depression among middle-aged adults with no psychiatric history. *BJPsych Open*, *9*(3), e85. https://doi.org/10.1192/bjo.2023.62

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

Boswell, J. F., Constantino, M. J., Coyne, A. E., & Kraus, D. R. (2022). For whom does a match matter most? Patient-level moderators of evidence-based patient–therapist matching. *Journal of Consulting and Clinical Psychology*, *90*(1), 61–74. https://doi.org/10.1037/ccp0000644

Cao, J., Tanana, M., Imel, Z. E., Poitras, E., Atkins, D. C., & Srikumar, V. (2019). Observing dialogue in therapy: Categorizing and forecasting behavioral codes. In Proceedings of the 57th conference of the Association for Computational Linguistics, (pp. 5599–5611). https://doi.org/10.18653/v1/P19-1563

Castillo-Sánchez, G., Marques, G., Dorronzoro, E., Rivera-Romero, O., Franco-Martín, M., & De la Torre-Díez, I. (2020). Suicide risk assessment using machine learning and social networks: A scoping review. *Journal of Medical Systems*, *44*(12), 205. https://doi.org/10.1007/s10916-020-01669-5

Castonguay, L. G., Barkham, M., Youn, S. J., & Page, A. C. (2021). Practice-based evidence- findings from routine clinical settings. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and garfield's handbook of psychotherapy and behavior change* (pp. 129–186). Wiley.

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., & Dwyer, D. (2021). *The promise of machine learning in predicting Treatment outcomes in psychiatry*. *World Psychiatr* 20 (2), 154–170.

Choi, K. W., Chen, C. Y., Stein, M. B., Klimentidis, Y. C., Wang, M. J., Koenen, K. C., Smoller, J. W., & Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. (2019). Assessment of bidirectional relationships between physical activity and depression among adults: A 2-sample mendelian randomization study. *JAMA Psychiatry*, *76*(4), 399–408. https://doi.org/10.1001/jamapsychiatry.2018.4175

Cohen, Z., Delgadillo, J., & DeRubeis, R. (2021). Personalized treatment approaches. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and garfield's handbook of psychotherapy and behavior change* (pp. 667–700). Wiley.

Cohn, J. F., Cummins, N., Epps, J., Goecke, R., Joshi, J., & Scherer, S. (2018). Multimodal assessment of depression from behavioral signals. In *The handbook of multimodal-multisensor interfaces: Signal processing, architectures, and detection of emotion and cognition-volume* 2, (pp. 375–417).

Constantino, M. J., Boswell, J. F., & Coyne, A. E. (2021). Patient, therapist, and relational factors. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and garfield's handbook of psychotherapy and behavior change* (pp. 225–262). Wiley.

Coyne, A. E., Constantino, M. J., Boswell, J. F., & Kraus, D. R. (2022). Therapist-level moderation of within- and between-therapist process-outcome associations. *Journal of Consulting and Clinical Psychology*, *90*(1), 75–89. https://doi.org/10.1037/ccp0000676

Cristea, I. A., Karyotaki, E., Hollon, S. D., Cuijpers, P., & Gentili, C. (2019). Biological markers evaluated in randomized trials of psychological treatments for depression: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *101*, 32–44. https://doi.org/10.1016/j.neubiorev.2019.03.022

Crits-Christoph, P. A. U. L., & Gibbons, M. B. C. (2021). *Psychotherapy process–outcome research: Advances in understanding causal connections. Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 263–296). Wiley.

Cuijpers, P., Ciharova, M., Quero, S., Miguel, C., Driessen, E., Harrer, M., Purgato, M., Ebert, D., & Karyotaki, E. (2022). The contribution of 'Individual participant data' meta-analyses of psychotherapies for depression to the development of personalized treatments: A systematic review. *Journal of Personalized Medicine*, *12*(1), 93. https://doi.org/10.3390/jpm12010093

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, *71*, 10–49. https://doi.org/10.1016/j.specom.2015.03.004

de Jong, K., Conijn, J. M., Gallagher, R. A. V., Reshetnikova, A. S., Heij, M., & Lutz, M. C. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review*, *85*, 102002. https://doi.org/10.1016/j.cpr.2021.102002

Deacon, B. J. (2013). The biomedical model of mental disorder: A critical analysis of its validity, utility, and effects on psychotherapy research. *Clinical Psychology Review*, *33*(7), 846–861. https://doi.org/10.1016/j.cpr.2012.09.007

Deisenhofer, A. K., Delgadillo, J., Rubel, J. A., Böhnke, J. R., Zimmermann, D., Schwartz, B., & Lutz, W. (2018). Individual treatment selection for patients with posttraumatic stress disorder. *Depression and Anxiety*, *35*(6), 541–550. https://doi.org/10.1002/da.22755

Delgadillo, J., Ali, S., Fleck, K., Agnew, C., Southgate, A., Parkhouse, L., Cohen, Z., DeRubeis, R., & Barkham, M. (2021). Stratified care vs stepped care for depression: a cluster randomized clinical trial. *JAMA Psychiatry*, *79*(2), 101–108. https://doi.org/10.1001/jamapsychiatry.2021.3539

Delgadillo, J., & Atzil-Slonim, D. (2022). Artificial intelligence, machine learning and mental health. In H. S. Friedman & C. H. Markey (Eds.), *Encyclopedia of Mental Health* (3rd Edition), (pp 132–142). Elsevier. https://doi.org/10.1016/B978-0-323-91497-0.00177-6

Delgadillo, J., & Lutz, W. (2020). A development pathway towards precision mental health care. *JAMA Psychiatry*, *77*(9), 889–890. https://doi.org/10.1001/jamapsychiatry.2020.1048

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The personalized advantage index: Translating research on prediction into individualized treatment recommendations. A demonstration. *PLOS ONE*, *9*(1), e83875. https://doi.org/10.1371/journal.pone.0083875

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pretraining of deep bidirectional transformers for language understanding. *arXiv Preprint* arXiv:1810.04805.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, *14*(1), 91–118, Retrieved from https://doi.org/10.1146/annurev-clinpsy-032816-045037

Esponda, G. M., Hartman, S., Qureshi, O., Sadler, E., Cohen, A., & Kakuma, R. (2020). Barriers and facilitators of mental health programmes in primary care in low-income and middle-income countries. *The Lancet Psychiatry*, *7*(1), 78–92. https://doi.org/10.1016/S2215-0366(19)30125-7

Ewbank, M. P., Cummins, R., Tablan, V., Catarino, A., Buchholz, S., & Blackwell, A. D. (2020). Understanding the relationship between patient language and outcomes in internet-enabled cognitive behavioural therapy: A deep learning approach to automatic coding of session transcripts. *Psychotherapy Research*, 1–13.

Flemotomos, N., Martinez, V. R., Chen, Z., Creed, T. A., Atkins, D. C., & Narayanan, S. (2021). Automated quality assessment of cognitive behavioral therapy sessions through highly contextualized

language representations. *PLOS ONE*, *16*(10), e0258639. https://doi.org/10.1371/journal.pone.0258639

Fletcher, S., Spittal, M. J., Chondros, P., Palmer, V. J., Chatterton, M. L., Densley, K., Potiriadis, M., Harris, H., Bassilios, B., Burgess, P., Mihalopoulos, C., Pirkis, J., & Gunn, J. (2021). Clinical efficacy of a decision Support Tool (Link-me) to guide intensity of mental health care in primary practice: A pragmatic stratified randomised controlled trial. *The Lancet Psychiatry*, *8*(3), 202–214. https://doi.org/10.1016/s2215-0366(20)30517-4

Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR* D study. *Journal of Affective Disorders*, *172*, 96–102. https://doi.org/10.1016/j.jad.2014.10.010

Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines*, *30*(3), 411–437. https://doi.org/10.1007/s11023-020-09539-2

Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc.

Girard, J. M., Cohn, J. F., Jeni, L. A., Sayette, M. A., & De la Torre, F. (2015). Spontaneous facial expression in unscripted social interactions can be measured automatically. *Behavior Research Methods*, *47*(4), 1136–1147. https://doi.org/10.3758/s13428-014-0536-1

Gomez Penedo, J. M., Constantino, M. J., Coyne, A. E., Bernecker, S. L., & Smith-Hansen, L. (2019). Patient baseline interpersonal problems as moderators of outcome in two psychotherapies for bulimia nervosa. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *29*(6), 799–811. https://doi.org/10.1080/10503307.2018.1425931

Gómez Penedo, J. M., Schwartz, B., Giesemann, J., Rubel, J. A., Deisenhofer, A. K., & Lutz, W. (2021). For whom should psychotherapy focus on problem coping? A machine learning algorithm for treatment personalization. *Psychotherapy Research*, 1–14. https://doi.org/10.1080/10503307.2021.1930242

Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H. C., & Jeste, D. V. (2019). Artificial intelligence for mental health and mental illnesses: An overview. *Current Psychiatry Reports*, *21*(11), 116. https://doi.org/10.1007/s11920-019-1094-0

Haque, F., Nur, R. U., Jahan, A., Mahmud, S. A., Z., & Shah, F. M. (2020). A transformer based approach to detect suicidal ideation using pre-trained language models. In 2020 23rd international conference on computer and information technology (ICCIT), (pp. 1–5). https://doi.org/10.1109/ICCIT51783.2020.9392692. IEEE Publications.

Harati, S., Crowell, A., Huang, Y., Mayberg, H., & Nemati, S. (2020). Classifying depression severity in recovery from major depressive disorder via dynamic facial features. *IEEE Journal of Biomedical and Health Informatics*, *24*(3), 815–824. https://doi.org/10.1109/JBHI.2019.2930604

He, L., Niu, M., Tiwari, P., Marttinen, P., Su, R., Jiang, J., Guo, C., Wang, H., Ding, S., Wang, Z., Pan, X., & Dang, W. (2022). Deep learning for depression recognition with audiovisual cues: A review. *Information Fusion*, *80*, 56–86. https://doi.org/10.1016/j.inffus.2021.10.012

Hermes, E. D. A., Lyon, A. R., Schueller, S. M., & Glass, J. E. (2019). Measuring the implementation of behavioral intervention technologies: Recharacterization of established outcomes. *Journal of Medical Internet Research*, *21*(1), e11752. https://doi.org/10.2196/11752

Hodgkinson, S., Godoy, L., Beers, L. S., & Lewin, A. (2017). Improving mental health access for low-income children and families in the primary care setting. *Pediatrics*, *139*(1), https://doi.org/10.1542/peds.2015-1175

Huibers, M. J., Lorenzo-Luaces, L., Cuijpers, P., & Kazantzis, N. (2021). On the road to personalized psychotherapy: A research agenda based on cognitive behavior therapy for depression. *Frontiers in Psychiatry*, *11*, 1551.

Imel, Z. E., Steyvers, M., & Atkins, D. C. (2015). Computational psychotherapy research: Scaling up the evaluation of patient-provider interactions. *Psychotherapy*, *52*(1), 19–30. https://doi.org/10.1037/a0036841

Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 65–135). Series in Affective Science Oxford University Press.

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, *63*(3), 146–159. https://doi.org/10.1037/0003-066X.63.3.146

Kazdin, A. E. (2021). Extending the scalability and reach of psychosocial interventions. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 763–790). Wiley.

Le Glaz, A., Haralambous, Y., Kim-Dufor, D. H., Lenca, P., Billot, R., Ryan, T. C., Marsh, J., DeVylder, J., Walter, M., Berrouiguet, S., & Lemey, C. (2021). Machine learning and natural language processing in mental health: Systematic review. *Journal of Medical Internet Research*, *23*(5), e15708. https://doi.org/10.2196/15708

Lim, S. M., Shiau, C. W. C., Cheng, L. J., & Lau, Y. (2022). Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: A systematic review and meta-regression. *Behavior Therapy*, *53*(2), 334–347. https://doi.org/10.1016/j.beth.2021.09.007

Lorenzo-Luaces, L., DeRubeis, R. J., van Straten, A., & Tiemens, B. (2017). A prognostic index (pi) as a moderator of outcomes in the treatment of depression: A proof of concept combining multiple variables to inform risk-stratified stepped care models. *Journal of Affective Disorders*, *213*, 78–85. https://doi.org/10.1016/j.jad.2017.02.010

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C., Noble, R., & Iveson, S. (2005). Predicting change for individual psychotherapy clients on the basis of their nearest neighbors. *Journal of Consulting and Clinical Psychology*, *73*(5), 904–913. https://doi.org/10.1037/0022-006X.73.5.904

Lutz, W., Schwartz, B., Hofmann, S. G., Fisher, A. J., Husen, K., & Rubel, J. A. (2018). Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: A methodological proof-of-concept study. *Scientific Reports*, *8*(1), 7819. https://doi.org/10.1038/s41598-018-25953-0

Lutz, W., Rubel, J. A., Schwartz, B., Schilling, V., & Deisenhofer, A. K. (2019). Towards integrating personalized feedback research into clinical practice: Development of the Trier Treatment Navigator (TTN). *Behaviour Research and Therapy*, *120*, 103438. https://doi.org/10.1016/j.brat.2019.103438

Lutz, W., De Jong, K., Rubel, J., & Delgadillo, J. (2021). Measuring, predicting, and tracking change in psychotherapy. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and garfield's handbook of psychotherapy and behavior change* (pp. 89–134). Wiley.

Lutz, W., Deisenhofer, A. K., Rubel, J., Bennemann, B., Giesemann, J., Poster, K., & Schwartz, B. (2022a). Prospective evaluation of a clinical decision support system in psychological therapy. *Journal of Consulting and Clinical Psychology*, *90*(1), 90–106. https://doi.org/10.1037/ccp0000642

Lutz, W., Schwartz, B., & Delgadillo, J. (2022b). Measurement-based and data-informed psychological therapy. *Annual Review of Clinical Psychology*, *18*, 71–98. https://doi.org/10.1146/annurev-clinpsy-071720-014821

Ma, X., Yang, H., Chen, Q., Huang, D., & Wang, Y. (2016). Depaudionet: An efficient deep model for audio based depression classification. In Proceedings of the 6th international workshop on audio/visual emotion challenge, (pp. 35–42). https://doi.org/10.1145/2988257.2988267

Nasser, S. A., Hashim, I. A., & Ali, W. H. (2020). A review on depression detection and diagnoses based on visual facial cues 2020 3rd International Conference on Engineering Technology and Its Applications (IICETA), (pp. 35–40). https://doi.org/10.1109/IICETA50496.2020.9318860

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neuroscience and Biobehavioral Reviews*, *36*(4), 1140–1152. https://doi.org/10.1016/j.neubiorev.2012.01.004

Pascual-Leone, A., & Greenberg, L. S. (2007). Emotional processing in experiential therapy: Why 'the only way out is through'. *Journal of Consulting and Clinical Psychology*, *75*(6), 875.

Paz, A., Rafaeli, E., Bar-Kalifa, E., Gilboa-Schectman, E., Gannot, S., Laufer-Goldshtein, B., Narayanan, S., Keshet, J., & Atzil-Slonim, D. (2021). Intrapersonal and interpersonal vocal affect dynamics during psychotherapy. *Journal of Consulting and Clinical Psychology*, *89*(3), 227–239. https://doi.org/10.1037/ccp0000623

Raket, L. L., Jaskolowski, J., Kinon, B. J., Brasen, J. C., Jönsson, L., Wehnert, A., & Fusar-Poli, P. (2020). Dynamic ElecTronic hEalth reCord deTection (DETECT) of individuals at risk of a first episode of psychosis: A case-control development and validation study. *The Lancet Digital Health*, *2*(5), e229–e239. https://doi.org/10.1016/S2589-7500(20)30024-8

Sajjadian, M., Lam, R. W., Milev, R., Rotzinger, S., Frey, B. N., Soares, C. N., Parikh, S. V., Foster, J. A., Turecki, G., Müller, D. J., & Strother, S. C. (2021). Machine learning in the prediction of depression treatment outcomes: A systematic review and meta-analysis. *Psychologie Medicale*, *51*(16), 2742–2751.

Santos, P. B., & Gurevych, I. (2018). Multimodal prediction of the audience's impression in political debates. In Proceedings of the 20th international conference on multimodal interaction: Adjunct, (pp. 1–6). https://doi.org/10.1145/3281151.3281157

Schwartz, B., Cohen, Z. D., Rubel, J. A., ZimMermann, D., Wittmann, W. W., & Lutz, W. (2021). Personalized treatment selection in routine care: Integrating machine learning and statistical algorithms to recommend cognitive behavioral or psychodynamic therapy. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *31*(1), 33–51. https://doi.org/10.1080/10503307.2020.1769219

Schwartz, B., Uhl, J., & Atzil-Slonim, D. (2023). Assessments and measures in psychotherapy research: going beyond self-report data. *Frontiers in Psychiatry*, *14*, 1276222. https://doi.org/10.3389/fpsyt.2023.1276222

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. *Nature Machine Intelligence*, *5*(1), 46–57.

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*, *49*(9), 1426–1448. https://doi.org/10.1017/S0033291719000151

Shen, X., Howard, D. M., Adams, M. J., Hill, W. D., Clarke, T. K., Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Deary, I. J., Whalley, H. C., & McIntosh, A. M. (2020). A phenome-wide association and mendelian randomisation study of polygenic risk for depression in UK Biobank. *Nature Communications*, *11*(1), 2301. https://doi.org/10.1038/s41467-020-16022-0

Soma, C. S., Baucom, B. R., Xiao, B., Butner, J. E., Hilpert, P., Narayanan, S., & Imel, Z. E. (2020). Coregulation of therapist and client emotion during psychotherapy. *Psychotherapy Research*, *30*(5), 591–603. https://doi.org/10.1080/10503307.2019.1661541

Stade, E., Stirman, S. W., Ungar, L. H., Yaden, D. B., Schwartz, H. A., Sedoc, J., DeRubeis, R., Willer, R., & Eichstaedt, J. C. (2023). *Artificial intelligence will change the future of psychotherapy: A proposal for responsible, psychologist-led development*.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Warikoo, N., Mayer, T., Atzil-Slonim, D., Eliassaf, A., Haimovitz, S., & Gurevych, I. (2022). NLP meets psychotherapy: Using predicted client emotions and self-reported client emotions to measure emotional coherence. arXiv preprint arXiv:2211.12512.

Warmerdam, L., Smit, F., van Straten, A., Riper, H., & Cuijpers, P. (2010). Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: Randomized trial. *Journal of Medical Internet Research*, *12*(5), e53. https://doi.org/10.2196/jmir.1436

World Health Organization. ICD-11. (2020). Website cited. Retrieved February 2020, from https://www.who.int/classifications/icd/en/

World Health Organization. (2022). *World mental health report*. Transforming mental health for all.

Yeung, K. (2018). A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. *MSI*-AUT, *5*.

Yim, S. J., Lui, L. M. W., Lee, Y., Rosenblat, J. D., Ragguett, R. M., Park, C., Subramaniapillai, M., Cao, B., Zhou, A., Rong, C., Lin, K., Ho, R. C., Coles, A. S., Majeed, A., Wong, E. R., Phan, L., Nasri, F., & McIntyre, R. S. (2020). The utility of smartphone-based, ecological momentary assessment for depressive symptoms. *Journal of Affective Disorders*, *274*, 602–609. https://doi.org/10.1016/j.jad.2020.05.116