

# Key considerations for incorporating conversational AI in psychotherapy

Adam Miner<sup>1,2,4\*</sup>, Nigam Shah<sup>3</sup>, Kim D Bullock<sup>1</sup>, Bruce A Arnow<sup>1</sup>, Jeremy Bailenson<sup>4</sup>, Jeff Hancock<sup>4</sup>

<sup>1</sup> Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA.

<sup>2</sup> Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, USA.

<sup>3</sup> Stanford Center for Biomedical Informatics Research, Stanford University School of Medicine, Stanford, CA, USA.

<sup>4</sup> Department of Communication, Stanford University, Stanford, CA, USA.

## \*Correspondence:

Adam Miner Psy.D.  
aminer@stanford.edu

**Keywords:** Natural language processing, artificial intelligence, expert systems, psychotherapy, conversational AI, chatbot, digital assistant, human-computer interaction

## Abstract

Conversational artificial intelligence (AI) is changing the way mental health care is delivered. By gathering diagnostic information, facilitating treatment, and reviewing clinician behavior, conversational AI is poised to impact traditional approaches to delivering psychotherapy. While this transition is not disconnected from existing professional services, specific formulations of clinician-AI collaboration and migration paths between forms remain vague. In this viewpoint, we introduce four approaches to AI-human integration in mental health service delivery. To inform future research and policy, these four approaches are addressed through four dimensions of impact: access to care, quality, clinician-patient relationship, and patient self-disclosure and sharing. Although many research questions are yet to be investigated, we view safety, trust, and oversight as crucial first steps. If conversational AI isn't safe it should not be used, and if it isn't trusted, it won't be. In order to assess safety, trust, interfaces, procedures, and system level workflows, oversight and collaboration is needed between AI systems, patients, clinicians, and administrators.

## 1. Introduction

Clinicians engage in conversations with patients to establish a patient-therapist relationship (i.e. alliance), make diagnoses, and provide treatment. In traditional psychotherapy, this conversation typically involves a single patient and a single clinician (Goldfried et al., 1990). This model of psychotherapy is being modified because software programs that talk like people (i.e. conversational artificial intelligence, chatbots, digital assistants) are now beginning to provide mental health care (Miner et al., 2017). Conversational artificial intelligence (AI) is gathering diagnostic information (Bickmore et al., 2005; Rizzo et al., 2014) and delivering evidence-based psychological interventions (Bickmore et al., 2010; Fitzpatrick et al., 2017; Oh et al., 2017). Additionally, conversational AI is providing clinicians with feedback on their psychotherapy (Imel et al., 2015) and talking to young people about suicide, sex, and drug use (Crutzen et al., 2011; Martínez-Miranda, 2017).

Conversational AI appears unlikely to achieve enough technical sophistication to replace human therapists anytime soon. However, it does not need to pass the Turing Test (i.e. able to hold human seeming conversations) to have a significant impact on mental health care (Miner et al., 2017). A more proximal challenge is to plan and execute collaborative tasks between relatively simple AI systems and human practitioners (Bailenson et al., 2004; Luxton 2014; Hyman, 2018). Although AI in mental health has been discussed broadly (for a review see Luxton, 2014), specific formulations of clinician-AI collaboration and migration paths between forms remain vague.

Articulating different forms of collaboration is important, because the deployment of conversational AI into mental health diagnosis and treatment will be embedded within existing professional services. Conversational AI will likely interact with traditional workers (i.e. clinicians), but how these roles and responsibilities will be allocated between them has not been defined. To guide future research, we outline four approaches and dimensions of care that AI will affect.

Within the four approaches of AI-human integration in mental health service delivery, one extreme is a view that any involvement by conversational AI is unreasonable, putting both patients and providers at risk of harmful unintended consequences. At the other extreme, we explore how conversational AI might uniquely serve a patient's needs and surpass the capacity of even the most experienced and caring clinician by overcoming entrenched barriers to access. Although embodiment (e.g. virtual avatars or robots) can have a significant impact on interactions with virtual systems, we focus exclusively on the potential benefits and challenges of verbal and written language-based conversation and ignore the implications of embodiment or presence (Rehm et al., 2016). Table 1 summarizes the four approaches and our related assumptions.

## 2. Care delivery approaches

It is unclear whether the path forward will involve simultaneous experimentation with all four degrees of digitization, or progression through these approaches. We first briefly describe how these compare to the way individual psychotherapy is most often delivered today. Perhaps

surprisingly, laws, norms and the ethics of data sharing represent a nonobvious but critical factor in how these alternative approaches can operate now or develop in the future.

Currently, psychotherapy sessions are rarely recorded except in training institutions for supervision. When they are, for example during training or to assess clinician fidelity during clinical trials, trained human clinicians with prescribed roles and responsibilities are the listeners and provide oversight. With few exceptions, such as immediate risk of serious harm to the patient or others, clinicians need explicit permission to share identifiable patient information. When one of these exceptions is invoked, there is an obligation to limit the sharing strictly to the extent needed to provide effective treatment and ensure safety (American Psychiatric Association, 2001; American Psychological Association, 2002). Against this backdrop, having conversational AI listen to psychotherapy sessions or talk directly with patients represents a departure from established practice.

In the “humans only” approach, psychotherapy remains unchanged. Most psychotherapy sessions are heard only by the patient and clinician who are in the room. If a session were recorded, the labor intensiveness of human review would ensure most sessions would never be analyzed (Imel et al., 2015). The second approach, “human delivered, AI informed,” introduces into the room a listening device connected to software that detects clinically relevant information (Althoff et al., 2016) such as symptoms or interventions (Xiao et al., 2015), and relays this information back to the patient or clinician. Quantitative analysis of recorded psychotherapy is in its early stages, but it shifts to software programs the burden of extracting relevant information from audio or text. In the third approach, “AI delivered, human supervised,” patients speak directly to a conversational AI with the goal of establishing diagnoses or providing treatment (Lucas et al., 2014). A human clinician would either screen patients and hand off specific tasks to conversational AI or supervise conversations between front-line conversational AI and patients. The fourth approach, “AI only,” would have patients talk to a conversational AI with no expectation of supervision by a human clinician.

One of the less developed but more alluring ideas of AI psychotherapy is “AI delivered, human supervised.” Even the most ardent supporters of AI will acknowledge that there are certain things humans do better than computers. Combining people and algorithms may potentially build on the best of both approaches, and AI-human collaboration has been suggested as a way to address limitations in planning treatment in other medical areas such as oncology (Goldstein et al., 2017). Indeed, the prevailing opinion of expert systems researchers in the 1980s argued that computer-human collaboration would outperform either people or computers alone (for a review see Metaxiotis and Samouilidis, 2000).

In assessing any system to augment the practice of psychotherapy the first consideration of its impact should be that it will ensure patients and clinicians are helped and not harmed (Martinez-Martin et al., 2018; Roberts, 2016). In the discussion below, we consider salient issues that impact the potential value and harm of different delivery mechanisms by focusing on four dimensions of impact: access to care, quality, clinician-patient relationship, and patient self-disclosure.

### 3. Dimensions of impact

#### 3.1. Access to care

Limited access to mental health treatment creates a demand for scalable and non-consumable interventions (Kazdin and Rabbitt, 2013; Kazdin and Blaze, 2011). Despite the high costs and disease burden associated with mental illness (Dieleman et al., 2016), we have a decreasing number of clinicians per capita available to provide treatment in the US (Olfson, 2016). Increasing the number of human clinicians is not currently feasible, in part because of the decline from 2008 to 2013 per capita for both psychologists (from a ratio of 1:3,642 to 1:3,802) and psychiatrists (from a ratio of 1:7,825 to 1:8,476) (Olfson, 2016). Conversational AI has the potential to help address insufficient clinician availability because it is not inherently limited by human clinician time or attention. Conversational AI could also bridge one of the current tensions in care delivery: although clinicians value patient conversations, they have no financial incentive to engage in meaningful but lengthy conversations (Kaplan et al., 2016).

The decreasing amount of time spent in meaningful conversations exacerbates the shortage of psychiatrists and psychologists. Psychiatrists' use of talk therapy has been consistently and steadily declining, meaning fewer patients are receiving talk therapy during psychiatric visits (Mojtabai and Olfson, 2008). In contrast to a human clinician's time and attention, conversational AI is relatively non-consumable, making it an attractive alternative to delivery of care by a human. If conversational AI is effective and acceptable to both patients and clinicians, it may address longstanding challenges to mental health access. These include the ability to accommodate rural populations and to facilitate increased engagement from people who may experience traditional talk therapy as stigmatizing (Perle et al., 2011).

#### 3.2. Quality

Technology has been highlighted as a way to better understand and disseminate high quality psychotherapy (Mohr et al., 2014; Patel et al., 2018). Clinicians are already using texting services to deliver mental health interventions (Schuab et al., 2015), which demonstrates a willingness by patients and clinicians to test new approaches to patient-clinician interaction. These new approaches facilitate novel measures of intervention quality. For example, innovations in computer science (e.g. natural language processing and machine learning) are being used to assess language patterns of successful crisis interventions in text-based mental health conversations (Althoff et al., 2016; Dinaker et al., 2015). Computational analysis of psychotherapy is encouraging researchers and companies to identify patterns of patient symptomology and therapist intervention (Owen and Imel, 2016; Iter et al., 2018). This approach may improve psychotherapy quality by better understanding what effective clinicians actually do. This assessment has historically occurred through clinicians' self-reports or time intensive human audits (e.g. Cook et al., 2010).

Although its efficacy is not definitively established, there are reasons to expect that conversational AI could constructively enhance mental health diagnosis and treatment delivery (Haque et al., 2018; Laranjo et al., 2018). A diagnostic interview aids the patient and clinician in understanding the patient's presenting problem and provides a working model of how problems are being maintained. Approaches vary from highly structured diagnostic interviews (e.g. Structured Clinical Interview for DSM-5 (First et al., 2016) to unstructured interviews in which the conversation develops based on the clinician's expertise, training, and the patient's features.

Conversational AIs have interviewed patients about symptoms for PTSD with a high level of patient acceptance (Lucas et al., 2014). Conversational AI has been piloted across numerous clinically relevant groups such as clinical depression (Fitzpatrick et al., 2017) and adolescent stress (Huang et al., 2015). In a study in which students believed they were speaking with a conversational AI, the students reported feeling better after talking about their problems following the encounter (Ho et al., 2018). Although these early findings point to potential benefits, there is a lack of rigorous clinical trial data and uncertainty about regulatory oversight (Miner et al., 2017).

Yet while there is reason for optimism, inflated or unsubstantiated expectations may frustrate patients and weaken their trust in psychotherapeutic interventions (Aboujaoude, 2018; Miner et al., 2016). Many current computation methods can be used to search for specific dialogue acts, but additional work is needed to map theoretically important constructs (e.g. therapeutic alliance) to assess causal relationships between language patterns and clinically relevant outcomes. Psychotherapy quality will be difficult to assess without disentangling causal inferences and confounding factors. Beyond computation, patients' attitudes matter in psychotherapy because those who have a negative experience compared with their expectations have worse clinical outcomes (Watsford and Rickwood, 2013). If a patient loses trust in a conversational AI, they may be less likely to trust human clinicians as well. As conversational AI becomes more sophisticated and expectations of benefit increase, there are growing concerns that users will transition from feeling let down to feeling betrayed (Brooker, 2018). These factors suggest that careful experimentation about sub-processes in AI-mediated communication merits research attention.

### **3.3. Clinician-patient relationship**

Modern medicine views the patient-clinician relationship as critical to patient health (Martin et al., 2000), and provider wellness (Rosenthal and Verghese, 2016). Indeed, appreciation of the importance of the patient-clinician relationship in modern medicine can be traced back to the influence of clinical psychology (Szasz and Hollender, 1956). Therapeutic alliance develops from clinicians' collaborative engagement with patients and reflects agreement on treatment goals, the tasks necessary to achieve such goals, and the affective bond between patient and provider (Horvath and Greenberg, 1989). Therapeutic alliance is consistently associated with symptom improvement in psychotherapy (Flückiger et al., 2012; Horvath et al., 2011; Norcross, 2002). Numerous approaches exist to create alliance during psychotherapy, including the use of supportive language, mirroring emotions, and projecting warmth. Although originally conceptualized for human-to-human conversations, users have reported experiencing a sense of therapeutic alliance when speaking directly with conversational AI, suggesting this bond may not necessarily be restricted to human-human relationships (Bickmore et al., 2005). If conversational AI can create and maintain a therapeutic alliance, the provision of psychotherapy will not be necessarily limited by human clinicians' time and attention.

Establishing therapeutic alliance with conversational AI may benefit both patients and providers. By allowing conversational AI to take over repetitive, time-consuming tasks, clinicians' attention and skill could be deployed more judiciously (Jha and Topol, 2016). Allowing clinicians to do less of the work that contributes to burnout, such as repetitive tasks performed with little autonomy, may improve clinicians' job satisfaction (Harvey et al., 2017).

Clinician burnout is associated with worse patient outcomes and is increasingly recognized as a problem which must be more adequately addressed (Delgadillo et al., 2018, Panagioti et al., 2017).

At the same time, software that augments clinical duties has been criticized for distancing clinicians from patient care (Verghese, 2008). In mental health, this risk is especially salient because the content of therapy is often quite intimate. Some of the repetitive, time-consuming tasks clinicians engage in with patients, such as reviewing symptoms or taking their history, are precisely the vehicles by which clinicians connect with and understand their patients' experiences and develop rapport. It is unknown whether having a conversational AI listen in on psychotherapy will significantly impact patients' and clinicians' sense of therapeutic alliance. This area merits further research.

### **3.4. Patient self-disclosure and sharing**

Patient self-disclosure of personal information is crucial for successful therapy, including sensitive topics such as trauma, substance use, sexual history, forensic history, and thoughts of self-harm. Patient self-disclosures during psychotherapy are legally and ethically protected (Roberts, 2016) and professional norms and laws have been established to set boundaries for what a clinician can share (Edwards, 2014). Unauthorized sharing of identifiable patient information can result in fines, loss of license, or even incarceration. Moreover, because of the natural limitations of human memory, patients are unlikely to expect a human clinician to remember entire conversations perfectly in perpetuity. This capacity is in stark contrast to conversational AI, which has near-limitless capacity to hear, remember, share, and analyze conversations as long as desired. Because humans and machines have such different capacities, patient expectations of AI capabilities may impact treatment decisions and consent to data sharing (Martinez-Martin et al., 2018).

In mental health, conversational AI has been shown to both facilitate and impede disclosure in different contexts. For example, users were more open with a conversational AI than with a human listener in reporting mental health symptoms (Lucas et al., 2014), and have been successfully used to treat persecutory delusions for people with psychosis (Craig et al., 2018). Conversely, users were more reluctant to disclose sensitive information such as binge drinking behavior to a conversational AI compared to a non-responsive questionnaire (Schuetzler et al., 2018). Because personal disclosures are central to diagnosis and treatment in psychotherapy, users' expectations and behavior towards technology-mediated conversations merit further assessment (Bickmore et al., 2018; French and Bazarova, 2017; Liu and Sundar; 2018).

Certain disclosures in a psychotherapy context carry specific ethical and legal mandates, such as reporting suicidal or homicidal ideation. In 1969, a therapist at the University of California did not share the homicidal ideation of a patient with the intended victim. The patient subsequently killed the named victim, and the victim's family sued. This case (*Tarasoff v. Regents of the University of California*, 1974) established clinicians' duty not only to protect the confidentiality of their patients but also to notify individuals their patient might harm. A failure to warn leaves a clinician liable to civil judgment (Swerdlow, 2018). Most case law and norms have been established on the premise of a dyadic relationship between patient and clinician. The

extent to which conversational AI inherits liability for harm is untested. As conversational AI takes on clinical duties and informs clinical judgment, expectations must be clarified about how and when these systems will respond to issues related to confidentiality, safety, and liability.

#### 4. Discussion

Experts in AI, clinicians, administrators, and other stakeholders recognize a need to more fully consider safety and trust in the design and deployment of new AI-based technologies (Bhugra et al., 2017; Stone et al., 2016). A recent Lancet commission on global mental health states that “technology-based approaches might improve the reach of mental health services but could lose key human ingredients and, possibly, lower effectiveness of mental health care” (Patel et al., 2018). To inform future research directions, we have presented four approaches to integrating conversational AI into mental health delivery and discussed the dimensions of their impact.

Because conversational AI may augment the work of psychotherapy, we seek to encourage product designers, clinicians, and researchers to assess the impact of new practices on both patients and clinicians. Other areas of medicine have seen success with AI, such as lung cancer imaging and building diagnostic or prognostic models (Avati et al., 2017; Jung et al., 2016; Gulshan et al., 2016; Pusiol et al., 2016; Yu et al., 2016), and conversational AI for health is an emerging field with limited research on efficacy and safety (Bai et al., 2017; Bickmore et al., 2018; Laranjo et al., 2018; Miner et al., 2016).

Before we deploy AI-mediated treatment, workflow changes must be considered in the context of other demands on clinician time and training. Clinicians are already being asked to be familiar with telehealth (Maheu et al., 2017) social media (Zalpuri et al., 2018), and mobile health (Hilty et al., 2019), while simultaneously being reminded of the need for self-care in light of clinician burnout (Panagioti et al., 2017). Before we insert new devices into clinical care, it will be crucial to engage clinicians and design evaluation strategies that appreciate the skills, attitudes, and knowledge of affected workers. Just as we can’t expect technology companies to easily understand healthcare, we can’t expect medical professionals to intuit or work in harmony with new technology without thoughtful design and training.

A limitation of this work is that we do not set out a specific research agenda, and some important considerations are beyond the scope of this work (e.g. the cost and feasibility of each approach). We propose instead that initiatives using conversational AI anticipate challenges and leverage lessons learned from existing approaches to deploying new technology in clinical settings that involve clinician training and patient protections from the start (Hilty et al., 2019, Mohr et al., 2014). We instead encourage those proposing to put AI into care settings to directly consider and measure impact on access, quality, relationships, and data sharing.

The potential benefits are clear for mental health. If diagnosis or treatment can be done by conversational AI, the societal burden of treating mental health could be diminished. Additionally, conversational AI could have a more long-term relationship with a patient than clinicians who rotate out of training centers. Despite these potential benefits, technology carries risks related to privacy, bias, coercion, liability, and data sharing that could harm patients in

expected (e.g. denial of health insurance) and unintended ways (Aboujaoude, 2018; Bai et al., 2017; Bickmore et al, 2018; Caliskan et al., 2017; De Choudhury et al., 2017; Martinez-Martin and Kreitmair, 2018; Patel et al., 2018). Conversations are valuable for patients and clinicians, and it is crucial to make sure they are delivered safely and effectively, regardless of who or what does the talking.



**Acknowledgements**

This work was supported by grants from the National Institutes of Health, National Center for Advancing Translational Science, Clinical and Translational Science Award (KL2TR001083 and UL1TR001085), the Stanford Department of Psychiatry Innovator Grant Program, and the Stanford Human-Centered AI Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We thank Nicole Martinez-Martin JD PhD, Victor Henderson MD MS, and Stan Fisher for their valuable feedback. Reference formatting assisted by Charlesworth Author Services.

**Conflict of interests**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Author contributions**

ASM and JH contributed to the initial conceptualization and design of the manuscript. ASM wrote the first draft; all authors contributed to manuscript revision, read and approved the submitted version.

**Data availability**

No datasets were generated for this study.

**Table 1.** Delivery approaches and dimensions of impact for conversational AI

Care delivery approach	Dimensions of impact			
	Access to care	Quality	Clinician-patient relationship	Patient self-disclosure and sharing
<b>Humans only</b>	Unchanged	Established	No disruption	Unchanged
<b>Human delivered, AI informed</b>	Unchanged	Potentially improved	Potentially disrupted <sup>a</sup>	Unknown
<b>AI delivered, human supervised</b>	Improved, but limited scalability	Unknown	Likely disrupted	Unknown
<b>AI only</b>	Improved, not restrained by human attention	Unknown	Nonexistent	Unknown

a: By ‘disrupt’ we do not mean to signal that the result will be necessarily good or bad.

## References

- Aboujaoude, E. (2018). Telemental health: Why the revolution has not arrived. *World Psychiatry*; 17(3):277.
- 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.
- American Psychiatric Association. (2001). *The Principles of Medical Ethics: With Annotations Especially Applicable to Psychiatry*. Amer Psychiatric Pub Incorporated.
- American Psychological Association. (2002). Ethical Principles of Psychologists and Code of Conduct. *American psychologist*; 57(12):1060-1073.
- Avati, A., Jung, K., Harman, S., Downing, L., Ng, A., Shah, NH. (2017). Improving palliative care with deep learning. *IEEE International Conference on Bioinformatics and Biomedicine*; Kansas City, MO. pp. 311-316. IEEE.
- Bai, G., Jiang, J.X., Flasher, R. (2017). Hospital risk of data breaches. *JAMA Internal Medicine*; 177(6):878-80.
- Bailenson, J.N., Beall, A.C., Loomis, J., Blascovich, J., Turk, M. (2004). Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence: Teleoperators & Virtual Environments*; 13(4):428-441.
- Bickmore, T., Gruber, A., Picard, R. (2005). Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*; 59(1):21-30.
- Bickmore, T.W., Puskar, K., Schlenk, E.A., Pfeifer, L.M., Sereika, S.M. (2010). Maintaining reality: Relational agents for antipsychotic medication adherence. *Interacting with Computers*; 22(4):276-288.
- Bickmore, T., Trinh, H., Asadi, R., Olafsson, S. (2018). Safety first: Conversational agents for health care. In: Moore, R., Szymanski, M., Arar, R., Ren, G.J. (eds) *Studies in Conversational UX Design*. Human-Computer Interaction Series. Springer, Cham.
- Bickmore, T.W., Trinh, H., Olafsson, S., O'Leary, T.K., Asadi, R., Rickles, N.M., et al. (2018). Patient and consumer safety risks when using conversational assistants for medical information: An observational study of Siri, Alexa, and google assistant *J Med Internet Res*; 20(9):e11510.
- Bhugra, D., Tasman, A., Pathare, A., et al. (2017). The WPA-lancet psychiatry commission on the future of psychiatry. *The Lancet Psychiatry*; 4(10):775-818.

- Brooker, N. "We Should Be Nicer to Alexa." *Financial Times*.  
<https://www.ft.com/content/4399371e-bcbd-11e8-8274-55b72926558f> (Accessed October 15, 2018).
- Caliskan, A., Bryson, J.J., Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*; 356(6334):183-186.
- Cook, J.M., Biyanova, T., Elhai, J., Schnurr, P.P., Coyne, J.C. (2010). What do psychotherapists really do in practice? An Internet study of over 2,000 practitioners. *Psychotherapy: Theory, Research, Practice, Training*; 47(2):260.
- Craig, T.K., Rus-Calafell, M., Ward, T., Leff, J.P., Huckvale, M., Howarth, E., Emsley, R., Garety, P.A. (2018). AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. *The Lancet Psychiatry*; 5(1):31-40.
- Crutzen, R., Peters, G.J.Y., Portugal, S.D., Fisser, E.M., Grolleman, J.J. (2011). An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. *Journal of Adolescent Health*; 48(5):514-519.
- De Choudhury, M., Sharma, S.S., Logar, T., Eekhout, W., Nielsen, R.C. (2017). Gender and cross-cultural differences in social media disclosures of mental illness. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*; Portland, OR. pp. 353-369.
- Delgadillo, J., Saxon, D., Barkham, M. (2018). Associations between therapists' occupational burnout and their patients' depression and anxiety treatment outcomes. *Depress Anxiety*; 35:844–850. <https://doi.org/10.1002/da.22766>
- Dieleman, J.L., Baral, R., Birger, M., Bui, A.L., Bulchis, A., Chapin, A., et al. (2016). US spending on personal health care and public health, 1996-2013; *JAMA*; 316(24):2627-2646.
- Dinakar, K., Chen, J., Lieberman, H., Picard, R., Filbin, R. (2015). Mixed-initiative real-time topic modeling & visualization for crisis counseling. *Proceedings of the 20th International Conference on Intelligent User Interfaces*; Atlanta GA. pp. 417-426. ACM.
- Edwards, G. (2014). Doing their duty: an empirical analysis of the unintended effect of Tarasoff v. Regents on homicidal activity. *The Journal of Law and Economics*; 57(2):321-348.

First, M.B., Williams, J.B.W., Karg, R.S., Spitzer, R.L. (2016). Structured clinical interview for DSM-5 disorders, clinician version (SCID-5-CV). Arlington, VA: American Psychiatric Association.

Fitzpatrick, K.K., Darcy, A., Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*; 4(2).

Flückiger, C., Del Re, A.C., Wampold, B.E., Symonds, D. and Horvath, A.O., (2012). How central is the alliance in psychotherapy? A multilevel longitudinal meta-analysis. *Journal of Counseling Psychology*, 59(1), p.10.

French, M., Bazarova, N.N. (2017). Is anybody out there?: Understanding masspersonal communication through expectations for response across social media platforms. *Journal of Computer-Mediated Communication*; 22(6):303-19.

Goldfried, M.R., Greenberg, L.S., Marmar, C. (1990). Individual psychotherapy: Process and outcome. *Annual Review of Psychology*; 41(1):659-688.

Goldstein, I.M., Lawrence, J., Miner, A.S. (2017). Human-machine collaboration in cancer and beyond: the centaur care model. *JAMA Oncology*; 3(10):1303-4.

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*; 316(22):2402-2410.

Haque A, Guo M, Miner AS, L Fei-Fei. (2018). Measuring Depression Symptom Severity from Spoken Language and 3D Facial Expressions. Paper presented at NeurIPS 2018 Workshop on Machine Learning for Health. Montreal, Canada.

Harvey, S.B., Modini, M., Joyce, S., Milligan-Saville, J.S., Tan, L., Mykletun, A., et al. (2017). Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. *Occup Environ Med*; 74(4):301-10.

Hilty, D. M., Chan, S., Torous, J., Luo, J., & Boland, R. J. (2019). A Telehealth Framework for Mobile Health, Smartphones, and Apps: Competencies, Training, and Faculty Development. *Journal of Technology in Behavioral Science*; 1-18.

Ho, A., Hancock, J., Miner, A.S. (2018). Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations with a Chatbot. *Journal of Communication*; 68(4):712-33. <https://doi.org/10.1093/joc/jqy026>

Horvath, A.O., Del Re, A.C., Flückiger, C. & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy*; 48(1), p.9.

Horvath, A.O., Greenberg, L.S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*; 36(2):223.

Huang, J., Li, Q., Xue, Y., Cheng, T., Xu, S., Jia, J., Feng, L. (2015). Teenchat: a chatterbot system for sensing and releasing adolescents' stress. *International Conference on Health Information Science*; Queensland, Australia. pp. 133-145. Springer, Cham.

Hyman, L. (2018). *Temp: How American Work, American Business, and the American Dream Became Temporary*. New York, New York. Penguin Random House. ISBN: 9780735224070.

Imel, Z.E., Steyvers, M., Atkins, D.C. (2015). Computational psychotherapy research: Scaling up the evaluation of patient-provider interactions. *Psychotherapy*; 52(1):19.

Iter, D., Yoon, J., Jurafsky, D. (2018). Automatic detection of Incoherent speech for diagnosing schizophrenia. *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*; New Orleans, LA. pp.136-146.

Jha, S., Topol, E.J. (2016). Adapting to artificial intelligence: radiologists and pathologists as information specialists. *JAMA*; 316(22):2353-4.

Jung, K., Covington, S., Sen, C.K., Januszyk, M., Kirsner, R.S., Gurtner, G.C. et al. (2016). Rapid identification of slow healing wounds. *Wound Repair and Regeneration*; 24(1):181-8.

Kaplan, R.S., Haas, D.A., Warsh, J. (2016). Adding value by talking more. *New England Journal of Medicine*; 17;375(20):1918-20.

Kazdin, A.E., Rabbitt, S.M. (2013). Novel models for delivering mental health services and reducing the burdens of mental illness. *Clinical Psychological Science*; 1(2):170-191.

Kazdin, A.E., Blase, S.L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on Psychological Science*; 6(1):21-37.

Laranjo, L., Dunn, A.G., Tong, H.L., Kocaballi, A.B., Chen, J., Bashir, R., et al. (2018). Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*; 25(9):1248-1258.

Liu, B., Sundar, S.S. (2018). Should machines express sympathy and empathy? experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking*; 21(10):625-636.

Lucas, G.M., Gratch, J., King, A., Morency, L.P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*; 1;37:94-100.

Luxton, D.D. (2014). Recommendations for the ethical use and design of artificial intelligent care providers. *Artificial Intelligence in Medicine*; Sep 1;62(1):1-0.

Luxton, D.D., editor. (2016). *Artificial Intelligence in Behavioral and Mental Health Care*. Elsevier/Academic Press. ISBN: 978-0-12-420248-1.

Maheu, M. M., Drude, K. P., Hertlein, K. M., Lipschutz, R., Wall, K., & Hilty, D. M. (2017). An interprofessional framework for telebehavioral health competencies. *Journal of Technology in Behavioral Science*; 2(3-4), 190-210.

Martin, D.J., Garske, J.P., Davis, M.K. (2000). Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *Journal of Consulting and Clinical Psychology*; 68(3):438.

Martinez-Martin, N., Kreitmair, K. (2018). Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. *JMIR Mental Health*; 5(2).

Martinez-Martin, N., Dunn, L. B., & Roberts, L. W. (2018). Is it ethical to use prognostic estimates from machine learning to treat psychosis? *AMA Journal of Ethics*; 20(9):804-811.

Martínez-Miranda, J. (2017). Embodied conversational agents for the detection and prevention of suicidal behaviour: Current applications and open challenges. *Journal of Medical Systems*; 41(9):135.

Metaxiotis, K.S., Samouilidis, J.E. (2000). Expert systems in medicine: academic illusion or real power? *Information Management & Computer Security*; 1;8(2):75-9.  
<https://doi.org/10.1108/09685220010694017>

Miner, A.S., Milstein, A., Hancock, J.T. (2017). Talking to machines about personal mental health problems; *JAMA*; 318(13):1217-8.

Miner, A.S., Milstein, A., Schueller, S., Hegde, R., Mangurian, C., Linos, E. (2016). Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Internal Medicine*; 176(5):619-25.

- Mohr, D.C., Schueller, S.M., Montague, E., Burns, M.N., Rashidi, P. (2014). The behavioral intervention technology model: An integrated conceptual and technological framework for eHealth and mHealth interventions. *Journal of Medical Internet Research*; 16(6).
- Mojtabai, R., Olfson, M. (2008). National trends in psychotherapy by office-based psychiatrists. *Archives of General Psychiatry*; 65(8):962-70.
- Norcross, J. C. (Ed.). (2002). Psychotherapy relationships that work: Therapist contributions and responsiveness to patient needs. New York: Oxford University Press.
- Oh, K.J., Lee, D., Ko, B., Choi, H.J. (2017). A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation. *Mobile Data Management (MDM), 2017 18th IEEE International Conference*; 29. pp. 371-375. IEEE.
- Olfson, M. (2016). Building the mental health workforce capacity needed to treat adults with serious mental illnesses. *Health Affairs*; 35(6):983-90.
- Owen, J., Imel, Z.E. (2016). Introduction to the special section “Big ‘er’Data”: Scaling up psychotherapy research in counseling psychology. *Journal of Counseling Psychology*; 63(3):247.
- Panagioti, M., Panagopoulou, E., Bower, P., Lewith, G., Kontopantelis, E., Chew-Graham, C., et al. (2017). Controlled interventions to reduce burnout in physicians: a systematic review and meta-analysis. *JAMA Internal Medicine*; 177(2):195-205.
- Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., et al. (2018). The Lancet Commission on global mental health and sustainable development; *The Lancet*; 392(10157):1553-1598.
- Perle, J.G., Langsam, L.C., Nierenberg, B. (2011). Controversy clarified: An updated review of clinical psychology and tele-health. *Clinical Psychology Review*; 31(8):1247-58.
- Pusiol, G., Esteva, A., Hall, S.S., Frank, M., Milstein, A., Fei-Fei, L. (2016). Vision-based classification of developmental disorders using eye-movements. *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Athens, Greece. pp. 317-325. Springer, Cham.
- Rehm, I.C., Foenander, E., Wallace, K., Abbott, J.A., Kyrios, M., Thomas, N. (2016). What role can avatars play in e-mental health interventions? Exploring new models of client-therapist interaction. *Frontiers in Psychiatry*; 7:186.



Rizzo, A., Scherer, S., DeVault, D. Gratch, J., Artstein, R., Hartholt, A., et al. (2014). Detection and computational analysis of psychological signals using a virtual human interviewing agent. *10<sup>th</sup> Intl Conf Disability, Virtual Reality & Associated Technologies*; Gothenburg, Sweden. Available at: <http://www.icdvrat.org/2014/index2014.htm> (Accessed 15 Oct. 2018).

Roberts, L.W. (2016). *A clinical guide to psychiatric ethics*. Arlington, VA: American Psychiatric Pub; ISBN: 978-1-61537-049-8.

Rosenthal, D.I., Verghese, A. (2016). Meaning and the nature of physicians' work. *New England Journal of Medicine*; 375(19):1813-1815.

Schaub, M.P., Wenger, A., Berg, O., Beck, T., Stark, L., Buehler, E., Haug, S. (2015). A web-based self-help intervention with and without chat counseling to reduce cannabis use in problematic cannabis users: Three-arm randomized controlled trial. *Journal of Medical Internet Research*; 17(10).

Schuetzler, R.M., Giboney, J.S., Grimes, G.M., Nunamaker, J.F. (2018). The influence of conversational agents on socially desirable responding. *Proceedings of the 51st Hawaii International Conference on System Sciences*; Waikoloa Village, Hawaii. pp. 283-292. ISBN: 978-0-9981331-1-9.

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R. Etzioni, O. Hager, G, et al. (2016). "Artificial Intelligence and Life in 2030." One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University, Stanford, CA, Doc: <http://ai100.stanford.edu/2016-report>. (accessed October 15, 2018).

Swerdlow, B.A. (2018). Tracing the Evolution of the Tarasoff Duty in California. *J Sociology & Social Welfare*; 45:25.

Szasz, T.S., Hollender, M.H. (1956). A contribution to the philosophy of medicine: the basic models of the doctor-patient relationship. *AMA Archives of Internal Medicine*; 97(5):585-592.

Verghese, A. (2008). Culture shock-patient as icon, icon as patient. *New England Journal of Medicine*; 359(26):2748-2751.

Watsford, C., Rickwood, D. (2013). Disconfirmed expectations of therapy and young people's clinical outcome, help-seeking intentions, and mental health service use. *Advances in Mental Health*; 12(1):75-86.

Xiao, B., Imel, Z.E., Georgiou, P.G., Atkins, D.C., Narayanan, S.S. (2015). "Rate my therapist": Automated detection of empathy in drug and alcohol counseling via speech and language processing. *PloS One*; 10(12):e0143055.

Yu, K.H., Zhang, C., Berry, G.J., Altman, R.B., Ré, C., Rubin, D.L., Snyder, M. (2016). Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nature Communications*; 7:12474.

Zalpuri I., Liu H.Y., Stubbe D., Wrzosek M., Sadhu J., Hilty D. (2018). Social media and networking competencies for psychiatric education: skills, teaching methods, and implications. *Academic Psychiatry*; 42(6):808-17.