Students’ attitudes towards AI in Psychiatry

Anne, Susanne, & Eesha

# Seed for random number generation  
set.seed(42)

# Introduction

The increasing number of people suffering from mental health conditions is a substantial concern for the world’s population. Psychotherapy is effective in tackling mental health problems. However, due to the high demand, psychotherapeutic interventions are not always available when needed. As a result, individuals suffering from severe mental health problems may remain untreated because psychotherapists cannot take on more patients. In addition, over the past decades, little has changed concerning how psychotherapy is delivered (Johnsen & Friborg, 2015). A lack of targeted feedback, care quality monitoring, and stagnation in terms of further education and training of (prospective) psychotherapists may hinder the implementation of effective interventions (Cummins et al., 2019; Hirsch et al., 2018; Schwalbe et al., 2014).

Specific artificial intelligence (AI) technologies for mental health care may address these problems. Intelligent systems have been designed to aid in diagnosing mental health diseases or recognizing the severity of a mental health condition to ensure timely and optimal care for severely affected individuals Stasak & Epps (2017). In their scoping review of machine learning in psychotherapy research, Aafjes-van Doorn et al. (2021) identified 51 studies that developed and tested machine learning algorithms focused on selecting appropriate treatment regimes and predicting treatment adherence, as well as therapist skill improvement. In addition, AI-enabled tools may improve the quality of psychotherapeutic training and education. To this end, AI systems analyze data gathered from therapist-patient conversations to provide performance-specific feedback for the therapist, thus potentially enhancing their motivational interviewing performance Imel et al. (2019).

Despite a large amount of academic knowledge and efforts to develop user-friendly applications, AI systems are still hardly utilized in clinical care (Sendak et al., 2020). Technical and administrative difficulties, such as the inaccuracy of predictions, accuracy-interpretability trade-offs, and data privacy and security concerns, complicate the widespread implementation of AI tools in diagnosis, prognosis, and the selection of treatment approaches (Aafjes-van Doorn et al., 2021; Chekroud et al., 2021; Chen et al., 2022; Kelly et al., 2019; Lee et al., 2021; Roth et al., 2021).

Next to general reasons for reluctance against using AI-enabled tools in mental healthcare, specific applications may be associated with specific forms of skepticism. For example, based on the belief that an algorithm may not judge human-provided care, psychotherapists may be hesitant to accept feedback regarding their therapeutic practice from a machine. Similarly, they may be reluctant to use an algorithm’s assessment of a patient’s depression level to schedule timely interventions for severely affected individuals. Much skepticism against AI-enabled tools in mental healthcare may be associated with a lack of understanding of how AI recommendations are generated. Despite attempts to enhance the explainability of AI, such as the Explainable Artificial Intelligence (XAI) Initiative, the complexity of deep learning approaches necessarily limits the extent to which they can be made accessible to a broader user group (Feldman et al., 2019). Especially in mental health care, where transparency and the explainability of clinical decision-making are highly valued, the black box problem of AI-based recommendations creates a significant obstacle to its adoption (Aafjes-van Doorn et al., 2021; Chekroud et al., 2021; Kelly et al., 2019).

The unified theory of acceptance and use of technology [UTAUT; Venkatesh (2022), Venkatesh et al. (2003), Venkatesh et al. (2016)] provides a theoretical framework that explains the relationship between these obstacles and the intention to use AI tools. The UTAUT includes four main predictors of the intention to use a specific technology: a) performance expectancy, defined as the degree to which an individual believes that using a system will enhance their performance, b) effort expectancy, as the degree of ease associated with using the technology, c) social influence, referring to the perception that important others believe that the system should be used, and d) facilitating conditions, as the belief that the infrastructure exists to support the use of the system. Next to these general predictors of the intention to use an AI tool, trust in the technology plays an important role for applications addressing sensitive matters, such as treatment recommendations or therapist feedback reports. Accordingly, we include trust in the technology

In addition, the UTAUT proposes context (e.g., location), user (e.g., age), and technology attributes (e.g., mode of delivery) as moderators of these relationships [Venkatesh et al. (2016); venkatesh\_etal03]. The UTAUT is widely used in technology acceptance research and has been utilized to explain medical staff and students’ attitudes towards AI technology (Fan et al., 2020; Tamori et al., 2022; Tran et al., 2021; Zhai et al., 2021). However, to our knowledge, only one study has investigated the predictors of technology acceptance and use in the mental health domain (Gado et al., 2022). Specifically, Gado et al. (2022) found that general perceived usefulness and perceived ease of use positively predicted intention to use AI tools. The effect was mediated through favorable attitudes towards AI. In addition, the authors identified knowledge about AI tools as a direct predictor of the intention to use the tool.

Based on previous research findings, the first goal of the current research is to test the applicability of the UTAUT in the mental health context to understand the factors that influence the willingness to accept AI-enabled recommendations (Gado et al., 2022; Venkatesh, 2022; Venkatesh et al., 2016; Venkatesh et al., 2003). In Study 1, we investigate the relevance of UTAUT predictors for the intention to use AI tools among samples of psychology students specialized in clinical psychology. In contrast to practicing psychotherapists, psychology students are less influenced by habits and established work processes that may hinder the adoption of new AI technologies (Venkatesh et al., 2016). Accordingly, several opportunities arise to improve students’ acceptance of AI technology (e.g., at university or during psychotherapy training).  
Psychotherapy includes multiple tasks, such as diagnosis, crisis intervention, selecting appropriate long-term treatment plans, and ensuring high-quality care. The acceptance of AI-enabled technology may differ between different psychotherapeutic tasks. For example, students with a positive attitude toward AI tools that provide targeted feedback based on the analysis of patient-practitioner interactions may still be skeptical about relying on AI support systems to assess patients’ disease severity. Accordingly, in Study 1, we extend previous research findings by investigating the predictors of the intention to use two different AI-enabled mental health tools already available to mental health practitioners. The first tool is a speech-based diagnostic device used to detect the severity of a mental health condition to deliver timely care to severely affected individuals [similar to the Sonde Health smartphone speech elicitation app; Huang et al. (2018)]. The second tool analyses therapeutic conversations between practitioner and patient to deliver targeted feedback to psychotherapists based on the principles of motivational interviewing [see the Therapy Insights Model (TIM); Cummins et al. (2019)].

Based on the concept of technology knowledge, skills, and attitudes [KSA; Seufert et al. (2021)], Gado et al. (2022) argue that students’ attitudes towards AI may be influenced by their knowledge and understanding of the technology itself. A basic understanding of how the AI recommendations are derived may leverage some ethical concerns and strengthen students’ competence in using the tool, thus potentially increasing their acceptance of the tool (Gado et al., 2022; Seufert et al., 2021). Accordingly, the second goal of the current research is to examine the effectiveness of a skill-based intervention on the intention to use the two previously described AI tools. Thus, based on the findings of Study 1, in Study 2, we test the effects of a skill-based intervention on students’ intention to use the two AI tools described above. By examining the influence of knowledge about AI tools in an experimental setting, we provide a detailed test of the knowledge hypothesis proposed by Gado et al. (2022).

# Theory Development

## Applications of AI in Psychotherapy Practice

## The Application of AI Tools to Assess the Severity of Mental Health Conditions

## The Application of AI Tools to Improve Psychotherapy Quality

Supervision and receiving performance feedback on their therapy sessions support psychotherapy trainees’ skills acquisition and increase retention (**moyers\_etal05?**). However, providing ongoing feedback is labor and cost intensive and thus rarely used in training and clinical practice. Accordingly, feedback is often based on trainees’ self-reports and is usually only available long after the session (Tanana et al., 2019b). Using AI technology for training purposes in mental health care may lhelp to reduce this problem by providing continuous, immediate, and performance-specific feedback to psychotherapists and trainees.

Most tools developed to improve psychotherapy quality rely on natural language processing-based feedback Tanana et al. (2019b). For example, *TIM* (Therapy Insights Model) uses real-time chat messages exchanged between therapists and patients to provide therapists with feedback regarding the topics that were sufficiently covered during the session and the topics that should be addressed in the following sessions (Cummins et al., 2019). *CORE-MI* (Counselor Observer Ratings Expert for Motivational Interviewing) uses audio recordings of motivational interviewing (MI)[[1]](#footnote-23) sessions to generate feedback on psychotherapists’ adherence to MI principles. The user receives feedback on six summary measures of MI fidelity: empathy, MI spirit, reflection-to-question ratio, percent open questions, percent complex reflections, and percent MI adherence. *CORE-MI* includes a visual summary of counseling sessions based on the fidelity assessment that the therapist may use to improve their MI performance (Hirsch et al., 2018). Similar tools include the *ClientBot*, a training tool that mimics typical patient responses to therapist questions and provides real-time feedback on therapists’ use of open questions and reflections (Tanana et al., 2019b); or *Partner*, a reinforcement learning agent that may increase the quality of mental health support conversations by suggesting sentence-level edits to posts that enhance the level of empathy while maintaining conversation quality (Sharma et al., 2021).

## The Unified Theory of Acceptance and Use of Technology (UTAUT) as a Theoretical Framework

## The Role of Specific Knowledge

# Study 1

## Methods

### Participants

(Pinto dos Santos et al., 2019): “The questionnaire was sent out via email and advertised on social media to undergraduate medical students at three major German universities (University of Cologne, University of Bonn, University of Mainz). Participation was voluntary and had no relation to the student’s curricular activities. The students were informed that the survey results would be used for further statistical evaluation and scientific publication. Respondent anonymity was guaranteed by design […].”

### Measurement Instruments

#### AI tools

##### Severity Detection Tool

|  |
| --- |
| Sonde Health voice detection |

##### Therapist Feedback Tool

|  |
| --- |
| Core MI Feedback Tool |

#### Independent Variables Based on UTAUT

* Perceived social norm Gado et al. (2022); all slightly adapted:
  + “I believe that when I work as a psychotherapist, people who will influence my professional behavior think that I should use [short tool description].”
  + “I believe that when I work as a psychotherapist, people who will be important to me think that I should use [short tool description].” (slightly adapted)
  + “I believe that when I work as a psychotherapist, my supervisors will help me in the use of [short tool description].” (slightly adapted)
  + “I believe that when I work as a psychotherapist, the institution I work at will support the use of [short tool description].” (slightly adapted)
* Performance expectancy (perceived usefulness) Gado et al. (2022); all slightly adapted:
  + “[short tool description] may be useful in my future job.”
  + “Using [short tool description] may enable me to accomplish tasks more quickly.”
  + “Using [short tool description] may increase my productivity as a psychotherapist.”
  + EXCLUDE: “If I use [short tool description], I will increase my chances of getting a raise.”
  + ADD: “Using [short tool description] may improve the quality of my care.”
  + CONTACT Gado et al. (2022) and ask for items!
* Effort expectancy (perceived ease of use) Gado et al. (2022); all slightly adapted:
  + “How to use [short tool description] would be clear and understandable.”
  + “It would be easy for me to become skillful at using [short tool description].”
  + “I would find [short tool description] easy to use.”
  + “Learning to operate [short tool description] would be easy for me.”

Facilitating conditions (Venkatesh et al., 2003); all slightly adapted: - “I believe that I will have the resource necessary to use [short tool description].” - “I believe that I will have the knowledge necessary to use [short tool description].” - “I believe that [short tool description] will be compatible with other tools I will use as a psychotherapist.” (was originally reverse coded, but I would frame this positively for methodological reasons) - “I believe that a specific person (or group) will be available for assistance with difficulties with the use of [short tool description].”

* Voluntariness of use (Venkatesh et al., 2003); all slightly adapted:
  + “Although it might be helpful, using [short tool description] will certainly not be compulsory in my job.”
  + “I believe that my boss or supervisor will not require me to use [short tool description].”
  + “I believe that my boss or supervisor will not expect me to use [short tool description].” (was originally reverse coded, but I would frame this positively for methodological reasons)
  + “My use of [short tool description] would be voluntary (as opposed to required by superiors/job).”
* Knowledge of the tool (Gado et al., 2022):
  + “Please rate your understanding of how the recommendations delivered by [short tool description] are derived [in comparison to your fellow students].”

#### Dependent Variable: Intention to Use the Tool

* Intention to use the tool Gado et al. (2022); all slightly adapted:
  + “I intend to use [short tool description] in my future job as a psychotherapist.”
  + “I predict I would use [short tool description] in my future job as a psychotherapist.”
  + “I plan to use [short tool description] in my future job as a psychotherapist.”

#### Additional Variables and Control Variables

* Technology readiness
* Perceived trust in the tool/ credibility
* Professional identity
* General technology affinity
* Computer self-efficacy
* Technostress
* Relevant education content (stats course)
* Personality (Park & Woo, 2022)
* Data privacy concerns
* affective, cognitive, behavioral attitudes towards AI (Park & Woo, 2022)

# References

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

Atkins, D. C., Steyvers, M., Imel, Z. E., & Smyth, P. (2014). Scaling up the evaluation of psychotherapy: Evaluating motivational interviewing fidelity via statistical text classification. *Implementation Science: IS*, *9*, 49. <https://doi.org/10.1186/1748-5908-9-49>

Can, D., Marín, R. A., Georgiou, P. G., Imel, Z. E., Atkins, D. C., & Narayanan, S. S. (2016). "It sounds like...": A natural language processing approach to detecting counselor reflections in motivational interviewing. *Journal of Counseling Psychology*, *63*(3), 343–350. <https://doi.org/10.1037/cou0000111>

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., & Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, *20*(2), 154–170. <https://doi.org/10.1002/wps.20882>

Chen, Z. S., Prathamesh, Kulkarni, Galatzer-Levy, I. R., Bigio, B., Nasca, C., & Zhang, Y. (2022). *Modern Views of Machine Learning for Precision Psychiatry* (No. arXiv:2204.01607). arXiv. <https://arxiv.org/abs/2204.01607>

Cummins, R., Ewbank, M., Martin, A., Tablan, V., Catarino, A., & Blackwell, A. (2019). *TIM: A Tool for Gaining Insights into Psychotherapy* (p. 3506). <https://doi.org/10.1145/3308558.3314128>

Fan, W., Liu, J., Zhu, S., & Pardalos, P. M. (2020). Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). *Annals of Operations Research*, *294*(1), 567–592. <https://doi.org/10.1007/s10479-018-2818-y>

Feldman, R., Aldana, E., & Stein, K. (2019). Artificial Intelligence in the Health care Space: How We Can Trust What We Cannot Know. *Policy Review*, *30*, 23.

Gado, S., Kempen, R., Lingelbach, K., & Bipp, T. (2022). Artificial intelligence in psychology: How can we enable psychology students to accept and use artificial intelligence? *Psychology Learning & Teaching*, *21*(1), 37–56. <https://doi.org/10.1177/14757257211037149>

Helge Rønnestad, M., & Ladany, N. (2006). The impact of psychotherapy training: Introduction to the special section. *Psychotherapy Research*, *16*(3), 261–267. <https://doi.org/10.1080/10503300600612241>

Hirsch, T., Soma, C., Merced, K., Kuo, P., Dembe, A., Caperton, D. D., Atkins, D. C., & Imel, Z. E. (2018). “It’s hard to argue with a computer:” Investigating Psychotherapists’ Attitudes towards Automated Evaluation. *DIS. Designing Interactive Systems (Conference)*, *2018*, 559–571. <https://doi.org/10.1145/3196709.3196776>

Huang, Z., Epps, J., Joachim, D., & Chen, M. (2018). *Depression Detection from Short Utterances via Diverse Smartphones in Natural Environmental Conditions* (p. 3397). <https://doi.org/10.21437/Interspeech.2018-1743>

Imel, Z. E., Pace, B. T., Soma, C. S., Tanana, M., Hirsch, T., Gibson, J., Georgiou, P., Narayanan, S., & Atkins, D. C. (2019). Design feasibility of an automated, machine-learning based feedback system for motivational interviewing. *Psychotherapy*, *56*(2), 318–328. <https://doi.org/10.1037/pst0000221>

Johnsen, T. J., & Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychological Bulletin*, *141*(4), 747–768. <https://doi.org/10.1037/bul0000015>

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, *17*(1), 195. <https://doi.org/10.1186/s12916-019-1426-2>

Lee, E. E., Torous, J., De Choudhury, M., Depp, C. A., Graham, S. A., Kim, H.-C., Paulus, M. P., Krystal, J. H., & Jeste, D. V. (2021). Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(9), 856–864. <https://doi.org/10.1016/j.bpsc.2021.02.001>

Park, J., & Woo, S. E. (2022). Who Likes Artificial Intelligence? Personality Predictors of Attitudes toward Artificial Intelligence. *The Journal of Psychology*, *156*(1), 68–94. <https://doi.org/10.1080/00223980.2021.2012109>

Pinto dos Santos, D., Giese, D., Brodehl, S., Chon, S. H., Staab, W., Kleinert, R., Maintz, D., & Baeßler, B. (2019). Medical students’ attitude towards artificial intelligence: A multicentre survey. *European Radiology*, *29*(4), 1640–1646. <https://doi.org/10.1007/s00330-018-5601-1>

Roth, C. B., Papassotiropoulos, A., Brühl, A. B., Lang, U. E., & Huber, C. G. (2021). Psychiatry in the Digital Age: A Blessing or a Curse? *International Journal of Environmental Research and Public Health*, *18*(16), 8302. <https://doi.org/10.3390/ijerph18168302>

Schwalbe, C. S., Oh, H. Y., & Zweben, A. (2014). Sustaining motivational interviewing: A meta-analysis of training studies. *Addiction*, *109*(8), 1287–1294. <https://doi.org/10.1111/add.12558>

Sendak, M. P., D’Arcy, J., Kashyap, S., Gao, M., Nichols, M., Corey, K., Ratliff, W., & Balu, S. (2020). A Path for Translation of Machine Learning Products into Healthcare Delivery. *EMJ Innovations*. <https://doi.org/10.33590/emjinnov/19-00172>

Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior*, *115*, 106552. <https://doi.org/10.1016/j.chb.2020.106552>

Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2021). Towards Facilitating Empathic Conversations in Online Mental Health Support: A Reinforcement Learning Approach. *Proceedings of the Web Conference 2021*, 194–205. <https://doi.org/10.1145/3442381.3450097>

Stasak, B., & Epps, J. (2017). Differential performance of automatic speech-based depression classification across smartphones. *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 171–175. <https://doi.org/10.1109/ACIIW.2017.8272609>

Tamori, H., Yamashina, H., Mukai, M., Morii, Y., Suzuki, T., & Ogasawara, K. (2022). Acceptance of the Use of Artificial Intelligence in Medicine Among Japan’s Doctors and the Public: A Questionnaire Survey. *JMIR Human Factors*, *9*(1), e24680. <https://doi.org/10.2196/24680>

Tanana, M. J., Soma, C. S., Srikumar, V., Atkins, D. C., & Imel, Z. E. (2019b). Development and Evaluation of ClientBot: Patient-Like Conversational Agent to Train Basic Counseling Skills. *Journal of Medical Internet Research*, *21*(7), e12529. <https://doi.org/10.2196/12529>

Tanana, M. J., Soma, C. S., Srikumar, V., Atkins, D. C., & Imel, Z. E. (2019a). Development and Evaluation of ClientBot: Patient-Like Conversational Agent to Train Basic Counseling Skills. *Journal of Medical Internet Research*, *21*(7), e12529. <https://doi.org/10.2196/12529>

Tran, A. Q., Nguyen, L. H., Nguyen, H. S. A., Nguyen, C. T., Vu, L. G., Zhang, M., Vu, T. M. T., Nguyen, S. H., Tran, B. X., Latkin, C. A., Ho, R. C. M., & Ho, C. S. H. (2021). Determinants of Intention to Use Artificial Intelligence-Based Diagnosis Support System Among Prospective Physicians. *Frontiers in Public Health*, *9*, 755644. <https://doi.org/10.3389/fpubh.2021.755644>

Venkatesh, V. (2022). Adoption and use of AI tools: A research agenda grounded in UTAUT. *Annals of Operations Research*, *308*(1-2), 641–652. <https://doi.org/10.1007/s10479-020-03918-9>

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478. <https://doi.org/10.2307/30036540>

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2016). *Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead* ({{SSRN Scholarly Paper}} No. 2800121).

Zhai, H., Yang, X., Xue, J., Lavender, C., Ye, T., Li, J.-B., Xu, L., Lin, L., Cao, W., & Sun, Y. (2021). Radiation Oncologists’ Perceptions of Adopting an Artificial Intelligence–Assisted Contouring Technology: Model Development and Questionnaire Study. *Journal of Medical Internet Research*, *23*(9), e27122. <https://doi.org/10.2196/27122>

1. A counseling method to enhance a patient’s motivation to change. [↑](#footnote-ref-23)