Attitudes towards the adoption of two AI-enabled mental health tools among prospective psychotherapists

Anne-Kathrin Kleine, Eesha Kokje, Eva Lermer, & Susanne Gaube

# Pre-registration statement

The hypotheses were pre-registered on the Open Science Framework ([osf.io/fqdzb](https://osf.io/fqdzb)). Exploratory hypotheses are identified as such.

# Introduction

Despite increasing efforts to develop user-friendly AI applications, they are still underutilized in clinical care [@sendak\_etal20]. Factors interfering with the adoption of clinical support tools can be found on the individual, organizational, and system level [@greenhalgh\_etal17; @yusof\_etal08]. Initial obstacles such as a lack of innovation culture, stakeholder interests, or financial risk may impede the introduction of AI into clinical care [@shachak\_etal19]. If the basic requirements are satisfied, the implementation of these tools highly relies upon the practitioner’s willingness to use them. Several frameworks and theories have been applied to explain the mechanisms influencing the implementation of clinical support systems in practice [@shachak\_etal19; @hsiao\_chen16; @kumar\_etal23; @wiljer\_etal21]. The two most relevant models for individual level predictors are the Unified Theory of Acceptance and Use of Technology [UTAUT, @venkatesh22, @venkatesh\_etal03, @venkatesh\_etal16] and the Technology Acceptance Model [TAM, @davis89].

Multiple research studies have demonstrated the applicability of the Unified Theory of Acceptance and Use of Technology (UTAUT) and the Technology Acceptance Model (TAM) to the individual-level context of clinical support systems [e.g., @arfi\_etal21; @fan\_etal20; @lin\_etal21; @zhai\_etal21; @tran\_etal21; @gado\_etal22]. UTAUT and TAM consider individual attitudes towards specific technologies, such as perceived usefulness and perceived ease of use (TAM), as relevant drivers of technology acceptance and use. However, only one study has thus far examined the predictors of the intention to use AI-enabled tools in mental health care [@gado\_etal22]. Results from this study, based on the UTAUT, suggest a link between perceived social norm, perceived ease of use, and perceived usefulness with psychology students’ intention to use AI-enabled tools in mental health practice.

Mental health practitioners may have varying levels of skepticism about the implementation of artificial intelligence (AI) in their practice. For example, when presented with AI-generated feedback regarding diagnostic or treatment decisions, practitioners may be cautious and take steps to ensure accuracy. At the same time, they may be open to incorporating AI-generated feedback into certain aspects of their therapeutic sessions to enhance the quality and effectiveness of care. Although research has begun to examine practitioners’ acceptance of AI-enabled tools, there is a lack of specificity, limiting the ability to use the current research to inform practice. This study seeks to address this caveat by examining the intention to use two specific mental health tools: a psychotherapy feedback tool and a treatment recommendation tool. The psychotherapy feedback tool is an AI system that is currently used to analyze data from therapist-patient conversations and provide performance-specific feedback for the therapist, in order to improve motivational interviewing [@cummins\_etal19; @hirsch\_etal18; @tanana\_etal19a; @imel\_etal19]. Similarly, the treatment recommendation tool uses voice recordings and mood scores to generate recommendations for psychotherapeutic support [@huang\_etal18].

This study builds upon and expands previous research regarding the intention to use AI-enabled mental health tools in four major ways. Firstly, we assess the predictors of an individual’s intention to use two specific mental health tools to potentially uncover factors related to the acceptance of tools characterized by specific technological features and designated for certain use cases. Secondly, based on previous findings, we extend the original UTAUT model by considering trust, specific understanding of the tools, and general AI knowledge as predictors of students’ intention to use the tools in their future jobs [@arfi\_etal21; @gado\_etal22]. Thirdly, we test the research model among a sample of psychology masters’s students and psychotherapists in training, increasing the practical relevance of the findings. This is because, unlike established psychotherapists, psychology students are required to complete in-depth training to become psychotherapists. Hence, the current research findings may provide useful starting points for incorporating elements into study curricula and psychotherapy training which enhance students’ intention to use AI-enabled tools in their future jobs. Finally, we utilize regularized structural equation modeling (RegSEM) to study our research model. Multicollinearity and associated suppression effects are frequently reported in studies investigating multiple UTAUT predictors at once [e.g., @bu\_etal21; @chimborazo-azogue\_etal21; @yoo\_etal15]. RegSEM can offer more reliable estimates and higher statistical power than non-regularized structural equation models, thus potentially overcoming issues associated with multicollinearity [@friemelt\_etal22; @scharf\_etal21]. In the following, we provide a brief description of the two tools investigated in this research, before introducing the research model.

# The AI-enabled feedback tool

The provision of supervision and performance feedback on psychotherapy sessions supports trainees’ skills acquisition and retention [@tanana\_etal19, @moyers\_etal05; @helgeronnestad\_ladany06]. However, providing ongoing feedback is labor and cost intensive and thus rarely used in training and clinical practice. Often, feedback is based on trainees’ self-reports and is only available after the therapy session has concluded [@tanana\_etal19]. AI technology may help reduce this problem by providing continuous, immediate, and performance-specific feedback to psychotherapists and trainees. For example, *TIM* (Therapy Insights Model) uses real-time chat messages exchanged between therapists and patients to provide feedback on topics covered in the session and those which should be addressed in the following session [@cummins\_etal19]. *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. This includes six summary measures of MI fidelity: empathy, MI spirit, reflection-to-question ratio, percent open questions, percent complex reflections, and percent MI adherence. A visual summary of the counseling sessions is also included, based on the fidelity assessment, which may be used to inform improvement [@hirsch\_etal18]. The tool chosen for the current study was developed based on *CORE-MI*. Participants are presented with information on how speech data recorded during a psychotherapy session is processed and analyzed using ML models to generate feedback for psychotherapists regarding their adherence to MI principles and possibilities for improvement, as shown in [Figure 1](#fig-feedback).

|  |
| --- |
| Figure 1: The output slide of the AI-enabled feedback tool |

# The AI-enabled treatment recommendation

Multiple studies have demonstrated the effectiveness of AI-enabled emotion analysis in assessing patients’ depressive states and recommending timely intervention, thus advancing mental healthcare [@jan\_etal18; @huang\_etal18]. In particular, systems that monitor or assess the mood of individuals with mental disorders, such as major depressive disorder (MDD) or bipolar disorder, using speech data have been developed over the past years [@karam\_etal17; @khan\_etal16]. These systems usually involve the patient recording voice samples through an application installed on their mobile phone, followed by the automated speech data classifier analyzing the data to assess the patient’s current mood [@karam\_etal17]. The generated mood score can then be used by mental health practitioners to decide whether urgent intervention is needed, particularly as patients with MDD hold a 40% risk of non-fatal lifetime suicide attempts [@sokero\_etal05]. Timely psychotherapeutic support may lower the risk of aggravation of depressive symptoms and suicidality [@calati\_courtet16]. An example of such a tool is the one developed by *SondeHealth*, which requests the patient to record a voice message answering a predetermined question using their mobile phone, and uses the voice data in a machine learning (ML) model to generate a mood score. Psychotherapists may use the mood score information to decide whether emergency intervention is necessary and whether a patient needs to be given preference in treatment [@sondehealth.com]. The tool chosen for the current study is based on the tool developed by *SondeHealth*. Participants are presented with information on how voice data recorded on a mobile device is processed and analyzed using ML models to generate a mood score that may be used for treatment-related decisions, as shown in [Figure 2](#fig-depression).

|  |
| --- |
| Figure 2: The output slide of the AI-enabled feedback tool |

# Research model and hypotheses development

The first goal of the current research is to test the applicability of a modified version of the UTAUT in the mental health context to understand the factors that influence the intention to use two specific AI-enabled mental healthcare tools [@gado\_etal22; @venkatesh22; @venkatesh\_etal03; @venkatesh\_etal16]. In line with the UTAUT, we propose tool-specific performance expectancy (i.e., the degree to which an individual believes that using a system will enhance their performance) and effort expectancy (i.e., the degree of ease associated with using the technology) to predict the behavioral intention to use the two tools in their future jobs.

*Hypothesis 1*: There is a positive relationship between perceived performance expectancy and the intention to use the tools in psychotherapy.

*Hypothesis 2*: There is a positive relationship between perceived effort expectancy and the intention to use the tools in psychotherapy.

In contrast to experienced psychotherapists, psychology students and psychotherapists in training may be less likely to be influenced by established habits or work procedures which could impede the adoption of new artificial intelligence (AI) technologies [@venkatesh\_etal16]. It has been suggested that students are more likely to be affected by their peers and the values and standards of their potential employers [@owusu\_etal22]. As a result, we propose that the UTUAT (Unified Theory of Use and Acceptance of Technology) variable, ‘social influence’ (i.e., the perception that significant others think the system should be used), should be considered a predictor of students’ intention to use the feedback tool.

*Hypothesis 3*: There is a positive relationship between social influence and the intention to use the tool in psychotherapy.

It has been suggested that trust may be a relevant predictor of the intention to use a technology if the level of risk associated with it is high [@arfi\_etal21]. This is particularly relevant for the treatment recommendation tool in this study, given the critical nature of the recommendations in comparison to the feedback tool. Thus, we hypothesize that trust may be a predictor of students’ intention to use the tools, and that levels of trust and the relationship between trust and the intention to use the tool may differ between the two tools.

*Hypothesis 4*: There is a positive relationship between trust in the tools and the intention to use them in psychotherapy.

*Exploratory Hypothesis 4*: a) The level of trust in the treatment recommendation tool will be lower than the level of trust in the feedback tool, and b) the relationship between trust and the intention to use the tool will be stronger for the treatment recommendation tool than for the feedback tool.

A lack of understanding of the underlying mechanisms of AI-enabled tools in mental healthcare has led to skepticism of their use [@aafjes-vandoorn\_etal21; @chekroud\_etal21]. In particular, the black box problem of AI-based recommendations has impeded adoption of such tools in mental healthcare due to the importance of transparency and explainability of clinical decision-making [@aafjes-vandoorn\_etal21; @chekroud\_etal21; @kelly\_etal19]. Building on the New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies [NASSS, @greenhalgh\_etal17], we propose that knowledge of the technology is a predictor of its perceived demand-side value. We suggest that students who possess the knowledge and skills to apply the tools and understand how the recommendations are derived are more likely to perceive them as useful and may be better equipped to address ethical concerns [@seufert\_etal21; @gado\_etal22]. To test this, we extend the UTAUT model by including cognitive technology readiness as an indicator of general AI knowledge and specific understanding of the tool as an indicator of specific AI knowledge as predictors of performance expectancy, effort expectancy, and trust. We pre-registered two research questions to test this relationship and tehrefore propose the following two matching exploratory hypotheses:

*Exploratory Hypothesis 5*: The positive relationship between cognitive technology readiness and the intention to use the tools is mediated through a) performance expectancy, b) effort expectancy, and c) trust in the tools.

*Exploratory Hypothesis 6*: The positive relationship between specific understanding of the tools and the intention to use the tools is mediated through a) performance expectancy, b) effort expectancy, and c) trust in the tools.

# Methods

## Participants

Psychology students and psychotherapists were recruited online through social media postings and contacts with administrative offices of universities and psychotherapy training centers. In addition, Prolific, a professional research-focused panel company was commissioned to recruit participants online. The data was collected from October 2022 until January 2023. In total, 362 individuals started answering the questionnaire. Of those, 208 provided answers on the behavioral intention to use the tools (42.54% dropout rate). Finally, rnrow(data\_all\_behav) - nrow(data\_all\_noatt)` participants failed at least two of the four attention check items [@Oppenheimer2009], leaving us with a final sample size of 206.

Of the final sample, *n* = 33 (16.02%) were male, *n* = 165 (80.1%) were female, and 8 (3.88%) non-binary defined. Participants’ age ranged from 18 to 54, with a mean age of 28.1 years (*SD* = 7.03. Most participants (*n* = 109, 52.91) studied in Germany, *n* = 44 (21.36) in the UK, *n* = 32 (15.53) in the USA, *n* = 13 (6.31) in Canada, and *n* = 35 (17.24) elsewhere. Most participants (*n* = 118, 58.13%) indicated that their studies were focused on clinical psychology, *n* = 50 (24.63%) indicated that they studied psychology with no specific focus, and *n* = NA (NA%) did not provide this information.

## Measurement instruments

### Independent variables

First, we assessed cognitive AI-readiness with five items of the cognition factor of the medical artificial intelligence readiness (MAIRS) scale [@karaca\_etal21]. The scale measures terminological knowledge about medical artificial intelligence applications. An example item reads: “I can explain how AI systems are trained” ( = 0.75, = 0.75). Next, participants were presented with slides that explained how recommendations for the AI-enabled feedback tool (henceforth, Tool 1) and the treatment recommendation tool (henceforth, Tool 2) were generated (the material is available from the first author upon request). Before seeing the slides, participants read the following short introduction: “On the following page, you will be presented with a tool that is used to [*Tool 1*: provide feedback to psychotherapists about what went well and what could be improved in their sessions; *Tool 2*: generate a mood score to rate the severity of patients’ depression. The mood score may be used by psychotherapists to decide which patient to treat first if multiple patients seek treatment and there is limited capacity]. Please read the information carefully and try to understand what the tool does and how it may be used in psychotherapy practice/ training. After the presentation, you will be asked a couple of questions about the tool.” After the presentation of the two tools, UTAUT variables were assessed for each tool separately. The introduction for the scales read: “To what extent do you agree or disagree with the following statements regarding the AI-enabled feedback tool?”. Answers were provided on a five-point Likert scale with options ranging from *1 = “Strongly disagree”* to *5 = “Strongly agree”*. Performance expectancy, effort expectancy, and social influence were measured with items adapted from @venkatesh\_etal03. *Performance expectancy* was assessed with five items (e.g., “using the AI tool would enable me to accomplish tasks more quickly”). Reliabilities are = 0.86 and = 0.91 for the first tool and = 0.91 and = 0.93 for the second tool. *Effort expectancy* was measured with four items (e.g., “my interaction with the AI tool will be clear and understandable”; = 0.84, = 0.89; = 0.89, = 0.93). *Social influence* was measured with five items (e.g., “in my future job as a psychotherapist, people who are important to me will think that I should use the AI tool”; = 0.88, = 0.94; = 0.91, = 0.95). Trust was measured with three items adapted from @venkatesh\_etal11 (e.g., “the AI tool will provide access to sincere and genuine feedback”; = 0.83, = 0.84; = 0.89, = 0.89). Finally, specific understanding of the AI-enabled tools was assessed with a single item (“Please rate your understanding of the AI-enabled feedback tool”), with answers ranging from *1 = “I don’t understand the tool at all”* to *6 = “I understand the tool extremely well”*.

### The behavioral intention to use the tools as dependent variable

The behavioral intention to use the tools was measured on a five-point Likert scale ranging from *1 = “Strongly disagree”* to *5 = “Strongly agree”* with three items adapted from @venkatesh\_etal03 (e.g., “I intend to use the AI tool in my future job as a psychotherapist”; = 0.95, = 0.95; = 0.96, = 0.96).

### Control variables

Data privacy concerns and AI anxiety as fears and a sense of insecurity regarding AI technology have repeatedly been identified as negative predictors of the intention to use AI technology [e.g., @mishra\_etal21, @chai\_etal20]. In addition, it has been shown that males have more positive attitudes toward AI technologies than females [e.g., @fietta\_etal22]. Finally, some evidence exists for associations of AI acceptance with age [@liang\_lee17] and country [@sindermann\_etal21]. Accordingly, we included data privacy and security concerns [@brady\_etal21, = 0.76, = 0.85; = 0.79, = 0.91], AI anxiety [@venkatesh\_etal03, = 0.78, = 0.81; = 0.76, = 0.79], gender (1 = not male, 0 = male), age, and study country as control variables.

# Data analysis

The data was analyzed using *R*. First, we calculated descriptive statistic summaries, including mean values, standard deviations, and correlations between study variables for each tool. Second, we conducted regularized structural equation modeling (RegSEM) using the *lavaan* and *regsem* [@jacobucci\_etal22] packages to examine the relationship between the predictor variables and the intention to use the tools to answer hypotheses 1 through 4. The use of regularization penalties to specific parameters in structural equation modeling has been introduced as a means to prevent overfitting the data as well as to reduce unnecessary model complexity, thus increasing the stability and generalizability of the findings from the fitted model [@jacobucci\_etal22; @ober\_etal21]. Through the implementation of lasso regularization penalties, irrelevant coefficient estimates become exactly equal to zero, which makes the regression model easier to interpret [@mcneish15; @melkumova\_shatskikh17]. Finally, we used the *regsem* and *lavaan* packages to test the exploratory mediation models [\*Hypotheses 5\* and \*6\*, @serang\_etal17].

# Results

lavaan 0.6-12 ended normally after 14 iterations  
  
 Estimator ML  
 Optimization method NLMINB  
 Number of model parameters 5  
  
 Number of observations 206  
  
Model Test User Model:  
   
 Test statistic 0.000  
 Degrees of freedom 0  
  
Parameter Estimates:  
  
 Standard errors Standard  
 Information Expected  
 Information saturated (h1) model Structured  
  
Covariances:  
 Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
 trust2 ~~   
 trust1 0.479 0.072 6.628 0.000 0.479 0.521  
  
Intercepts:  
 Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
 trust2 (m\_t2) 2.982 0.072 41.145 0.000 2.982 2.867  
 trust1 (m\_t1) 3.374 0.062 54.793 0.000 3.374 3.818  
  
Variances:  
 Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
 trust2 1.082 0.107 10.149 0.000 1.082 1.000  
 trust1 0.781 0.077 10.149 0.000 0.781 1.000  
  
Defined Parameters:  
 Estimate Std.Err z-value P(>|z|) Std.lv Std.all  
 diff 0.392 0.066 5.905 0.000 0.392 0.951

Paired t-test  
  
data: comp\_df2$trust1 and comp\_df2$trust2  
t = 5.8904, df = 205, p-value = 1.56e-08  
alternative hypothesis: true mean difference is not equal to 0  
95 percent confidence interval:  
 0.2605153 0.5226562  
sample estimates:  
mean difference   
 0.3915858

[1] 3.373786

[1] 0.8858924

[1] 2.982201

[1] 1.042826

Table 1 shows the means, standard deviations, and correlations between the study variables. To test hypotheses 1 through 4, we specified an SEM with the behavioral intention to use Tool 1 and Tool 2 to be predicted by the respective UTAUT variables (i.e., performance expectancy, effort expectancy, social influence, trust), AI anxiety, data privacy concerns, cognitive readiness, specific tool understanding, age, gender, and study country. Next, we applied regularization to the SEM. We first determined the optimal penalty value () by comparing Bayesian information criteria (BIC) among a set of 90 penalty values. We retained the value of associated with the best fitting model. [Figure 3](#fig-BIC) shows the parameter trajectory plot for the model estimates against the values of . The dots represent the size of the estimates where is optimal according to BIC. In total, there were ten parameters (five for the prediction of the intention to use Tool 1 and Tool 2, respectively) for which the estimates were shrunken to zero according to the results of the RegSEM. Table 2 shows the results of the SEM and RegSEM. Performance expectancy, social influence, and trust (but not effort expectancy) were positively associated with the intention to use Tool 1 and Tool 2, supporting hypotheses 1, 3, and 4 (but not Hypothesis 2). AI anxiety was negatively associated with the intention to use both tools. Age, gender, and country showed no relationship with the intention to use the tools. In preparation for the test of Hypotheses 5 and 6, we also examined the direct relationship of cognitive readiness and specific tool understanding with the intention to use the tool. Cognitive readiness was positively associated with the intention to use Tool 1, but was unrelated to the intention to use Tool 2. Specific tool understanding was unrelated

The results of structural equation modeling (SEM) paired t-tests showed that trust in Tool 1 (*M* = 3.37, *SD* = 0.062) was higher than in Tool 2 (*M* = 2.98, *SD* = 1.043; *d* = 0.392, *t*(205) = 5.890, p < .001), supporting Hypothesis 4 a). Finally, the standardized estimate () for the coefficient of the relationship between trust and the intention to use the tool was higher for Tool 1 than for Tool 2, supporting Hypothesis 4 b).

|  |
| --- |
| Figure 3: **?(caption)** |

# Appendix

## Discussion

# References

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