International Platform of Registered Systematic Review and Meta-analysis Protocols

INPLASY

Psychiatric Principles, Affiliation, and Patient Safety in Mental Health Chatbots: A Systematic Review

INPLASY202570035
doi: 10.37766/inplasy2025.7.0035
Received: 9 July 2025
Published: 9 July 2025

Ehlers, J; Ostermann, T; Hecht-Shin, M.

Corresponding author:
Mandy Hecht-Shin

mandy.hecht-shin@uni-wh.de

Author Affiliation:
n/a.

ADMINISTRATIVE INFORMATION

Support - n/a.

Review Stage at time of this submission - The review has not yet started.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202570035

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 9 July 2025 and was last updated on 9 July 2025.

INTRODUCTION

Review question / Objective Recent studies have examined and appraised the technical implementation of artificial intelligence (AI) in psychotherapy, with the objective of ascertaining the extent to which and the way AI can be employed in therapeutic settings. These studies have also identified the opportunities and challenges that emerge from this integration. A determination should be made regarding the extent to which the application complies with psychotherapeutic standards and guidelines, including ethical principles. Additionally, there appears to be a paucity of research investigating whether AI solutions should be utilized to reduce waiting times for therapy appointments or whether they can be incorporated into ongoing therapeutic interventions. This prompts the question of the extent to which AI solutions can be integrated into both self-therapeutic treatment and therapeutic practice. Additionally, it is relevant to determine the feasibility of replacing the therapeutic relationship with an alternative arrangement. This study undertakes an examination of the current state of research in the field of psychology in connection with AI. Its objective is to furnish a comprehensive overview of the extant research activities in this domain.

Condition being studied The incidence and prevalence of mental disorders, including depression and anxiety disorders, have been observed to increase at a significant rate. The recommended course of action for the treatment of these disorders is psychotherapeutic intervention. While acute depression and anxiety frequently necessitate prompt psychotherapeutic intervention, in certain instances, a protracted wait for a treatment slot may ensue. In certain instances, this phenomenon manifests itself regionally in a more pronounced form in rural areas, where the range of psychotherapeutic treatment options may be constrained.

Notwithstanding the prior negative trend, a positive trend is emerging in the domain of technical development. One such example is the ongoing development of AI solutions. A select number of companies have already introduced AI chatbots that offer psychotherapeutic counseling. A notable benefit of AI applications is their potential to enhance acceptance and reduce barriers to engagement for individuals grappling with mental health challenges, owing to their anonymous nature. Furthermore, there is a high probability that not only generative AI but also image and video execution will undergo significant improvement. Subsequently, the efficacy of replacing a psychotherapeutic session with an image and sound combination will be assessed. A more thorough examination is therefore necessary to determine the extent to which technical solutions are suitable for therapeutic treatments and whether therapeutic guidelines, standards, and ethical principles are being met.

METHODS

Search strategy A comprehensive and interdisciplinary search strategy was developed for the planned systematic work, taking into account psychological, medical, and technological aspects of digital interventions. As part of the study, central databases such as the Open Science Framework (OSF), APA PsycINFO, and PsycArticles were utilized to identify studies that were supported by empirical evidence and focused on psychological interventions, diagnostics, and patient interaction. Additionally, a comprehensive review of medical databases, including PubMed and MEDLINE, was conducted to identify relevant clinical studies, psychiatric research, and the application of AI in diagnostics and therapy. To address the rapidly evolving landscape of technical developments in the fields of artificial intelligence, natural language processing, and digital health tools, a comprehensive review of relevant platforms was conducted. This included well-regarded resources such as IEEE Xplore, SpringerLink, and Scopus. These resources facilitate the analysis of citations and encompass the interfaces between psychology, computer science, and medicine. To identify systematic reviews and ongoing studies, specialized platforms such as Cochrane Library, INPLASY, ResearchGate, and JMIR were consulted. Furthermore, a supplementary analysis of databases such as the WHO Digital Health Atlas was conducted. The objective of this analysis is to obtain a comprehensive global overview of digital health applications, with a particular emphasis on mental health and artificial intelligence.

Participant or population None declared.

Intervention None declared.

Comparator None declared.

Study designs to be included None declared.

Eligibility criteria The following inclusion criteria were considered in this study: The following sources are relevant for this research: systematic reviews, meta-analyses, umbrella reviews, systematic literature reviews (SLRs), systematic reviews of reviews, scoping reviews according to PRISMA-ScR, scoping reviews according to Arksey & O'Malley, scoping reviews with thematic analysis, narrative reviews, literature reviews, and literature-based analysis. The following research methodologies have been employed: conceptual analysis, randomized controlled trials (RCTs), pilot RCTs, exploratory RCTs, quasi-experimental studies, mixed-method studies, user studies, qualitative interviews, qualitative content analysis, review mining, natural language processing (NLP) analysis, design science approach, Delphi method, framework development, and conceptual modeling. Additionally, this paper will examine the applicable laws, regulations, and psychotherapeutic guidelines.
The sole exclusion criterion for this study was the exclusion of studies that considered only theoretical results without evaluating empirical data.

Information sources The following bibliographic databases are relevant for this research: APA PsycINFO, APA PsycArticles, the Psychology and Behavioral Science Collection (PBSC), PubMed/MEDLINE, PSYNDEX (via ZPID), IEEE Xplore, and Scopus. A comprehensive list of relevant databases includes Web of Science, SpringerLink, Google Scholar, ResearchGate, the Directory of Open Access Journals (DOAJ), Cochrane Library, INPLASY, and Digital Mental Health Databases (e.g., JMIR, BMC Digital Health). The WHO Digital Health Atlas is also a pertinent resource.

In accordance with the investigative framework, the search terms were methodically categorized into five distinct classifications: (a) technology and system architecture, (b) specific AI applications in psychotherapy, (c) effectiveness, outcomes, and study findings, (d) comparison between humans and AI, and (e) quality and ethics.

In the course of the investigation, various queries were used in which the terms were combined or retained. The search terms employed in this study

are as follows: The following terms are relevant to the field of artificial intelligence and related areas:

artificial intelligence, machine learning, deep learning, natural language processing, large language model or LLM, chatbot architecture, psychological prompts, GPT-4, ChatGPT, blind methodology, hybrid models, AI-based system design, mental health chatbot, AI therapy app, virtual therapist, CBT chatbot, digital mental health, AI in psychotherapy, e-mental health AI, AI-assisted psychotherapy. The following terms are relevant to the field of digital interventions for mental health: chatbots, mobile health (mHealth), remote patient monitoring, mental telehealth, healthcare ecosystem, clinical effectiveness, treatment outcomes, efficacy of AI therapy, symptom reduction, and AI. The following terms are relevant to randomized controlled trials (RCTs) of AI therapy: outcomes, engagement, validity, adherence, evaluation, depression, anxiety, PHQ-9, core outcome set, and meta-analysis. A comprehensive examination of the extant literature reveals a plethora of studies addressing a myriad of topics relevant to the intersection of artificial intelligence (AI) and mental health. These include, but are not limited to, the following: systematic and scoping reviews, workplace well-being, the phenomenon of burnout, user satisfaction, the comparison of AI chatbots and human therapists, the comparison of AI therapists and psychotherapists, human-AI comparisons in mental health, empathy AI versus human, the therapeutic alliance AI, diagnostic accuracy AI, performance comparisons, ethical considerations in AI mental health, quality standards in psychotherapy, the digital landscape of regulatory frameworks for AI health, patient safety AI, and ethical concerns and safety issues in the context of trust in AI psychotherapy.

Main outcome(s) The planned study will provide a comprehensive overview of the mental illnesses for which the utilization of AI applications in treatment is pertinent. This will include an explanation of the implementation of these technologies in therapeutic measures and a review of guidelines, standards, and ethical principles in the relevant studies. Furthermore, an evaluation of extant quality standards for psychotherapy will be conducted to ascertain the extent to which they can be utilized to assess the implementation of AI in therapy sessions.

Quality assessment / Risk of bias analysis None declared.

Strategy of data synthesis None declared.

Subgroup analysis None declared.

Sensitivity analysis None declared.

Language restriction English.

Country(ies) involved Germany.

Keywords Artificial Intelligence (AI), Chatbots, Psychotherapy, Ethical Principles, Psychotherapeutic Standards, Therapeutic Relationship, Meta-analysis, Review, Mental, AI in Mental Health, Feasibility Assessment.

Contributions of each author

Author 1 - Jan Ehlers.
Email: jan.ehlers@uni-wh.de
Author 2 - Thomas Ostermann.
Email: thomas.ostermann@uni-wh.de
Author 3 - Mandy Hecht-Shin.
Email: mandy.hecht-shin@uni-wh.de