

As a preliminary search into my topic, I used BrowZine to assess the scope and availability of existing research. Since my topic focuses on AI chatbots, which are a relatively new and rapidly evolving technology, I first needed to determine if sufficient literature existed to answer my research question. I began with general search terms such as "artificial intelligence in health", but it yielded results related to physical health rather than mental health. In response, I refined my search to "mental health and technology", which led me to the journal *Internet Interventions*. The journal exclusively publishes work that examines the application of technology in the mental and behavioral health contexts.

Through a brief initial search of the journal by only using the search term "AI chatbots", I found numerous relevant literature reviews and primary research studies that aligned with my topic of interest. After establishing that there was sufficient background on the topic, I refined and expanded my research question to add greater depth. While my initial focus was on AI chatbot effectiveness as a whole, I wanted to break it down to examine effectiveness through multiple perspectives, such as those of the user, clinician, and independent researchers.

I took a deeper look through *Internet Interventions* using the phrase "AI chatbots AND effectiveness". I also experimented with terms such as "conversational agents" and "digital mental health tools"; however, the results for these searches were mostly unrelated. I then reviewed abstracts to assess relevance and selected studies that reflected the range of perspectives I was interested in. I attempted to prioritize newer research that had been peer reviewed; however, a few older papers were included in the end due to their content still being relevant.

One day in class, we had a librarian come in to help us find relevant databases. Through that, I was introduced to the database PsycINFO. Using PsycINFO, I focused on gathering

qualitative data related to user experience by using specific search terms such as “user perspective” and “interviews”. I specifically selected observational studies to better understand how individuals perceive the effectiveness of AI chatbots. To assess the ethical component of my research question, I searched the phrase “AI chatbots AND medical ethics”. This yielded a combination of literature reviews that outlined the primary ethical concerns up to date and empirical studies focused on the experiences of users.

In addition to this, I wanted sources that provided measurable clinical outcomes. Thus, I utilized PubMed to obtain quantitative results of effectiveness. This brought me the perspectives of clinicians and outside researchers alike. I utilized search terms such as AI chatbots, ethics, effectiveness, users, and clinicians. I also prioritized sources that utilized common psychiatric evaluation tools like the PHQ-9 (measures depressive symptoms) and GAD-7 (screens for anxiety and its severity).

After gathering my sources, I organized them via a research matrix in which I recorded a quick summary, strengths and weaknesses, and key comparisons between the studies. This allowed me to identify patterns across studies, evaluate consistency and accuracy, and determine key themes that emerged.

I briefly explored AI research tools such as Elicit and Google Scholar Labs, where I typed my research question and directly asked for relevant sources. While the articles provided were relevant to my research, I noticed that a lot of them were older and inconsistent with current research, so I chose not to include any of the recommended papers.

A final look through my sources was then conducted. I assessed relevance by ensuring that all sources addressed my research topic, meaning they specifically used the term “AI chatbots” and it was in reference to an AI technology that engages in continuous dialogue with

the user. I also ensured that sources specifically talked about effectiveness or ethical concerns. Accuracy and credibility were assessed by ensuring that all sources were peer-reviewed and from a reputable journal. I included both literature reviews and primary research studies; however, I placed an emphasis on the latter. Currency was prioritized as publications in the past three years were heavily favored. However, a few older ones were still utilized due to their relevance in current discussions. I ensured both scope and multiple perspectives by combing through multiple databases, ensuring that different types of studies were included, and including the perspectives of users, clinicians, and independent researchers alike. Finally, I ensured all studies included human perspectives, whether from clinicians, patients, or independent researchers. Any studies that used AI to evaluate results were excluded to avoid potential bias.

In accordance with my professor's and course guidelines, I used AI to support the creation of this review. All ideas and writing were my own original work. AI was used to review grammar, sentence structure, flow, and help improve clarity. Any suggestions made by AI were reviewed by me before integration. I utilized both ChatGPT (<https://chatgpt.com/>) and Grammarly (<https://app.grammarly.com/>). ChatGPT mainly served to evaluate my work against the assignment's rubric, and it offered suggestions for areas of improvement. Occasionally, I asked the AI to reword awkward sentences or give me a different word or phrase to use. Grammarly was used to edit my spelling and grammar, and on occasion, it would restructure my sentence to improve clarity and flow. A final look at my paper was conducted by My Essay Feedback AI, which was recommended by our instructor. It provided specific feedback for areas of improvement.. I utilized its suggestion of adding depth to my methods section and incorporating a greater discussion between sources.