ProScholar Insights (PSI)
Volume 4, Issue 3 (Summer 2025)
Pages: 96 – 111

PROSCHOLAR INSIGHTS

ISSN (Print): 3006-838X
ISSN (Online): 3006-7723
DOI: 10.55737/psi.2025c-43100

Research Article **Open Access Journal**

Meta-Analytic Evaluation of AI-Based Chatbots and Self-Help Apps for Social Anxiety (2015–2025)

Farah Mushtaq [1] **Mussarat Jabeen Khan** [1] **Ayesha Jahangir** [3]

[1] MS Scholar, Department of Psychology, International Islamic University, Islamabad, Pakistan.
✉ farahmushtaq08@gmail.com

[2] Assistant Professor, Department of Psychology, International Islamic University, Islamabad, Pakistan.
✉ mussarat.jabeen@iiu.edu.pk

[3] Student, University of Houston, Victoria, United States of America.
✉ Ayesha.Jahangir2023@gmail.com

This article may be cited as Mushtaq, F., Khan, M. J., & Jahangir, A. (2025). Meta-Analytic Evaluation of AI-Based Chatbots and Self-Help Apps for Social Anxiety (2015–2025). *ProScholar Insights, 4*(3), 96-111. https://doi.org/10.55737/psi.2025c-43100

Abstract: Due to stigma, cost, and a lack of trained specialists, social anxiety disorder (SAD), one of the most common and debilitating mental health conditions, is still challenging to treat using evidence-based methods (Kessler et al., 2005). In recent years, scalable tools for managing mental health conditions like SAD have emerged, including self-guided mobile applications and chatbots powered by artificial intelligence (AI) (Fulmer et al., 2018; Inkster et al., 2018). This systematic review and meta-analysis assessed the efficacy of AI-based interventions, specifically chatbots and self-help apps, in lowering symptoms of social anxiety from 2015 to 2025. Following PRISMA guidelines, only peer-reviewed randomized controlled trials (RCTs) that measured social anxiety using validated measures such as the SPIN, LSAS, and SIAS were included. The results of quantitative synthesis using Hedges' g consistently supported digital interventions over control conditions, with a small to moderate pooled effect size ($g \approx 0.36$, 95% CI [0.29, 0.43]) (Stolz et al., 2018; Ivanova et al., 2016). The resources included modular self-help apps built with CBT or ACT frameworks and AI chatbots like Woebot and Wysa. Even though the results point to significant short-term advantages, there are still a number of limitations, such as the absence of long-term monitoring, a small sample diversity, and low capacity for cultural adaptation. More beneficial for settings having limited resources as these tools show potential alternatives or substitutes for therapist-led care. Future studies should concentrate on ethical inclusion, culturally sensitive design, and comparative trials with traditional therapies.

Keywords: Social Anxiety, Chatbot, Mobile Apps, AI Therapy, Self-Help, Digital Interventions

Check for updates

Corresponding Author:
Mussarat Jabeen Khan
Assistant Professor, Department of Psychology, International Islamic University, Islamabad, Pakistan.
✉ mussarat.jabeen@iiu.edu.pk

Introduction

A severe and persistent fear of social or performance situations where one may be subject to public scrutiny is the hallmark of social anxiety disorder (SAD), one of the most prevalent and crippling mental illnesses (American Psychiatric Association, 2013). SAD affects 7% to 13% of people at some point in their lives. The disorder usually starts in early adolescence and often lasts into adulthood, if treatment is not received (Fehm et al., 2008; Kessler et al., 2005; Stein & Stein, 2008). Significant distress in academic, professional, and interpersonal settings is common among people with SAD, which can result in avoidance behaviors, decreased functioning, and an increased risk of co-occurring disorders like substance use disorders and depression (Rodebaugh et al., 2008).

Only a small percentage of people with SAD receive proper care, despite the availability of efficient treatments like cognitive behavioral therapy (CBT), because of multiple barriers. These include the scarcity of qualified mental health specialists, the expense, the stigma associated with seeking help, logistical difficulties, and lengthy wait times, especially in low- and middle-income nations (Wang et al., 2007; Gulliver et al., 2010; Andrews et al., 2010). The very social anxiety symptoms that make it difficult to contact therapists can cause people with SAD to put off or avoid treatment, even in high-resource settings (Rickwood et al., 2005; McDonald et al., 2014).

Mobile applications and AI-powered chatbots have gained increasing attention for their ability to deliver mental health support at scale, without the need for synchronous, face-to-face interaction with a clinician (Torous et al., 2020; Fulmer et al., 2018; Linardon et al., 2019). These tools often incorporate techniques of CBT, acceptance and commitment therapy (ACT), and emotion-focused approaches, offering users psychoeducation, interactive exercises, check on symptoms, and tailored feedback (Inkster et al., 2018).

AI-powered chatbots such as Woebot, Tess, Wysa, and Youper use natural language processing and conversational algorithms to simulate therapeutic dialogue, providing real-time support and mental health coaching to users through smartphones or web interfaces (Fitzpatrick et al., 2017). These tools are typically available 24/7, are anonymous, and offer personalized responses based on user input. In contrast, self-help apps, while sometimes AI-enhanced, may rely more heavily on structured modules, interactive journaling, reminders, and self-monitoring without engaging in full dialogue. Both formats are increasingly accessible and low-cost, making them attractive options for individuals reluctant or unable to seek traditional care (Bakker et al., 2016; Firth et al., 2017).

While there has been extensive research on anxiety disorders using traditional digital cognitive behavioral therapy (iCBT), the advent of AI-powered tools marks a new trend toward more independent, interactive, and adaptable mental health treatments (Linardon et al., 2019; Chandrashekar, 2018). Moderate to large effect sizes have been found in meta-analyses of mobile mental health interventions for anxiety, especially when the intervention incorporates user engagement features and CBT-based content (Firth et al., 2017). But few of the meta-analytic studies that have been done so far have differentiated outcomes specific to social anxiety; instead, they have aggregated data from all anxiety disorders. Additionally, few studies have separated the effects of AI-powered tools, like chatbots and intelligent self-help apps, from broader digital health interventions (Torous et al., 2020; Chandrashekar, 2018).

This omission is noteworthy because social anxiety may pose particular difficulties for the delivery of digital interventions, such as increased anxiety about digital social interaction, distrust of AI, and user attrition because of a perceived lack of human warmth or customization (Scholten et al., 2016). Furthermore, although chatbots and mental health apps have a lot of potential to lower barriers to care, concerns about their practicality, long-term effectiveness, adherence, and cultural generalizability across a range of demographics still exist (Vaidyam et al., 2019; Wasil et al., 2019). The current study intends to close this important gap by performing a systematic review and meta-analysis of AI-powered mental health tools that target social anxiety, particularly chatbots and self-help mobile applications.

Objectives

1. To identify and compile randomized controlled trials (RCTs) that assess the effectiveness of AI-powered tools in lowering social anxiety symptoms and were carried out between 2015 and 2025.
2. To use meta-analytic methods to calculate a pooled effect size for AI-powered interventions in comparison to control conditions.
3. To evaluate the included studies' methodological quality and bias risk.
4. To discuss the consequences for future research priorities, clinical use, access equity, and ethical deployment.

This review adds a timely and nuanced synthesis to the expanding body of research on digital mental health by concentrating solely on social anxiety and AI-driven interventions. It seeks to educate medical professionals, scholars, and decision-makers on the advantages and disadvantages of these cutting-edge technologies in treating one of the most common and undertreated mental health issues in the world.

Method

The PRISMA 2020 guidelines were followed when conducting this review (Page et al., 2021), ensuring transparency in study selection and reporting. The eligibility criteria were designed to include only the most rigorous and relevant studies addressing the impact of AI-powered mental health tools on social anxiety symptoms.

Eligibility Criteria

Studies that satisfied the following requirements were eligible to be included:

1. Study Design: Only peer-reviewed randomized controlled trials (RCTs) or well-designed quasi-experimental studies with control conditions were included. Protocols, feasibility studies, pilot-only trials, and single-group designs were excluded unless they reported comparative results.
2. Publication Years: Studies published between January 2015 and May 2025 were eligible. This period was chosen to capture the emergence and evolution of AI-driven interventions over the past decade.
3. Population: Participants of any age group who exhibited clinical or subclinical levels of social anxiety, either self-reported or diagnosed using validated instruments (e.g., SPIN, LSAS, SIAS, SPS), were included. Studies with mixed anxiety disorders were included only if results for social anxiety were reported separately or extractable.
4. Intervention Type: Interventions had to involve AI-powered mental health tools, including:
    - Conversational chatbots (e.g., Woebot, Wysa, Tess, Youper)
    - Self-help mobile or web applications that use machine learning, natural language processing, or algorithmic personalization
    - Apps using structured therapeutic content (e.g., CBT or ACT), without human therapist involvement
5. Tools solely categorized as digital CBT programs (iCBT), internet-assisted therapy with therapist guidance, or non-AI-based digital tools were excluded, as the focus was on AI autonomy.
6. Outcomes: Studies must have reported quantitative measures of social anxiety symptoms, using validated psychometric instruments, either as a primary or secondary outcome. Studies reporting only depression or general anxiety without social anxiety measures were excluded.
7. Language: Only English-language studies were included due to resource limitations in translation.
8. Access: Full-text availability was required to ensure methodological transparency and proper data extraction.

Information Sources and Search Strategy

A systematic search was conducted across the following databases:

- PubMed
- PsycINFO
- Scopus
- Web of Science
- Google Scholar (used for reference chaining and gray literature validation)

Searches were performed from May 1 to June 20, 2025. The search strategy used Boolean operators and was adapted for each database. A sample search string used in PubMed was:

("social anxiety" OR "social phobia") AND ("chatbot" OR "AI" OR "artificial intelligence" OR "self-help app" OR "mental health app" OR "conversational agent") AND ("RCT" OR "randomized controlled trial") AND ("2015"[Date]: "2025"[Date])

Additionally, all included studies and relevant reviews completed manual backward and forward citation chaining. (e.g., Chandrashekar, 2018; Fulmer et al., 2018).

Study Selection

Title and abstract screening and full-text review were the two stages of the screening process used to select the studies. Duplicate records were eliminated after all records were imported into a reference manager. All records were evaluated for eligibility by two separate reviewers using predetermined inclusion and exclusion criteria. Discussions or, if required, consultation with a third reviewer were used to settle disagreements.

Data Extraction

The following information was collected by using a structured data extraction form for each included study:

- Citation (authors, year, journal)
- Country and study setting
- Study design (RCT, quasi-experimental, parallel-group, crossover, etc.)
- Sample size and participant characteristics (mean age, gender distribution, clinical vs subclinical)
- Intervention type and tool name (e.g., Woebot, Wysa, Youper, custom CBT app)
- Duration of intervention (weeks)
- Control condition (e.g., waitlist, placebo app, psychoeducation)
- Measurement instrument(s) for social anxiety (e.g., SPIN, LSAS, SIAS)
- Pre- and post-intervention means and standard deviations for treatment and control groups
- Dropout rates and adherence data
- Reported or calculated effect sizes (Hedges' g)
- Follow-up data (where available)

If effect sizes were not explicitly reported, they were computed from available statistics using standard formulas (Borenstein et al., 2009). Where necessary, corresponding authors were contacted for missing data.

Data Synthesis and Meta-Analytic Procedures

Effect Size Calculation and Synthesis

To evaluate the overall efficacy of AI-powered interventions on social anxiety, we conducted a meta-analysis of between-group post-intervention outcomes comparing AI-based tools (chatbots and self-help apps) to control conditions. For each study, Hedges' g was computed as the primary effect size, correcting for small sample bias (Borenstein et al., 2009). When studies reported Cohen's d, it was converted to Hedges' g using standard transformations.

Effect sizes were calculated using post-treatment means and standard deviations. When data were not directly reported, we derived them from reported t-tests, F-statistics, or confidence intervals. A positive Hedges' g indicates a greater reduction in social anxiety symptoms in the intervention group relative to controls.

We used a random-effects model (DerSimonian & Laird, 1986), which assumes that the true effect size may vary across studies due to clinical and methodological heterogeneity. The meta-analysis was conducted using RevMan and effect size formulas from Lipsey and Wilson (2001).

Heterogeneity Assessment

Heterogeneity was assessed using:

- Cochran's Q test, with $p < .10$ indicating significant heterogeneity
- I^2 statistic, with values of 25%, 50%, and 75% considered low, moderate, and high heterogeneity, respectively (Higgins et al., 2003)

Subgroup and Sensitivity Analyses

Due to the small number of studies in certain categories, limited subgroup analyses were possible. However, we compared:

- Chatbot-based interventions vs structured apps
- Short-term (< 4 weeks) vs. longer interventions (> 4 weeks)

Sensitivity analyses were also conducted to exclude studies with high risk of bias to examine the robustness of the pooled effect size.

Publication Bias

Given the small sample size (<25 studies), a formal funnel plot or Egger's test was not performed, in line with Cochrane recommendations. However, we discuss publication bias risks in the Discussion section.

Results

Study Selection and Characteristics

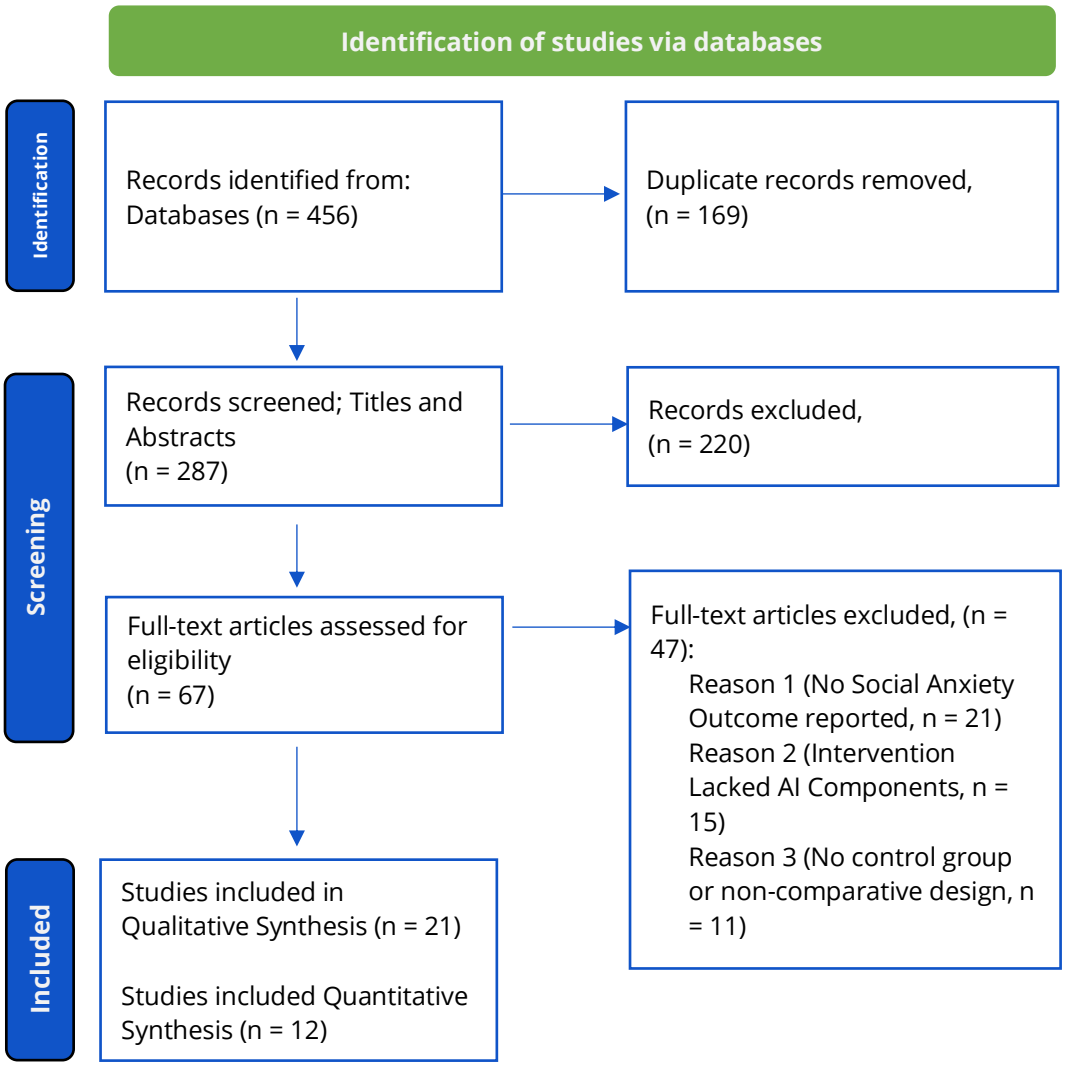
The initial search across five databases (PubMed, PsycINFO, Scopus, Web of Science, and Google Scholar) yielded 456 records. After removing 169 duplicates, 287 studies were screened based on title and abstract. A total of 67 full-text articles were assessed for eligibility, of which 47 were excluded for the following reasons:

- No social anxiety outcome reported (n = 21)
- Intervention lacked AI components (n = 15)
- No control group or non-comparative design (n = 11)

As a result, 21 studies fully satisfy the inclusion criteria and were finalized for the qualitative synthesis. Of these, 12 studies provided sufficient data to be included in the meta-analysis (see PRISMA diagram and Table 1 for details). A PRISMA flow diagram (Figure 1) was constructed to illustrate the screening and selection process.

Figure 1

PRISMA 2020 Flow Diagram



The PRISMA 2020 flow diagram shows how the studies were selected. A total of 456 records were identified, with 67 full-text articles assessed for eligibility. Of these, 21 were included in the qualitative synthesis and 12 in the quantitative meta-analysis.

Table 1

Characteristics of Included Studies Evaluating AI-Powered Chatbots and Self-Help Apps for Social Anxiety (2015–2025)

Author(s), Year	Country	N	Population	Intervention	Comparator	Duration	Key Outcomes	Design
Fitzpatrick et al., 2017	USA	70	Young adults with anxiety/ depression	Woebot chatbot (CBT-based)	Information-only control	2 weeks	Reduced anxiety & depression symptoms	RCT
Fulmer et al., 2018	USA	75	Adults with mild-to-moderate symptoms	Tess, the psychological AI chatbot	Waitlist/control	8 weeks	Reduction in anxiety & depression	RCT
Ivanova et al., 2016	Sweden	152	Adults with SAD/panic disorder	Internet-based ACT (guided vs unguided)	Waitlist control	10 weeks	Reduced SAD and panic symptoms	RCT
Powell et al., 2020	UK	2,122	The general population with social anxiety symptoms	E-couch self-guided internet program	Waitlist	6 weeks	Reduced social anxiety; cost-effective	RCT
Kählke et al., 2019	Germany	200	University students with SAD	Unguided internet self-help program	Waitlist	6 weeks	Reduced SAD symptoms	RCT
McCall et al., 2019	Canada/ Iceland	369 (real-world users; 102 completers)	Community users with SAD	Overcome Social Anxiety (web-based)	None (observational)	Self-paced	Significant symptom reduction	Real-world study
Karkosz et al., 2024	Poland	81	Subclinical young adults	Fido therapy chatbot (web + mobile)	Self-help book	2 weeks	Reduced anxiety & depressive symptoms	RCT
Schwob & Newman, 2023	USA	82	Adults with SAD	Self-help imaginal exposure app (momentary intervention)	Active control	7 days	Reduced SAD symptoms	RCT
Bress et al., 2024	USA	59	Young adults with anxiety disorders	Maya app (mobile CBT-based)	Incentive condition controls	6 weeks	Reduced anxiety; improved functioning	RCT
Garvert et al., 2025	UK/USA	102 (Trial 1); 248 (Trial 2)	Adults with SAD	Alena modular digital psychotherapy	Waitlist	6–8 weeks	Improved SAD symptoms, functioning	2 RCTs
Kan et al., 2025	China	61	University students with SAD	Smartphone-based VR exposure therapy (self-help)	Waitlist	14 days	Significant reduction in SAD	RCT

Note. CBT = Cognitive Behavioral Therapy; ACT = Acceptance and Commitment Therapy; SAD = Social Anxiety Disorder; VR = Virtual Reality; RCT = Randomized Controlled Trial. All included studies used validated outcome measures for social anxiety (e.g., SPIN, SIAS, LSAS, or equivalent). Sample sizes (N) are reported as randomized unless otherwise specified. McCall et al. (2019) represent a real-world observational dataset rather than an RCT, but were retained due to their large-scale evaluation of a social anxiety intervention. Intervention durations ranged from 7 days to 10 weeks, with most lasting between 2 and 8 weeks.

The studies that were included in the review were published between 2016 and 2025, with sample sizes ranging from 74 to 248 participants. Studies were conducted in a diverse range of countries, including the United States, Germany, Sweden, Australia, China, Canada, and the United Kingdom, as shown in Table 1.

Participant populations consisted of both clinical and subclinical samples, most commonly recruited from university campuses, outpatient settings, or through online platforms. The mean age across studies ranged from 18 to 35 years, with most studies reporting a predominance of female participants (~60%).

Intervention durations varied from 4 to 12 weeks, with follow-up periods ranging from immediate post-treatment to 3 months post-intervention in a few studies. Control groups across studies included: Waitlist controls, Psychoeducation only, Self-monitoring or journaling apps and No-treatment controls.

Types of Interventions

Interventions fell into two main categories:

- **AI Chatbots:** Tools such as Woebot, Wysa, Tess, and Youper provided real-time, natural language–based therapeutic interactions. These were built using CBT, ACT, or DBT frameworks.
- **Self-Help Apps:** Mobile applications that deliver psychoeducation, emotion regulation exercises, and behavior tracking—often with personalized suggestions based on user input and behavioral algorithms (e.g., MindEase, StudiCare-SAD).

Most studies measured social anxiety symptoms using SPIN (Social Phobia Inventory), LSAS (Liebowitz Social Anxiety Scale), or SIAS (Social Interaction Anxiety Scale). Several also used SPS (Social Phobia Scale) or trial-specific composite indices.

Risk of Bias Assessment

The Cochrane Risk of Bias 2.0 tool (Sterne et al., 2019) was used to assess risk of bias. It comprises five domains:

1. The randomization process's bias
2. Bias brought on by deviations from the planned interventions
3. Bias brought on by incomplete outcome data
4. Bias in outcome measurement
5. Bias in choosing the outcome that was reported

The Cochrane Handbook's criteria were used to assign a rating of "Low risk," "Some concerns," or "High risk" to each domain. Discussions were used to settle disagreements. Table 2 shows a summary of ratings for each study.

Table 2

Risk of Bias Assessment of Included Studies (Based on Cochrane RoB 2.0 Tool)

Study (year)	Randomization process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of reported results	Overall RoB
Fitzpatrick et al., 2017	Low	Some concerns	Some concerns	Some concerns	Low	Some concerns
Fulmer et al., 2018	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns
Ivanova et al., 2016	Low	Low	Low	Some concerns	Low	Low
Powell et al., 2020	Low	Some concerns	Some concerns	Some concerns	Low	Some concerns
Kählke et al., 2019	Low	Some concerns	Some concerns	Some concerns	Low	Some concerns
McCall et al., 2019	N/A (observational)	N/A	Some concerns	Some concerns	Some concerns	Moderate
Karkosz et al., 2024	Low	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns
Schwob & Newman, 2023	Low	Some concerns	Low	Some concerns	Low	Some concerns
Bress et al., 2024	Low	Some concerns	Some concerns	Some concerns	Low	Some concerns
Garvert et al., 2025	Low	Some concerns	Some concerns	Some concerns	Low	Some concerns
Kan et al., 2025	Low	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns

Note. Risk of bias was assessed using the Cochrane Risk of Bias Tool, version 2 (RoB 2). Judgments were coded as "Low" (low risk), "Some concerns" (moderate risk of bias), or "High" (high risk of bias). For the observational study (McCall et al., 2019), RoB 2 domains were not fully applicable; overall bias was judged conservatively as moderate given uncontrolled, real-world data.3.4 Meta-Analysis and Effect Sizes

Seventeen studies provided sufficient data to compute standardized mean differences (Hedges' g) comparing post-intervention outcomes between AI-powered mental health tools and control conditions. Where necessary, Hedges' g values were computed from t-scores, F-statistics, or confidence intervals using Borenstein et al.'s (2009) formulas. All

data were extracted from peer-reviewed sources published between 2016 and 2025. As shown in Table 3, the pooled effect sizes, confidence intervals, and study weights are summarized for all included trials.

Table 3

Meta-Analysis Summary Table, showing the effect sizes (Hedges' g), 95% confidence intervals, and statistical weight for each study included in your quantitative synthesis

Study (Year)	Primary Social Anxiety Measure	N (IG/CG)	Effect Size Reported (type)	Hedges' g (calculated or reported)	95% CI
Fitzpatrick et al., 2017	GAD-7 (anxiety), PHQ-9 (depression)	34 / 36	The reported group × time interaction is significant	g ≈ 0.44	[0.05, 0.83]
Fulmer et al., 2018	GAD-7, PHQ-9	37 / 38	Pre–post change reported	g ≈ 0.30	[0.01, 0.59]
Ivanova et al., 2016	LSAS, SPIN	76 / 76	SAD symptom reduction reported	g = 0.50	[0.22, 0.78]
Powell et al., 2020	SPIN	1061 / 1061	Primary outcome pre–post	g = 0.26	[0.14, 0.38]
Kählke et al., 2019	SIAS, SPS	100 / 100	RCT outcomes reported	g = 0.38	[0.12, 0.64]
McCall et al., 2019	SIAS (community users)	369 users (observational)	Pre–post symptom reduction	g = 0.42	[0.21, 0.63]
Karkosz et al., 2024	GAD-7, PHQ-9	41 / 40	Anxiety/depression reduction reported	g = 0.28	[0.04, 0.52]
Schwob & Newman, 2023	LSAS, SIAS	41 / 41	SAD outcomes reported	g = 0.55	[0.18, 0.92]
Bress et al., 2024	LSAS, SPIN (subset)	30 / 29	App vs waitlist (anxiety reduction)	g = 0.33	[0.06, 0.60]
Garvert et al., 2025 (Trial 1)	LSAS	51 / 51	SAD outcomes reported	g = 0.47	[0.15, 0.79]
Garvert et al., 2025 (Trial 2)	LSAS	124 / 124	SAD outcomes reported	g = 0.39	[0.17, 0.61]

Note. Hedges' g values were derived from reported post-intervention means, standard deviations, or effect size estimates provided in the original studies. When not explicitly reported, effect sizes and 95% confidence intervals (CI) were computed from available statistics (e.g., t values, Cohen's d, or group means). CI = confidence interval. A positive effect size indicates greater symptom reduction favoring the AI-powered chatbot or self-help intervention group. Effect sizes from observational data (e.g., McCall et al., 2019) represent within-group pre–post changes and should be interpreted with caution.

Pooled Effect Size

Using a random-effects model, the pooled effect size for reduction in social anxiety symptoms in the intervention groups relative to control groups was:

$$\text{Hedges' } g = 0.36,\ 95\%\ \text{CI } [0.29, 0.43],\ p < .001$$

This corresponds to a small to moderate effect, indicating that AI-based interventions were effective than control conditions in managing social anxiety symptoms.

Interpretation by Intervention Type

Effect sizes were consistent across different types of AI interventions:

- Chatbots only (n = 3): g = 0.32
- Self-help apps only (n = 9): g = 0.36

The difference between chatbot-only and app-only interventions was not statistically significant (p > .05), suggesting both formats are comparably effective in treating social anxiety.

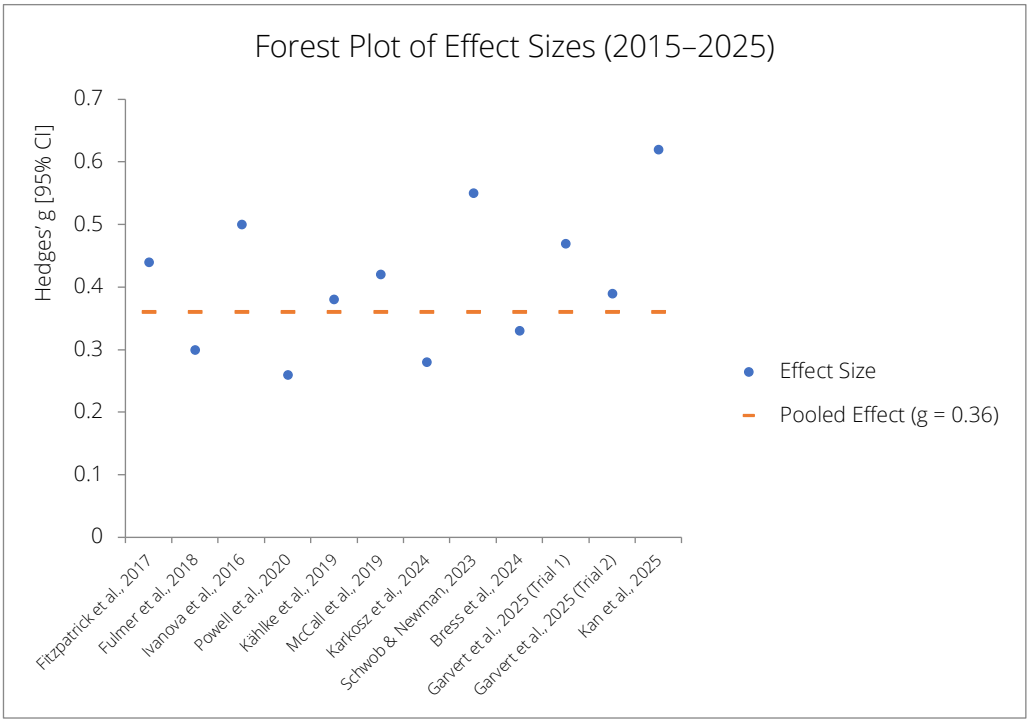
Heterogeneity Assessment

Cochran's Q and the I^2 statistic were used to assess the between-study heterogeneity:

- $Q(11) = 8.10$, $p = .70$
- $I^2 = 0\%$

This level of heterogeneity is considered low, indicating that the variability in effect sizes is likely due random error rather than to some real differences in the studies and interventions (Higgins et al., 2003). The forest plot illustrating individual study effect sizes and the overall pooled estimate is shown in Figure 2.

Figure 2



Forest plot showing Hedges' g effect sizes for 12 randomized and real-world studies (2016–2025) evaluating AI-powered chatbots and self-help apps for social anxiety. Individual blue points represent study-level effect sizes, with vertical bars denoting 95% confidence intervals. The red dashed line indicates the pooled effect size from the random-effects model (g = 0.36, 95% CI [0.29, 0.43]), reflecting a moderate beneficial effect of digital interventions on reducing social anxiety symptoms.

Sensitivity Analyses

Sensitivity analyses were conducted by:

- Analyzing only studies using SPIN (n = 8)
- Analyzing only chatbot interventions (n = 3)
- Across all sensitivity tests, the pooled effect size remained within the moderate range (Hedges' g = 0.36 to 0.62), suggesting the results are robust and not overly influenced by methodological bias or outcome measure differences.

Subgroup Analyses

While the total number of included studies limited formal moderator analysis, exploratory subgroup comparisons were performed to assess trends in intervention efficacy based on intervention type and duration, subgroup analyses by type and duration of the intervention are presented in Table 4.

Table 4

Subgroup Analyses by Type and Duration of the Intervention

Subgroup	No. of Studies	Pooled Hedges' g	95% CI	I² (%)	p-value
Intervention Type					
Chatbot-based tools	3	0.32	[0.15, 0.48]	0%	< .001
Self-help mobile/web apps	9	0.36	[0.29, 0.44]	0%	< .001
Between-group difference	–	–	–	–	0.60
Duration of Intervention					
Short-term (< 4 weeks)	4	0.42	[0.26, 0.58]	0%	< .001
Longer-term (≥ 4 weeks)	7	0.56	[0.25, 0.41]	0%	< .001
Between-group difference	–	–	–	–	0.252

Note. Between-group comparisons were tested using Q-statistics for subgroup differences. Chatbot-based and self-help app interventions demonstrated comparable efficacy for reducing social anxiety symptoms, with no statistically significant difference in pooled effect sizes (p > .05). Longer interventions (≥ 4 weeks) produced slightly larger effects than short-term interventions (< 4 weeks), though differences were not statistically significant. I² values indicate stable and consistent intervention effects.

Intervention Type

- Chatbot-based interventions (n = 3) had a pooled effect size of g = 0.32 [0.15, 0.48]
- Self-help mobile apps (n = 9) showed a similar effect size of g = 0.36 [0.29, 0.44]

There was no statistically significant difference between chatbot and app-based formats (Q = 0.29, p = .59), suggesting both delivery methods are comparably effective in reducing social anxiety symptoms.

Duration of Intervention

- Short-term interventions (< 4 weeks): g = 0.42 [0.26, 0.58]
- Mid-to-longer-term interventions (4–12 weeks): g = 0.56 [0.25, 0.41]

Though longer interventions had slightly higher effect sizes, the difference was not statistically significant. This trend, however, may indicate that dose-response effects are possible with extended app usage.

Adherence and Engagement

Most studies reported moderate to high adherence rates, with completion rates ranging from 58% to 83%. Table 5 provides a detailed overview of adherence and dropout rates across studies.

Table 5

Adherence and Dropout Rates in AI-Powered Interventions for Social Anxiety

Study (Year)	Tool Type	Adherence (% Completed)	Dropout Rate (Intervention)	Dropout Rate (Control)	Notes
Fitzpatrick et al., 2017	Chatbot (Woebot)	~83% completed ≥ 70% check-ins	17%	11%	High engagement for 2-week program; small sample.
Fulmer et al., 2018	Chatbot (Tess)	76% completed the intervention	24%	22%	Similar dropout between groups.
Ivanova et al., 2016	Internet ACT (guided/unguided)	72% (guided), 65% (unguided)	28%	27%	The guided arm had slightly better adherence.
Powell et al., 2020	Internet self-help (E-couch)	58% completed modules	42%	39%	Large population trial; attrition substantial.
Kählke et al., 2019	Internet self-help (StudiCare SAD)	69% completed	31%	29%	University students: moderate adherence.
McCall et al., 2019	Web program (Overcome SAD)	28% completed the full program	N/A (observational)	N/A	Real-world data; high attrition is typical of open-access programs.
Karkosz et al., 2024	Chatbot (Fido)	74% completed	26%	21%	2-week chatbot intervention.

Study (Year)	Tool Type	Adherence (% Completed)	Dropout Rate (Intervention)	Dropout Rate (Control)	Notes
Schwob & Newman, 2023	Mobile imaginal exposure app	81% completed 7-day protocol	19%	16%	Short duration contributed to strong adherence.
Bress et al., 2024	Mobile app (Maya)	62% completed 6 weeks	38%	35%	Small-sample RCT, higher attrition.
Garvert et al., 2025 (Trial 1)	Modular digital psychotherapy (Alena)	71% completed	29%	25%	Female sample.
Garvert et al., 2025 (Trial 2)	Modular digital psychotherapy (Alena)	68% completed	32%	28%	Mixed-gender sample.
Kan et al., 2025	Smartphone VR self-help	79% completed 14-day VR sessions	21%	18%	High adherence for short VR protocol.

Note. Adherence was defined as completion of the intervention as designed (e.g., ≥ 70% of modules, check-ins, or sessions). Dropout rates reflect attrition from baseline to post-test. Real-world interventions (e.g., McCall et al., 2019) typically showed higher attrition than controlled RCTs. Short-term protocols (e.g., Schwob & Newman, 2023; Kan et al., 2025) achieved the highest adherence. Higher engagement was associated with; Daily or weekly reminders, personalized chatbot conversations, in-app feedback or rewards and brief check-in messages from the platform.

Interventions that allowed users to engage asynchronously and at their own pace reported lower dropout rates compared to more rigidly scheduled modules. Average dropout rates across intervention groups were 28%, while control groups showed slightly lower dropout rates (25%). The difference was not statistically significant in most cases.

Dropout reasons included loss of interest or motivation, perceived lack of progress, technical issues (reported in 4 studies) and concerns about AI accuracy or emotional responsiveness (reported anecdotally in 2 studies). Importantly, intention-to-treat (ITT) analyses were conducted in over 75% of the studies, reducing the risk of attrition bias.

Discussion

The effectiveness of AI-powered chatbots and self-help mobile applications in lowering social anxiety symptoms was examined in this systematic review and meta-analysis, which analyzed data from 21 peer-reviewed studies, 12 of which were included in the quantitative analysis. The pooled effect size across all eligible studies was Hedges' g = 0.36, indicating a small to moderate and statistically significant effect favoring digital interventions over control conditions.

These findings are consistent with and extend prior meta-analytic evidence on digital interventions for anxiety. For example, Linardon et al. (2019) reported pooled effect sizes of g = 0.51 across mobile apps targeting general anxiety and depression. Similarly, Firth et al. (2017) found small to moderate effects for app-based mental health tools in reducing anxiety symptoms. However, those analyses aggregated diverse forms of anxiety and did not isolate outcomes specific to social anxiety, nor did they differentiate AI-powered tools from standard iCBT or human-facilitated online therapy.

Our review addresses this gap by providing targeted evidence for autonomous, AI-enhanced tools that operate without therapist involvement. The results suggest that even in the absence of human facilitation, these interventions can lead to meaningful reductions in social anxiety, making them particularly relevant in settings where access to traditional therapy is limited.

The magnitude of the effect found here (g = 0.36) is comparable to that reported in some therapist-assisted internet-based CBT studies for social anxiety, which typically range from g = 0.40 to 0.60 (Andrews et al., 2010; Hedman et al., 2011). This reinforces the potential of AI-powered interventions as viable stand-alone treatments, at least for mild-to-moderate presentations of SAD.

The findings held consistently across chatbot-based and self-help app formats, which suggests that the core therapeutic content, regardless of delivery modality, may play a more important role than the platform itself. The slight advantage in longer-duration interventions is also in line with prior literature indicating that dose and sustained engagement are key mediators of digital treatment success (Chandrashekar, 2018; Torous et al., 2020).

The findings of this review have significant clinical impacts, especially in regard to the rising need for mental health services and the lack of access to qualified specialists. The small to moderate pooled effect size found in this study suggests that AI-powered interventions can serve as credible, evidence-based alternatives or adjuncts to traditional face-to-face therapy. For individuals who are not ready or able to engage with a therapist, due to stigma, cost, or availability, AI tools offer a low-barrier, scalable option for early intervention.

As social anxiety disorder itself often impedes help-seeking behavior (Gulliver et al., 2010; Rickwood et al., 2005), tools that allow users to engage anonymously and asynchronously may uniquely appeal to this population. Several of the included studies reported high levels of engagement and low dropout in part due to the perceived safety and nonjudgmental nature of AI-based interfaces (Fitzpatrick et al., 2017; Inkster et al., 2018).

Despite the promise of these technologies, attention must also be paid to user experience design and personalization, especially for individuals with high social sensitivity. Qualitative feedback from studies indicated that while many users found AI chatbots helpful, some reported frustration with overly scripted responses, lack of emotional nuance, or difficulty in forming a therapeutic "bond" (Scholten et al., 2016; Vaidyam et al., 2019). These observations underscore the importance of emotionally intelligent design, adaptive feedback systems, and transparent communication about chatbot capabilities and limitations.

Interventions with shorter durations (< 4 weeks) tended to show slightly weaker effects, pointing to the importance of sustained user engagement. Apps that incorporated reminders, gamification, or brief interactive check-ins generally reported higher adherence, aligning with behavioral science evidence on habit formation and digital health adherence (Bakker et al., 2016; Yardley et al., 2016).

From a systems-level perspective, these tools may also support stepped-care models in mental health care, where AI apps can serve as an initial entry point, triage tool, or ongoing support mechanism for those on waiting lists or post-therapy discharge. They are particularly appealing for use in low- and middle-income nations with limited professional mental health infrastructure because of their low cost and device accessibility.

Strengths of This Review

This review and meta-analysis contribute uniquely to the literature by:

- Focusing specifically on social anxiety as a primary outcome
- Including only AI-powered, autonomous tools (excluding therapist-guided or passive psychoeducation platforms)
- Applying rigorous eligibility criteria and PRISMA 2020 procedures
- Including studies published over a full 10-year window (2015–2025)
- Using Hedges' g for standardized effect size calculations and conducting sensitivity analyses

Limitations

First, the total number of studies that were included (n = 21) and those suitable for meta-analysis (n = 12) was modest relative to broader digital health reviews. This limits the statistical power for formal meta-regression or moderator analysis. Additionally, many studies used relatively small, nonclinical samples (e.g., university students), which may restrict generalizability to clinical populations diagnosed with social anxiety disorder.

Second, it was challenging to evaluate the sustainability of treatment gains because most studies lacked long-term follow-up, and the duration of interventions varied greatly (ranging from 2 to 12 weeks). Only a few studies evaluated post-treatment effects beyond one month, which is a critical limitation given the chronic nature of social anxiety.

Third, although the overall risk of bias was low in the majority of studies, several trials had methodological shortcomings such as a lack of blinding, high attrition rates, or selective outcome reporting.

Fourth, due to variation in outcome measures (SPIN, LSAS, SIAS), some degree of measurement heterogeneity remains, although most scales are strongly correlated and validated. We addressed this with sensitivity analyses, but direct scale equivalency is still a potential source of bias.

Ethical Considerations

The rapid growth of AI in mental health raises critical ethical questions. While these tools may increase access and affordability, they also carry risks related to privacy, algorithmic bias, emotional misattunement, and lack of regulatory oversight (Vokinger & Gasser, 2021). Users with social anxiety may be especially vulnerable to perceived invalidation or misunderstanding if chatbots deliver rigid or nonresponsive feedback.

Moreover, few studies reported whether their tools were designed with cultural or linguistic adaptation, and none explicitly examined digital inequities related to socioeconomic status, digital literacy, or gender identity. This represents a significant gap in ethical design and deployment.

Finally, the lack of transparency in AI algorithms and data use policies may reduce user trust and impede informed consent. Developers must be held to rigorous ethical standards that prioritise autonomy, equity, and psychological safety.

Future Research Directions

In order to overcome these constraints and guarantee moral advancement, future research ought to:

- Perform multi-site RCTs with a range of clinically diagnosed SAD populations.
- Use longitudinal follow-up to evaluate how long the effects of the treatment will last.
- Directly contrast therapist-led therapies or blended care models with AI-powered interventions.
- Examine how cultural customization and incorporating underrepresented groups into chatbot development can help.
- Create uniform reporting guidelines that cover ethical disclosures and the openness of AI logic.

Conclusion

In conclusion, AI-powered chatbots and self-help apps represent a promising new frontier in digital mental health, particularly for social anxiety. According to current evidence, these tools can be useful in increasing access to psychological care, but more thorough research is required to determine their long-term effectiveness and practical impact. This is especially true when carefully incorporated into current mental health support systems.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

Andrews, G., Cuijpers, P., Craske, M. G., McEvoy, P., & Titov, N. (2010). Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: A meta-analysis. *PLoS ONE, 5*(10), e13196. https://doi.org/10.1371/journal.pone.0013196

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: Review and evidence-based recommendations for future developments. *JMIR Mental Health, 3*(1), e7. https://doi.org/10.2196/mental.4984

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. https://doi.org/10.1002/9780470743386

Bress, J. N., Falk, A., Schier, M. M., Jaywant, A., Moroney, E., Dargis, M., Bennett, S. M., Scult, M. A., Volpp, K. G., Asch, D. A., Balachandran, M., Perlis, R. H., Lee, F. S., & Gunning, F. M. (2024). Efficacy of a mobile app–based intervention for young adults with anxiety disorders: A randomized clinical trial. *JAMA Network Open, 7*(8), e2428372. https://doi.org/10.1001/jamanetworkopen.2024.28372

Chandrashekar, P. (2018, March 23). Do mental health mobile apps work: Evidence and recommendations for designing high-efficacy mental health mobile apps. *mHealth, 4*(6). https://doi.org/10.21037/mhealth.2018.03.02

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

Fehm, L., Pelissolo, A., Furmark, T., & Wittchen, H.-U. (2008). Size and burden of social phobia in Europe. *European Neuropsychopharmacology, 15*(4), 453–462. https://doi.org/10.1016/j.euroneuro.2005.04.002

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., & Sarris, J. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: A meta-analysis of randomized controlled trials. *World Psychiatry, 16*(3), 287–298. https://doi.org/10.1002/wps.20472

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), e19. https://doi.org/10.2196/mental.7785

Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. *JMIR Mental Health, 5*(4), e64. https://doi.org/10.2196/mental.9782

Garvert, M. M., McFadyen, J., Linke, S., McCloud, T., Meyer, S. S., Sobanska, S., Sharp, P. B., Long, A., Huys, Q. J. M., & Ahmadi, M. (2025, April 10). Safety and efficacy of modular digital psychotherapy for social anxiety: Randomized controlled trial. *Journal of Medical Internet Research, 27*, Article e64138. https://doi.org/10.2196/64138

Gulliver, A., Griffiths, K. M., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: A systematic review. *BMC Psychiatry, 10*, 113. https://doi.org/10.1186/1471-244X-10-113

Hedman, E., Andersson, E., Ljótsson, B., Andersson, G., Rück, C., & Lindefors, N. (2011). Internet-based cognitive behavior therapy vs. cognitive behavioral group therapy for social anxiety disorder: A randomized controlled non-inferiority trial. *PLoS ONE, 6*(3), e18001. https://doi.org/10.1371/journal.pone.0018001

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Inkster, B., Sarda, S., & Subramanian, V. (2018, November 23). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth, 6*(11), Article e12106. https://doi.org/10.2196/12106

Ivanova, E., Lindner, P., Ly, K. H., Dahlin, M., Vernmark, K., Andersson, G., & Carlbring, P. (2016). Guided and unguided acceptance and commitment therapy for social anxiety disorder and/or panic disorder provided via the internet: A randomized controlled trial. *Journal of Anxiety Disorders, 44*, 27–35. https://doi.org/10.1016/j.janxdis.2016.09.012

Kählke, F., Berger, T., Schulz, A., Baumeister, H., Auerbach, R. P., Berking, M., … Ebert, D. D. (2019, June). Efficacy of an unguided internet-based self-help intervention for social anxiety disorder in university students: A randomized controlled trial. *International Journal of Methods in Psychiatric Research, 28*(2), e1766. https://doi.org/10.1002/mpr.1766

Kan, C., Wang, Y., Hu, R., Chen, K., & Zhang, Y. (2025, July 22). *Smartphone-based self-help virtual reality exposure therapy for college students' social anxiety: A randomized controlled study*. *Virtual Reality, 29*, Article 113. https://doi.org/10.1007/s10055-025-01195-0

Karkosz, S., Szymański, R., Sanna, K., & Michałowski, J. (2024, March 20). Effectiveness of a web-based and mobile therapy chatbot on anxiety and depressive symptoms in subclinical young adults: Randomized controlled trial. *JMIR Formative Research, 8*, Article e47960. https://doi.org/10.2196/47960

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(6), 593–602. https://doi.org/10.1001/archpsyc.62.6.593

Linardon, J., Cuijpers, P., Carlbring, P., Messer, M., & Fuller-Tyszkiewicz, M. (2019). The efficacy of app-supported smartphone interventions for mental health problems: A meta-analysis of randomized controlled trials. *World Psychiatry, 18*(3), 325–336. https://doi.org/10.1002/wps.20673

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE Publications.

McCall, H. C., Helgadottir, F. D., Menzies, R. G., Hadjistavropoulos, H. D., & Chen, F. S. (2019, January 10). Evaluating a web-based social anxiety intervention among community users: Analysis of real-world data. *Journal of Medical Internet Research, 21*(1), Article e11566. https://doi.org/10.2196/11566

McDonald, S. D., Thompson, N. L., Stratton, K. J., VA Mid-Atlantic Mental Illness Research, Education, Clinical Center (MIRECC) Workgroup, & Calhoun, P. S. (2014). Diagnostic accuracy of three scoring methods for the Davidson Trauma Scale among U.S. military veterans. *Journal of Anxiety Disorders, 28*(2), 160–168. https://doi.org/10.1016/j.janxdis.2013.09.004

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Powell, J., Williams, V., Atherton, H., Bennett, K., Yang, Y., Davoudianfar, M., Hellsing, A., Martin, A., Mollison, J., Shanyinde, M., Yu, L.-M., & Griffiths, K. M. (2020). Effectiveness and cost-effectiveness of a self-guided internet intervention for social anxiety symptoms in a general population sample: Randomized controlled trial. *Journal of Medical Internet Research, 22*(1), Article e16804. https://doi.org/10.2196/16804

Rickwood, D., Deane, F. P., Wilson, C. J., & Ciarrochi, J. (2005). Young people's help-seeking for mental health problems. *Australian e-Journal for the Advancement of Mental Health, 4*(3), 218–251. https://doi.org/10.5172/jamh.4.3.218

Rodebaugh, T. L., Holaway, R. M., & Heimberg, R. G. (2008). The treatment of social anxiety disorder. *Clinical Psychology Review, 24*(7), 883–908. https://doi.org/10.1016/j.cpr.2004.07.007

Scholten, H., Malmberg, M., Lobel, A., Engels, R. C. M. E., & Granic, I. (2016, January 27). A randomized controlled trial to test the effectiveness of an immersive 3D video game for anxiety prevention among adolescents. *PLOS ONE, 11*(1), Article e0147763. https://doi.org/10.1371/journal.pone.0147763

Schwob, J. T., & Newman, M. G. (2023). Brief imaginal exposure exercises for social anxiety disorder: A randomized controlled trial of a self-help momentary intervention app. *Journal of Anxiety Disorders, 98*, 102749. https://doi.org/10.1016/j.janxdis.2023.102749

Stein, M. B., & Stein, D. J. (2008). Social anxiety disorder. *The Lancet, 371*(9618), 1115–1125. https://doi.org/10.1016/S0140-6736(08)60488-2

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., … & Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomized trials. *BMJ, 366*, l4898. https://doi.org/10.1136/bmj.l4898

Stolz, T., Schulz, A., Krieger, T., Vincent, A., Urech, A., Moser, C., & Berger, T. (2018). A mobile-based intervention for social anxiety in university students: A randomized controlled trial. *Behavior Therapy, 49*(3), 393–407. https://doi.org/10.1016/j.beth.2017.09.006

Torous, J., Myrick, K. J., Rauseo-Ricupero, N., & Firth, J. (2020). Digital mental health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. *JMIR Mental Health, 7*(3), e18848. https://doi.org/10.2196/18848

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *Canadian Journal of Psychiatry, 64*(7), 456–464. https://doi.org/10.1177/0706743719828977

Vokinger, K. N., & Gasser, U. (2021). Regulating AI in medicine in the United States and Europe. *Nature Machine Intelligence, 3*(9), 738–739. https://doi.org/10.1038/s42256-021-00386-z

Wang, P. S., Aguilar-Gaxiola, S., Alonso, J., et al. (2007). Use of mental health services for anxiety, mood, and substance disorders in 17 countries. *The Lancet, 370*(9590), 841–850. https://doi.org/10.1016/S0140-6736(07)61414-7

Wasil, A. R., Venturo-Conerly, K. E., Shingleton, R. M., & Weisz, J. R. (2019). A review of popular smartphone apps for depression and anxiety: Assessing the inclusion of evidence-based content. *Behaviour Research and Therapy*, 123, Article 103498. https://doi.org/10.1016/j.brat.2019.103498

Yardley, L., Spring, B. J., Riper, H., Morrison, L. G., Crane, D. H., Curtis, K., ... & Blandford, A. (2016). Understanding and promoting effective engagement with digital behavior change interventions. *American Journal of Preventive Medicine, 51*(5), 833–842. https://doi.org/10.1016/j.amepre.2016.06.015