ISRCTN11343758 https://doi.org/10.1186/ISRCTN11343758

Can artificial intelligence provide effective feedback in online therapy for social anxiety?

Submission date	Recruitment status	[ ] Prospectively registered
08/04/2026	Recruiting	[ ] Protocol
Registration date	**Overall study status**	[ ] Statistical analysis plan
09/04/2026	Ongoing	[ ] Results
Last Edited	**Condition category**	[ ] Individual participant data
09/04/2026	Mental and Behavioural Disorders	[X] Record updated in last year

Plain English summary of protocol

Background and study aims
Social anxiety disorder affects about 12% of the population. People with the condition fear being judged or embarrassed in social situations, and this often interferes with work, education, and relationships. Waiting lists for treatment in Sweden currently exceed six months.

Internet-based cognitive behavioural therapy (ICBT) is an online treatment that has been shown to work as well as face-to-face therapy for social anxiety. The main limitation is that each patient needs regular written feedback from a therapist, which limits how many patients a therapist can treat at once. AI systems can now produce therapeutic feedback that trained clinicians have difficulty telling apart from human-written feedback in blinded evaluations.

This study tests whether AI-generated feedback is at least as good as human therapist feedback in ICBT for social anxiety (a non-inferiority design). A second question is whether what patients believe about who wrote their feedback matters for how well the treatment works, independent of who actually wrote it. The study also compares the therapeutic relationship, engagement, and negative effects across conditions.

Who can participate?
Adults aged 18 or older with social anxiety symptoms.

What does the study involve?
Participants are randomly assigned to one of six groups. The treatment is delivered online through the Iterapi platform (managed at Linkoeping University). Participants take part from home on their own computer or device. Four groups get the online treatment with weekly written feedback. In two of these, the feedback is generated by AI and checked by a psychologist before delivery. In the other two, a human therapist writes the feedback. To test whether expectations matter, some participants are told their feedback comes from AI when it actually comes from a human, or the other way around. A fifth group gets the same treatment programme without any personalised feedback. A sixth group waits and is offered treatment afterwards.

The treatment runs for nine weeks. It covers understanding social anxiety, identifying and changing unhelpful thoughts, gradually facing feared situations, reducing self-focused attention, and planning for the future. Participants fill in questionnaires at 12 time points: before treatment, weekly during the nine-week treatment, at the end of treatment, and again at 3 and 6 months (spanning approximately 8 months in total). Anyone whose feedback was attributed to a different source than the real one is told the truth after treatment.

What are the possible benefits and risks of participating?
Participants in the treatment groups may see their social anxiety symptoms decrease. In 15 previous trials using this same treatment programme (with over 2,400 participants), improvements have been large and lasted up to five years. Not everyone will benefit. Some people experience a temporary increase in anxiety during the exposure exercises; this is expected and usually passes. Participants who find out after treatment that their feedback came from a different source than they were told may react to this; a psychologist is available during the debriefing. The waiting list group does not receive treatment during the study but gets the same programme after.

Where is the study run from?
Stockholm University, Department of Psychology, Sweden.

When is the study starting and how long is it expected to run for?
Recruitment started in September 2025 and is expected to run until May 2026. Treatment is nine weeks long. Follow-up assessments are at 3 and 6 months after treatment ends. The last data collection is expected in February 2027.

Who is funding the study?
This study is investigator initiated and funded.

Who is the main contact?
Professor Per Carlbring
Department of Psychology, Stockholm University
Email: per.carlbring@psychology.su.se

Contact information

Type(s)
Public, Scientific, Principal investigator

Contact name
Prof Per Carlbring

ORCID ID
https://orcid.org/0000-0002-2172-8813

Contact details
Department of Psychology, Stockholm University, Stockholm, Sweden
Stockholm
Sweden
10691
+468163920
per.carlbring@psychology.su.se

Additional identifiers

Study information

Scientific Title
PSY-CHAI: a six-arm double-blind randomized controlled non-inferiority trial with factorial deception design comparing ai-generated and human therapist feedback in internet-based cognitive behavioural therapy for adults with social anxiety disorder

Acronym
PSY-CHAI

Study objectives
1. To test whether AI-generated therapist feedback produces non-inferior outcomes compared to human therapist feedback on social anxiety symptoms in internet-based cognitive behavioural therapy (ICBT) for social anxiety disorder, using a pre-specified non-inferiority margin of 0.30 standard deviations on the Social Phobia Inventory (SPIN) total score at post-treatment (one-sided alpha = 0.025)
2. To decompose treatment effects into the contributions of actual feedback quality and patients' beliefs about the feedback source, using a 2x2 factorial design crossing actual source (AI-generated versus human) with perceived source (told AI versus told human)
3. To evaluate treatment process variables including therapeutic alliance trajectories, treatment credibility, negative effects, and treatment engagement across conditions

Ethics approval required
Ethics approval required

Ethics approval(s)
approved 17/03/2025, Swedish Ethical Review Authority (Etikproevningsmyndigheten), Gothenburg Division 1 Medicine (Box 2110, Uppsala, 750 02, Sweden; +46 10 475 08 00; registrator@etikprovning.se), ref: 2025-01058-01

Primary study design
Interventional

Allocation
Randomized controlled trial

Masking
Blinded (masking used)

Control
Active

Assignment
Parallel

Purpose
Treatment

Study type(s)

Health condition(s) or problem(s) studied
Social anxiety disorder

Interventions
Participants are randomized to one of six parallel groups. Four active groups form a 2x2 factorial design: (1) AI-generated feedback, participants told AI (transparent AI, n=80); (2) AI-generated feedback, participants told human therapist (deceptive AI, n=80); (3) human therapist feedback, participants told human therapist (transparent human, n=80); (4) human therapist feedback, participants told AI (deceptive human, n=80). Two control groups complete the design: (5) unguided self-help with full treatment content but no personalized feedback (n=80); (6) waitlist control (n=80).

All active groups (1-5) receive the 9-week SOFIE internet-based cognitive behavioural therapy protocol delivered via the Iterapi treatment platform. The protocol is based on the Clark and Wells cognitive model and consists of nine modules: psychoeducation, the cognitive model of social anxiety, cognitive restructuring, behavioural experiments, graded exposure (two modules), attention and self-focus training, social skills, and relapse prevention. Treatment content is identical across groups 1-5.

Groups 1-4 receive personalized asynchronous written feedback from their therapist 1-2 times per week (about 200-350 words per message). For the AI groups, a multi-agent system generates candidate feedback drawing on three knowledge sources: the current treatment module manual, the participant's symptom trajectory (weekly SPIN scores with clinical thresholds), and the conversation and homework history. A licensed psychologist reviews and approves every AI-generated message before it is delivered to the participant. For the human groups, psychologists or clinical psychology student's under supervision write feedback following the same clinical guidelines to ensure comparability.

Randomization uses permuted blocks of unknown size, automated through the Iterapi platform, to ensure approximately equal group sizes throughout recruitment. The waitlist group receives no treatment during the study period and is offered treatment after the waitlist period ends.

The informed consent procedure discloses that the study involves multiple conditions, some of which include deception regarding the feedback source, without revealing individual allocation. All participants in deception conditions are fully debriefed on demand at the conclusion of the treatment period.

Assessments are conducted at 12 time points across 6 months: Day 0 (Baseline), Day 7 (Week 1), Day 14 (Week 2), Day 21 (Week 3), Day 28 (Week 4), Day 35 (Week 5), Day 42 (Week 6), Day 49 (Week 7), Day 56 (Week 8), Day 64 (Post-treatment, week 9), Day 154 (3-month follow-up), Day 244 (6-month follow-up).

The sample size of 480 (80 per group) provides >85% power for the primary non-inferiority comparison (n=160 AI-guided vs n=160 human-guided, margin 0.30 SD) using linear mixed models with 12 repeated measures, >99% power for active treatment versus waitlist, and approximately 96% power for guided versus self-help (expected d = 0.46).

Intervention Type
Behavioural

Primary outcome(s)
1. Social anxiety symptoms measured using the Social Phobia Inventory (SPIN) measured using Self-report at Social anxiety symptoms measured using the Social Phobia Inventory (SPIN) at baseline, weekly during treatment (weeks 1-8), post-treatment (week 9), and at 3-month and 6-month follow-up

Key secondary outcome(s))
1. Therapeutic alliance measured using the Session Alliance Inventory (SAI) at weeks 1-8 (weekly during treatment)

2. Depressive symptoms measured using the Patient Health Questionnaire-9 (PHQ-9), and the Patient Health Questionnaire-2 (PHQ-2) at 3- and 6-month follow-ups, and weeks 1-8 (weekly during treatment)

3. Generalized anxiety measured using the Generalized Anxiety Disorder-7 (GAD-7), and the Generalized Anxiety Disorder-2 (GAD-2) at baseline, post-treatment, and at 3-and 6-month follow-ups and weeks 1-8 (weekly during treatment)

4. Social anxiety severity measured using the Liebowitz Social Anxiety Scale, Self-Report version (LSAS-SR) at baseline

5. Negative effects of treatment measured using the Negative Effects Questionnaire (NEQ-20) at post-treatment

6. Treatment credibility and expectancy measured using the Credibility/Expectancy Questionnaire (CEQ) with parallel AI and human versions at baseline

7. Treatment engagement measured using module completion rates, number of logins, worksheets completed, and messages sent, recorded throughout at the 9-week treatment period

Completion date
01/02/2027

Eligibility

Key inclusion criteria
1. Age 18 years or older
2. Primary diagnosis of social anxiety disorder according to DSM-5 criteria, assessed via structured self-report screening
3. Score of 30 or above on the Liebowitz Social Anxiety Scale, Self-Report version (LSAS-SR), indicating at least moderate social anxiety
4. Stable psychotropic medication for at least 3 months, or no psychotropic medication
5. No concurrent psychological treatment
6. Ability to read and write Swedish
7. Access to the internet

Healthy volunteers allowed
No

Age group
Mixed

Lower age limit
18 years

Upper age limit
120 years

Sex
All

Total final enrolment
0

Key exclusion criteria
1. Severe depression as indicated by a score of 20 or above on the Patient Health Questionnaire-9 (PHQ-9)
2. Active suicidal ideation as indicated by PHQ-9 item 9 > 2 points at screening
3. Non-serious or invalid responses at clinical screening

Date of first enrolment
01/09/2025

Date of final enrolment
30/05/2026

Locations

Countries of recruitment
Sweden

Sponsor information

Organisation
Stockholm University

ROR
https://ror.org/05f0yaq80

Funder(s)

Funder type

Funder Name

Investigator initiated and funded

Results and Publications

Individual participant data (IPD) sharing plan
Available upon request. Contact: Per Carlbring, per.carlbring@psychology.su.se. De-identified participant-level data will be made available to qualified researchers upon reasonable request, following study completion and publication of the primary results.

IPD sharing plan summary
Available on request