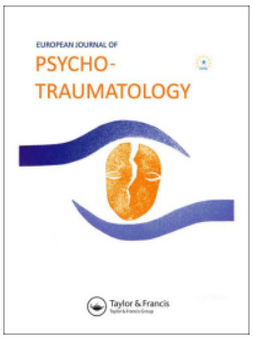

European Journal of Psychotraumatology

ISSN: 2000-8066 (Online) Journal homepage: www.tandfonline.com/journals/zept20

Deepfake perpetrator conversation for adults with sexual abuse-related posttraumatic stress disorder: intervention development and multiple baseline study protocol

Jackie June ter Heide, Iva Bicanic, Brieneke Holtrop , Larissa Maassen , Inge van Nistelrooij, Maryëlle Hoogeveen, Mathilde Overbeek, Eline Voorendonk, Sven Zebel, Theo Gevers & Agnes van Minnen

To cite this article: Jackie June ter Heide, Iva Bicanic, Brieneke Holtrop , Larissa Maassen , Inge van Nistelrooij, Maryëlle Hoogeveen, Mathilde Overbeek, Eline Voorendonk, Sven Zebel, Theo Gevers & Agnes van Minnen (2026) Deepfake perpetrator conversation for adults with sexual abuse-related posttraumatic stress disorder: intervention development and multiple baseline study protocol, European Journal of Psychotraumatology, 17:1, 2668166, DOI: 10.1080/20008066.2026.2668166

To link to this article: https://doi.org/10.1080/20008066.2026.2668166

Published online: 26 May 2026.

EUROPEAN JOURNAL OF PSYCHOTRAUMATOLOGY
2026, VOL. 17, NO. 1, 2668166
https://doi.org/10.1080/20008066.2026.2668166

STUDY PROTOCOL

OPEN ACCESS Check for updates

Deepfake perpetrator conversation for adults with sexual abuse-related posttraumatic stress disorder: intervention development and multiple baseline study protocol

Jackie June ter Heide [a], Iva Bicanic [b], Brieneke Holtrop[c], Larissa Maassen[a], Inge van Nistelrooij [d], Maryëlle Hoogeveen [e], Mathilde Overbeek [a], Eline Voorendonk [e], Sven Zebel [f], Theo Gevers [g] and Agnes van Minnen [h]

[a]ARQ Centrum'45, Diemen, The Netherlands; [b]National Psychotrauma Center for Children and Youth, Utrecht University Medical Center, Utrecht, The Netherlands; [c]Independent expert by experience, Nijmegen, The Netherlands; [d]Independent care ethicist, Gouda, The Netherlands; [e]Research Department PSYTREC, Zeist, The Netherlands; [f]Department of Psychology of Conflict, Risk and Safety, Faculty of Behavioural, Management and Social Sciences, University of Twente, and Dutch Private Law, Faculty of Law, Enschede, The Netherlands; [g]University of Amsterdam, Amsterdam, The Netherlands; [h]Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

ABSTRACT

Background: A conversation between a victim and a perpetrator of sexual abuse has the potential to reduce posttraumatic stress disorder (PTSD) symptoms in the victim. However, an actual conversation may not always be possible or feasible. Deepfake technology may offer a way to simulate a conversation in a clinical context. Therefore, a feasible, acceptable and potentially effective intervention protocol is required.

Objective: We aimed to develop an intervention protocol for a deepfake victim-perpetrator conversation to be used in the treatment of patients with sexual abuse-related PTSD, that would be technologically, legally, ethically, and clinically feasible as well as acceptable. In addition, we aimed to design a single-case experimental design (SCED) multiple baseline study to evaluate its efficacy.

Method: The intervention protocol was developed through an iterative multidisciplinary process following established guidelines for complex intervention development. Based on a literature review and interviews with 15 experts and patient representatives, a draft intervention protocol was created and subsequently evaluated, piloted and refined. A consequent study protocol was then designed in accordance with the SPIRIT guidelines.

Results: The development process resulted in an intervention protocol for a half-day perpetrator conversation using deepfake technology. The intervention is grounded in cognitive–behavioural therapy and emotional processing theory, and consists of a preparatory session, a deepfake session, and a debriefing session, to be conducted by two therapists. The intervention is tested in a SCED multiple baseline study of 10 patients with sexual abuse-related PTSD who are randomly assigned to a baseline period of 10, 15 or 20 days. Outcomes (changes in negative cognitions, PTSD, guilt and shame, forgiveness, and empowerment) are assessed through 20–30 daily measurements and four main assessments, and analysed using Bayesian statistics.

Conclusion: A feasible and acceptable intervention protocol for a deepfake intervention was developed. Its efficacy is tested in a multiple-baseline study.

Conversación simulada con un perpetrador para adultos con trastorno de estrés postraumático relacionado con abuso sexual: desarrollo de la Intervención y protocolo de estudio de línea de base múltiple

Objetivo: El objetivo fue desarrollar un protocolo de intervención para una conversación simulada entre víctima y perpetrador para ser utilizada en el tratamiento de pacientes con TEPT relacionado con abuso sexual, que pudiera se factible tecnológica, legal, ética y clínicamente, así como también aceptable. Además, nos propusimos diseñar un estudio de línea de base múltiple con diseño experimental de caso único (SCED por sus siglas en inglés) para evaluar su eficacia.

Método: El protocolo de intervención se desarrolló a través de un proceso multidisciplinario iterativo siguiendo las guías establecidas para el desarrollo de intervenciones complejas. Basados en una revisión de la literatura y entrevistas con 15 expertos y representantes de pacientes, se creó un borrador del protocolo de intervención que posteriormente se evaluó, se piloteó y se perfeccionó. A continuación, se diseñó un protocolo de estudio definitivo de acuerdo con las directrices SPIRIT.

Resultados: El proceso de desarrollo dio como resultado en un protocolo de intervención para una conversación de medio día con el perpetrador utilizando tecnología *deepfake*. La

ARTICLE HISTORY
Received 3 April 2026
Accepted 15 April 2026

KEYWORDS
Deepfake; posttraumatic stress disorder; PTSD; sexual abuse; sexual violence; feasibility; acceptability; intervention development; study protocol; artificial intelligence

PALABRAS CLAVE
Deepfake; trastorno de estrés postraumático; TEPT; abuso sexual; violencia sexual; factibilidad; aceptabilidad; desarrollo de una intervención; protocolo de estudio; inteligencia artificial

HIGHLIGHTS

- We developed an intervention protocol for a deepfake perpetrator conversation for patients with sexual abuse-related PTSD.
- The intervention protocol appeared ethically, legally, and clinically feasible, as well as acceptable to patients with PTSD.
- Its efficacy is examined in a multiple baseline study.

CONTACT Jackie June ter Heide j.ter.heid@ecentrum45.nl ARQ Centrum'45, Nienoord 5, Diemen 1112 XE, The Netherlands

intervención se basa en la terapia cognitivo conductual y la teoría del procesamiento emocional y consiste en una sesión preparatoria, una sesión *deepfake* y una sesión evaluación, dirigida por dos terapeutas. La intervención se prueba en un estudio de línea de base múltiple SCED para 10 pacientes con TEPT relacionado con abuso sexual que son asignados en forma aleatoria a un periodo basal de 10,15 o 20 días. Los resultados (cambios en las cogniciones negativas, TEPT, la culpa y la vergüenza, el perdón y el empoderamiento) se evalúan mediante 20–30 mediciones diarias y cuatro evaluaciones principales y se analizan utilizando estadística Bayesiana.
Conclusión: Se desarrolló un protocolo de intervención factible y aceptable para una intervención con *deepfake*. Su eficacia se evaluó en un estudio de línea de base múltiple.

1. Introduction

Generative artificial intelligence (AI) is an increasing part of public and daily life. While some uses of generative AI are potentially psychologically harmful (e.g. Kraaijeveld & Ivanova, 2026), other uses may potentially contribute to psychological wellbeing and mental health (e.g. Becker & Laycock, 2023). Within mental health care, the use of interventions facilitated by generative AI is still limited (Craig et al., 2018; Petkovski et al., 2026; Van Minnen et al., 2022). The responsible development and evaluation of clinical interventions that use generative AI is therefore an important matter. In this paper, we describe the development and study protocol of a deepfake perpetrator conversation for patients with posttraumatic stress disorder (PTSD; American Psychiatric Association [APA], 2013) after sexual abuse.

1.1. Posttraumatic stress disorder after sexual abuse

Exposure to sexual abuse, defined as completed or attempted forced penetration, unwanted sexual contact, or non-contact unwanted sexual experiences (Basile et al., 2014), is a widespread potentially traumatic event (Dworkin et al., 2021). A global analysis shows that in 2023, around 19% of girls and 15% of boys ages 0–18 were sexually abused (Cagney et al., 2025). Most victims are abused by male perpetrators (Cortoni et al., 2017), who are usually known to their victims (Centraal Bureau voor de Statistiek, 2024). Almost half (49%) of rape victims and almost a quarter (24%) of sexual assault victims develop PTSD (Breslau et al., 1998).

Among those who develop PTSD after sexual abuse, the symptom cluster of negative alterations in cognitions and mood (NACM) appears central to their symptomatology (McBride et al., 2020). This cluster involves, among other symptoms, strong negative beliefs about oneself, distorted cognitions about being to blame, and strong negative feelings such as anger, guilt and shame. NACM symptoms such as negative cognitions and blame play a role in the onset, maintenance, and reduction of PTSD (Kip et al., 2022; Kooistra et al., 2023), underscoring the relevance of targeting these symptoms in psychological treatment.

Trauma-focused psychological treatments are generally effective in treating adults with PTSD (Hoppen et al., 2023). However, residual symptoms are common (Semmlinger et al., 2024). Among patients treated for rape-related PTSD, symptoms from the NACM cluster, including detachment, self-blame, and guilt related to lack of justification for actions, are among the most common residual symptoms (Larsen et al., 2019). Consequently, there is a need for additional interventions to alleviate these symptoms and aid in further psychological recovery.

1.2. Perpetrator conversation

In addition to trauma-focused treatment, clinicians sometimes recommend that patients have a conversation with the perpetrator (e.g. Schouten, 2016). Such a conversation may enable victims to relay the impact of the sexual abuse, request acknowledgement of the abuse, and ask questions. These conversations may take place within the treatment room or within a formal context of restorative justice.

Restorative justice is a criminological approach that prioritises restoration of both victim and offender over retribution in response to crime. It can be helpful for victims who wish to communicate the impact of the sexual abuse to the perpetrator and gain a better understanding of why the sexual abuse was committed (Bolívar, 2013). Restorative justice interventions include mediation through one-way communication, indirect or direct mediation, and conferencing (Sherman & Strang, 2010). These interventions may promote the psychological wellbeing of victims (Nascimento et al., 2023), leading to an increased sense of justice and empowerment (Hansen & Umbreit, 2018) and trust in self and others (Sherman & Strang, 2010). In addition, they may decrease feelings of anger, anxiety and guilt (Lloyd & Borrill, 2020; Zebel, 2012), negative cognitions about the world (Lloyd & Borrill, 2020), and fear of revictimization and desire for revenge (Sherman et al., 2015). While these effects are mainly related to the NACM

cluster, restorative justice interventions can also reduce the severity of intrusions, avoidance, and heightened threat (Lloyd & Borrill, 2020).

Although restorative justice interventions may be helpful, actual communication may not always be possible. At some providers, only 1 in 3 applications concerning sexual offence cases leads to actual contact (Perspectief Herstelbemiddeling (2020). Offenders' contact details may not always be made available by the public prosecutor's office, offenders may not be ready for contact or may have died, or the mediator may feel it is better to abstain from mediation (e.g. Bolívar, 2013; Hansen & Umbreit, 2018). Within a therapeutic setting, therapists may then resort to techniques in which contact is imaginary, such as empty chair techniques. However, not all patients may be able to benefit from these techniques as it requires the conjuring of vivid mental imagery (e.g. King et al., 2024).

1.3. Deepfake perpetrator conversation

In recent years, a technique has become available that enables patients to conduct an imaginary perpetrator conversation using generative AI. In this technique, a frontal photo of the perpetrator is converted into a deepfake animation that can be used to simulate interpersonal interactions. During the session the patient, seated behind a laptop, may conduct a conversation with a 'deepfaked' perpetrator, voiced by a therapist who is seated behind a laptop in another room (see Petkovski et al., 2026, for a visualisation).

The first phase in the development of this technique involved an exploration of its feasibility in two patients with residual PTSD symptoms following trauma-focused treatment (Van Minnen et al., 2022). In this phase, a preliminary, brief intervention protocol was used. Development then moved on to a second phase aimed at developing a full intervention protocol that would be feasible, acceptable, and potentially efficacious. Subsequently, a multiple baseline study was designed aimed at evaluating the intervention's safety and efficacy. In this paper, we describe the development of the intervention protocol as well as the design of the study protocol.

2. Intervention development

We developed an intervention protocol, guided by the framework for the development of complex healthcare interventions (O'Cathain et al., 2019). In accordance with this guidance, the following steps were taken (not necessarily in this order): (1) planning the development process, (2) involving stakeholders, (3) bringing together a team and establishing a decision-making process, (4) reviewing published research evidence, (5) drawing on existing theories, (6) articulating a programme theory, (7) undertaking primary data collection, (8) understanding context; (9) paying attention to future implementation, and (10) designing and refining the intervention.

Steps 7 and 10 were approved by the Ethics Review Board of the Faculty of Social & Behavioural Sciences of Utrecht University, #2024-0010, on February 1, 2024. Participants were informed about the project through a participant information letter and signed an informed consent form.

2.1. Procedure

2.1.1. Step 1, planning the development process

The development process included funding application, legal and faculty ethics review, and intervention development, and was conducted from June 2022 until April 2025.

2.1.2. Step 2, involving stakeholders

Stakeholders were involved as either project team members or advisors. The project team consisted of three scientist-practitioners working in the field of psychotraumatology, one AI scientist working in the field of deepfake technology, an ethicist, and a patient representative with lived experience of prolonged sexual abuse, trauma-focused treatment, and victim-offender mediation. Four experts in the field of restorative justice and young adults' mental health served as advisors.

2.1.3. Step 3, bringing together a team and establishing a decision-making process

The project team made all final decisions regarding protocol development and study design. The project team and advisors met regularly to provide input and feedback. In addition, advisors were consulted separately on relevant issues. Their advice contributed significantly to decision-making.

2.1.4. Step 4, reviewing published research evidence

First, all team members and advisors were asked to recommend research and theoretical papers from their specific discipline, yielding literature on sexual abuse, restorative and epistemic justice, and ethical aspects of deepfake interventions. Second, several semi-structured searches were conducted in PsycINFO for various topics including deepfake or virtual reality therapy for PTSD, sexual abuse and its psychological outcomes, and perpetrator typology. Third, additional papers were recommended by experts interviewed during step 7.

2.1.5. Step 5, drawing on existing theories

Development was based on four theories. First, in *cognitive behavioural therapy*, trauma-related fear structures are activated within a safe therapeutic

environment, through controlled exposure to trauma cues. This process enables corrective experiential learning by disconfirming maladaptive cognitions and feared expectations (Cooper et al., 2017). Second, the *needs-based model of reconciliation* states that moral transgressions threaten victims' sense of agency, and consequently, lead to a heightened need for empowerment. Through restorative exchanges between victim and perpetrator this need may be met (Shnabel & Adler, 2008; Shnabel & Nadler, 2015). Third, *empty chair techniques* commonly follow a structure of expressing unresolved feelings of hurt or guilt, followed by a response of validation, acceptance or explanation. Such an exchange then leads to a shift in the representation of oneself and/or the other person (Elliott et al., 2004; Litz et al., 2016). Fourth, *perpetrator typology* states that perpetrators of sexual offenses may be classified according to type of offense, motive, and ways of coercing the victim, leading to a distinction between preferential and situational perpetrators, and between manipulative, introverted ('grooming') and sadistic forms of coercion (Van de Weijer et al., 2023).

2.1.6. Step 6, articulating a programme theory

Deepfake therapy can be framed within the cognitive–behavioural therapy model as a novel form of exposure intervention informed by emotional processing theory. The patient is exposed to the deepfake video of the perpetrator within a controlled, non-threatening environment. By counteracting avoidance behaviours, the patient's fear structure is activated, fostering emotional engagement. This process allows for the integration of new, adaptive cognitive information. As a result, the patient's fear response will diminish. In addition, the deepfake technology aims at optimizing the vividness of the experience by bringing the perpetrator to life through movement and speech, engaging multiple sensory modalities. Despite knowing that the 'perpetrator' is artificially generated, the combination of heightened realism and fear network activation is theorized to elicit a similar psychological response as an actual conversation with the perpetrator, but within a safe setting. The patient is encouraged to express sexual abuse-related emotions and cognitions, and may direct questions to the deepfake perpetrator. In turn, the deepfaked perpetrator's responses are designed to validate the patient's experiences and address individual maladaptive cognitions identified prior to the session. This approach allows for the targeted correction of dysfunctional beliefs. Importantly, in contrast with the sexual abuse itself, the patient is now in control of the interaction with the perpetrator. The patient decides when to start and to end the conversation, determines its content, and makes eye contact with the simulated perpetrator. Consequently, the patient may regain a sense of control.

2.1.7. Step 7, undertaking primary data collection

The primary data collection consisted of interviews with different stakeholders and experts. A topic list was designed consisting of general questions on the content and expected feasibility, acceptability and efficacy of the intervention. Where applicable, questions were added pertaining to the interviewee's specific expertise. Interviews were conducted online by the first and fourth author and took 60 min on average. Interviews were recorded and key findings were documented during and after the interview.

Interviews were held with in total 15 people (six men and nine women): four patient representatives with lived experience of sexual abuse-related PTSD, PTSD treatment, and restorative justice or deepfake interventions; one offender with lived experience of sexual perpetration, psychological treatment for sexual perpetration, and restorative justice interventions; two experts in the field of forensic psychology; three therapists with clinical experience of deepfake interventions; two restorative justice experts; one ethicist; one linguist; and one expert in the field of deepfake technology. One person declined to participate due to time constraints.

Findings were then organized using the following themes: (1) indication criteria, (2) preparation and feasibility (preparatory session, ethical and legal aspects, technology), and (3) deepfake conversation (practical requirements, role of patient, role of coaching therapist, role of deepfake therapist, potential questions, and don'ts).

2.1.8. Step 8, understanding context

The intervention is intended to be used in a clinical context, to aid in the treatment of adult patients with sexual abuse-related PTSD. As trauma-focused treatment is treatment-of-choice for PTSD in adults (e.g. National Institute for Health and Care Excellence [NICE], 2018), the intervention is intended as an add-on intervention after the minimum recommended number of trauma-focused sessions has been followed (i.e. eight).

2.1.9. Step 9, paying attention to future implementation

Attention to future implementation aimed at ensuring that all technological, legal and clinical requirements for implementation of the intervention in a clinical and research setting could be met. Regarding legal feasibility, three independent lawyers and one privacy officer were consulted on the legal admissibility of the intervention under Dutch law.

2.1.10. Step 10, designing and refining the intervention

A draft protocol (version 1) was designed which was then evaluated by the project team members, who

rated feasibility and acceptability of the different parts on a scale from 1 to 10 and suggested improvements needed to reach a score of 10. Based on these evaluations, the protocol was modified (version 2) and then presented to project team members, advisors, and additional experts, using the same rating scale. In addition, two trial runs of the intervention were performed using a simple version of so-called 'verbal protocol': a research method that involves participants thinking aloud while they perform a task (Fonteyn et al., 1993, as recommended by O'Cathain et al., 2019). As our protocol involved three users (participant, coaching therapist, and deepfake therapist) simultaneously thinking aloud turned out to be impractical. Instead, feedback was collected after the preparatory session, after the deepfake session, and after one week. Subsequently, a final version of the protocol was developed using several iterations.

2.2. Outcome

The development process resulted in an intervention protocol for a conversation between a patient and a 'deepfaked' perpetrator, to be used in the psychological treatment of adults with sexual abuse-related PTSD. The intervention aims to reduce the severity of residual PTSD symptoms (especially NACM), as well as to increase feelings of empowerment in relation to the perpetrator. The intervention protocol consists of three parts, four attachments, and two forms (see Table 1). The intervention requires two therapists: a coaching therapist and a deepfake therapist.

A preparatory form is sent to the patient one week before the session. It enquires about symptoms, goals, topics, questions, wishes, and needs, helping the patient to prepare for the session. The patient brings the form to the preparatory session. The preparatory session follows a scripted protocol that may be read aloud, adapting sentences where needed for a natural conversation. The session addresses the patient's symptoms; explains the intervention; clarifies the patient's goals; gathers relevant information about the perpetrator; enquires about what the patient wishes to say, ask, and hear; and asks what the patient needs during and after the deepfake session in terms of guidance and support. Answers are noted on a therapist note form. The preparatory session takes approximately 60 min.

The deepfake session instructions for the coaching therapist cover practical matters and the therapist's tasks. These range from helping the patient get into the session, supporting the patient, and helping the patient express emotions and hold the perpetrator accountable, to conferring with the patient about session content, helping the patient to confront and overcome avoidance behaviour, and facilitating engagement with questions. The deepfake session takes approximately 60-90 min and is followed by a debriefing of approximately 30 min.

The deepfake session instructions for the deepfake therapist also explain practical matters before discussing the therapist's role. This role is explicitly stated as facilitating, benevolent, validating, and excusing the patient (i.e. taking the blame), through attentive listening and responses that are both empathetic and tailored to the patient's specific needs. To ensure authenticity, during the preparatory session the deepfake therapist is given information about the perpetrator, including gender, age, relationship to the victim, general information about the sexual abuse, and ways of forcing or grooming the victim. In addition, the deepfake therapist may refer to two attachments that provide information about different types of perpetrators as well as sample sentences that match these different types (based on McLeod & Dodd, 2022; Van de Weijer et al., 2023; U.S. Centers for Disease Control and Prevention [CDC], 2024).

For both therapists, the last attachment reminds them of the different stages in the deepfake session, allowing them to adapt their coaching and responses accordingly. Table 2 shows an example of an actual deepfake dialogue published with permission of the patient and therapists.

2.3. Feasibility

2.3.1. Technological feasibility

The intervention requires a secure online communication platform. We use a programme called 'DeepTherapy.ai', which enables users to create lifelike video animations based on a single photo using generative AI. During the session, the therapist controls the video animation with their voice, which is sychronised with the mouth movements of the animation to enable a realistic and interactive conversation. The platform operates through a secure videoconference. No special software or hardware is required, except a web browser.

2.3.2. Legal feasibility

The uploading of a perpetrator photo in the deepfake programme involves the processing of personal data for which a legal foundation is needed. Asking the perpetrator for permission might provide such a foundation but requires the patient to contact the perpetrator and disclose being in treatment, which would constitute an invasion of the patient's privacy as well as potential exposure to risk. Alternatively, legal justification can be found in the patient's

Table 1. Deepfake intervention protocol: structure and examples.

Part	Structure	Example
Part I	Symptoms	*"What are your main symptoms?"*
Preparatory session	Explanation of intervention	*"I will coach you."*
	Clarification of goals	*"What do you hope to achieve?"*
	Information about perpetrator	*"Which information is important for us to know?"*
	Inventory of topics, questions, and wishes	*"Do you have any questions?' 'What would be helpful?"*
	Inventory of needs during the session	*"How can I best coach you?"*
Part II	Requirements	What do you need?
Coaching therapist	Preparation	Check the sound before the session.
	Session structure	Take your time, up to 90 min.
	Use of deepfake programme	Open the programme by clicking the link.
	Starting the session	*"First take a look. How does that feel?"*
	Making contact	*"Tell [the perpetrator]."*
	Supporting	*"You're doing great."*
	Expressing of emotions	*"How does that feel? Tell [the perpetrator]."*
	Holding accountable	*"Feel free to say: you had no right to do that."*
	Conferring	*"Anything else you'd like to say?"*
	Focussing	*"You wanted to express your anger but you seem to be holding back."*
	Dealing with questions	*"What's it like for you that you will never know the answer?'*
	Ending the session	The participant can always decide to stop the deepfake conversation.
	Discussion of session	*"What was this session like for you?"*
Part III	Requirements	What do you need?
	Use of deepfake programme	Find the instructions in attachment I.
Deepfake therapist	Preparation	Read attachment II and mark any applicable example sentences.
	Session structure	Make sure to be ready when the patient starts the deepfake session.
	Role	Your role is to help the patient express themselves.
	Greeting	It may be helpful to greet the patient at the start of the session.
	Listening	*"I hear you."*
	Responding	*"I understand."*
	Answering	*"Whatever the answer, I never should have done this."*
	Taking the blame	*"I chose to do this."*
	Ending the session	The participant can always decide to stop the deepfake conversation.
	Discussion of session	You may ask additional questions or give additional responses.
Attachment I	General instruction	This attachment provides a step-by-step instruction.
Deepfake programme	Gaining access	Log in using the following password.
	Preparation of deepfake and planning	Click on 'create animation'.
	Deepfake session	Clik on 'open deeptherapy session'.
Attachment II	Abuse	*"This never should have happened."*
Conversation topics	Excusing the patient	*"It wasn't your fault."*
	Taking the blame	*"I was responsible as parent/adult."*
	Consequences	*"If anyone should feel bad about this, it should be me."*
	Wishes and future	*"I will think about this."*
Attachment III	Preferential	*"I have always been attracted to children."*
Sexual violence motives	Ange, aggression and power	*"I wanted to take it out on somebody."*
	Broader antisocial behaviour	*"I didn't care what you wanted."*
	Situational factors	*"You were just available."*
	Unintentional perpetration	*"I thought it was alright."*
	Financial or other gain	*"I wanted to make money."*
	Forced or accompanied by male perpetrator	*"I didn't dare say no."*
	Relational female perpetrator	*"I wanted to show you that I cared about you."*
Attachment IV	Starting	Opens conversation and express emotions
Stages of conversation	Expressing emotions	Deepens emotional expression and distinguishes between emotions
	Expressing unmet needs	Indicates which needs have not been met by the perpetrator
	Shift in representation of perpetrator	Changes understanding of perpetrator, e.g. as a less powerful person.
	Self-affirmation	Affirms self and ends conversation, e.g. by holding perpetrator responsible.
Form I. Patient preparatory form	Follows structure of preparatory session	*"Is there anything that worries you?"*
Form II. Therapist note form	Follows structure of preparatory session	

legitimate interest (i.e. allowing them to further process their trauma). The legitimate interest of the patient may then, under Dutch law, rightfully prevail over the perpetrator's right to privacy.

2.3.3. Clinical feasibility

The intervention requires two therapists: one who coaches the patient and one who plays the role of the deepfake perpetrator. Therapists need to be knowledgeable about sexual abuse dynamics and trauma-focused cognitive behavioural therapy. In addition, they need to be trained in applying the protocol.

2.4. Acceptability

2.4.1. Clinical acceptability

The protocol is carefully worded to increase patient acceptability and acknowledges inclusiveness and diversity of different patients. To support patient empowerment, the term 'victim' is not used. Relatively short sentences and consistent terminology are used to increase clarity. The protocol repeatedly stresses that patients can take their time and are in full control. A patient representative was involved in writing the intervention protocol, the study

Table 2. Example patient-therapist dialogue.

Speaker	Communication
P	*(Looking straight at the deepfake image)*: Well the big question is of course, why?
DT	Yes. I thought you would ask that. And when I think back on it now I think that that really had to do with a feeling of power.
P	But the things you have done, that, that you also just forced to have things done, do you realize that?
DT	Well now that I'm older I have started to realize that a bit more. Looking back, I think it's bizarre that I did that. Uhm.
P	But you just, you humiliated me.
DT	*(Looking pensive)*: Yes, it's terrible to hear that.
P	But you did that so often.
DT	Yes, I was, I am that disturbed.
P	*(Interrupts)*: Time and time again.
DT	Yes, there is no way of justifying that.
	(P raises hands and then bows head with hands in front of face)
CT	Would you like to close that for a second?
P	*(Sniffs, closes laptop and removes headphone, sighs, puts hands in front of face)*: Sorry.
CT	*(Whispers)*: Really well done, well done [name].

Note. P = patient, DT = deepfake therapist, CT = coaching therapist.

protocol, and this paper, with a special focus on acceptability (see Box 1).

Box 1. Reflection from a patient representative.

As a child, I was sexually abused by my father, which resulted in complex PTSD. I received extensive trauma treatment and engaged in letter-based contact with my father through restorative justice mediation. As a patient representative, I raised the following points during the intervention development.

(1) During the deepfake session, the deepfake perpetrator responds to questions that the real perpetrator might have answered differently or not at all. This raised ethical questions for me: What message does it send to the patient when someone else speaks on behalf of the perpetrator? I found it problematic if patients are left with the impression that, because the perpetrator committed sexual violence, their voice no longer matters. In my own recovery process, I found that learning to accept reality – even when painful or unsatisfying – was valuable. Processing these emotions helped me move forward. In my view, responses from a deepfake perpetrator that do not align with the attitude or behaviour of the real perpetrator may be unhelpful. Also, although most patients will cognitively understand that the session involves a deepfake rather than a real perpetrator, I still have concerns. If the therapeutic effect partly relies on the brain responding as if the confrontation is real, then the intervention involves a blending of reality and deepfake on a subconscious level. I believe this should not be overlooked too easily.

(2) My impression is that incongruent responses from the deepfake perpetrator may lead to annoyance or resistance in the patient, which could undermine the therapeutic effect. These concerns were addressed by emphasizing speaking out rather than asking questions by the patient. The sample questions were also critically reviewed. Finally, the deepfake conversation is not intended as a first-line intervention.

(3) In my view, the deepfake session may evoke intense and difficult emotions and thoughts, and confusion during and after the session, despite its expected efficacy. I emphasized the importance of discussing what the patient needs to feel safe and confident when returning home, and to cope with whatever the session may trigger. This was subsequently addressed in the protocol.

(4) Together with the team, we reflected carefully on the language used – aiming to do justice to what patients have gone through and what they experience during trauma treatment, while also presenting a realistic and empowered image of these individuals to a broader audience.

2.4.2. Ethical acceptability

A first issue concerns the use of 'perpetrator' versus 'offender'. While 'offender' is commonly used in restorative justice interventions, not all those accused of committing sexual abuse may have actually been convicted. Consequently, in the protocol, 'perpetrator' is used to refer to the person identified by the patient as having committed the sexual abuse, regardless of whether they have been convicted.

A second issue concerns the patient's understanding that the perpetrator is not actually present in the conversation. This is related to issues of culpability (i.e. the therapist's admission of guilt is not an admission of guilt by the perpetrator) and safety (i.e. the therapist's benevolence does not imply an actual benevolence by the perpetrator). Consequently, during the preparatory session, the protocol clearly stresses the absence of the perpetrator. In the study (see below), patients who are unable to understand that the perpetrator is not actually present are excluded from participation.

A third issue involves the words spoken by the deepfake therapist and the extent to which they should reflect the attitude of the actual perpetrator. While in real life perpetrators might be malevolent, this is considered unethical as it might amount to verbal abuse or victim-blaming. However, unjustifiedly suggesting that the perpetrator is benevolent might also be considered unethical. This issue is solved by carefully asking the patient for information about the perpetrator, instructing the deepfake therapist to explain any malevolent tendencies but not actively expressing them, and stressing that the intervention revolves primarily around expression by the patient rather than responses by the deepfake therapist.

A fourth issue concerns the potential psychological harm of the intervention. The intervention may be distressing for both the patient and the deepfake therapist. In response to this concern, care is taken to fully explain the intervention to the patient during the preparatory session, and to offer peer support and supervision to the therapists.

Having described the intervention development, we now turn to the study protocol designed to evaluate its safety and efficacy.

3. Study protocol

We developed a study protocol following the SPIRIT guidelines (see consort-spirit.org).

3.1. Study design

The study concerns a non-concurrent multiple baseline Single-Case Experimental Design (SCED) involving 10 participants. Non-concurrency allows for inclusion of participants at different times. Participants are randomly assigned to a baseline length of 10, 15 or 20 days to ensure internal validity and to demonstrate that symptom change is related to the intervention rather than to passage of time, maturation or other external factors. Daily measures of three main symptoms are conducted during the baseline phase as well as until 10 days after the intervention, ensuring substantial statistical power. In addition, main assessments are conducted at the start (T1) and end (T2) of the baseline phase, and one week (T3) and three weeks (T4) after the intervention. This type of design is fitting to bridge the gap between fully experimental research in a laboratory setting and a randomized controlled clinical trial.

3.2. Setting

The study takes place at ARQ Centrum'45, a Dutch national center for diagnostics and treatment of complexly traumatized patients, in collaboration with researchers from Psychotrauma Expertise Centrum PSYTREC and Landelijk Psychotraumacentrum voor Kinderen en Jongeren (National Psychotrauma Center for Children and Adolescents) UMC Utrecht.

3.3. Participants

3.3.1. Inclusion criteria

Patients are included who (1) are at least 18 years old, (2) are being treated for PTSD related to exposure to sexual abuse, (3) have had at least eight sessions of trauma-focused treatment, (4) score ≥3 on 'cued psychological distress', 'exaggerated negative beliefs or expectations', 'distorted cognitions leading to blame' and/or 'persistent negative emotional state' as assessed with the Clinician-Administered PTSD Scale for DSM-5 Revised (CAPS-5-R; Weathers et al., 2022) and (5) have a frontal photo of the perpetrator.

3.3.2. Exclusion criteria

Patients are excluded who (1) have a visual or auditory impairment that prevents them from seeing the deepfake image and/or hearing the deepfake therapist, and (2) are unable to fully comprehend, for whatever reason, that the image of the perpetrator is deepfaked, and that the conversation is with a therapist and not with the actual perpetrator.

3.3.4. Sample size

There is no set rule for the number of participants needed for a multiple baseline design. Some recommend a minimum of three participants (Kratochwill et al., 2012), others four to eight (Ferron et al., 2009). Based on this range, we decided on 10 participants.

3.5. Intervention

The intervention consists of an at-home preparation by the patient, a 60-min preparatory session, a 60-90-min deepfake session, and a 30-min debriefing session, to be delivered in one half-day. For a full description, see 'Development'.

3.5.1. Strategies to improve adherence

Therapists follow an online training; read literature on sexual abuse, trauma-focused cognitive behavioural therapy, and perpetrator typologies; watch a webinar on sexual abuse dynamics; and practice the intervention using roleplay. During the study, they receive supervision by two clinical psychologists to support clinical integrity.

3.5.2. Concomitant care

Participants are asked to refrain from trauma-focused treatment (see NICE, 2018) during the study. Supportive treatment is allowed. In addition, participants are requested to not change the type or dosage of any psychotropic medication.

3.5.3. Criteria to guide discontinuation

Participants can leave the study at any time for any reason should they so wish, without any consequences. The investigator can withdraw a participant from the study for urgent medical or psychological reasons.

3.6. Outcomes

The primary outcomes are those on an individual level (daily measurements). These concern the main posttraumatic cognition, main symptom of guilt and shame, and main level of empowerment, rated on a ten-point scale ranging from 1 (*no problem at all*) to 10 (*it bothers me extremely*). Secondary outcomes are those on a group level (main assessments). These concern posttraumatic cognitions, guilt and shame, level of empowerment, PTSD symptom severity, and forgiveness.

Dutch versions of the following, psychometrically sound measures are administered (in alphabetical order; see also Table 3).

3.6.1. Credibility & Expectancy Questionnaire (CEQ)

The CEQ (Devilly & Borkovec, 2000) is a 6-item self-report instrument for measuring treatment expectancy and rationale credibility. The scale is divided

Table 3. Participant timeline.

Order	Timepoint	Content	Estimated duration in minutes
1	Information	Receiving and reading participant information form	N.a.
2	Telephone call	Discussion of study, questions, and interest	30
3	Inclusion	Demographic and clinical questions, CAPS-5-R, CEQ	60
4	10, 15 or 20 days pre-intervention (T1)	PTCI, PCL-5, GSQ-8,TRIM-20, SCS	50
		Start of daily assessments	5 per day
5	7 days pre-intervention	Participant receives preparatory form	60
6	1 day pre-intervention (T2)	PTCI, PCL-5, GQ-8, TRIM-20, SCS	50
7	Preparatory session	Preparation for deepfake session with therapists	60
8	Deepfake session	Deepfake perpetrator conversation	60–90
9	Debriefing	Debriefing of deepfake session	30
10	1 week post-intervention (T3)	PTCI, PCL-5, GSQ-8, TRIM-20, SCS, CSQ-8, qualitative interview	55
11	10 days post-intervention	End of daily assessments	
12	3 weeks post-intervention (T4)	PTCI, PCL-5, GSQ-8, TRIM-20, SCS	80
13	Evaluation	Looking back at study, travel expenses, gift coupon	20

into two sections: thinking and feeling. Items are rated either on a 9-point scale or as a percentage, with a higher rating or percentage reflecting a higher treatment credibility and expectancy. Timepoint: at inclusion.

3.6.2. Client Satisfaction Questionnaire-8 (CSQ-8)

The CSQ-8 (Larsen et al., 1979) is an 8-item self-report instrument that measures global satisfaction with an intervention. Items are rated on a 4-point scale, with a higher score indicating a higher level of overall satisfaction. Timepoint: T3.

3.6.3. Clinician-administered PTSD scale for DSM-5 revised (CAPS-5-R)

The CAPS-5-R (Weathers et al., 2022) is a 20-item clinical interview that assesses the presence and severity of PTSD symptoms according to DSM-5. Items are rated on an 11-point scale ranging from 0 (*absent*) to 10 (*extreme*). A rating of 3 is considered the threshold for a clinically significant problem. In this study, only the following items are administered: exposure to sexual violence (A); 'cued psychological distress' (B4), 'exaggerated negative beliefs or expectations' (D2); 'distorted cognitions leading to blame' (D3); and 'persistent negative emotional state' (D4). Timepoint: at inclusion.

3.6.4. Guilt and Shame Questionnaire (GSQ-8)

The GSQ-8 (Hoppen et al., 2022) is an 8-item self-report instrument that assesses shame and guilt over the past four weeks. Items are rated on a five-point scale ranging from 0 (*never present*) to 4 (*daily present*), with higher scores indicating a higher frequency. For this study, the scale was adapted to enquire about the past week, and consequently rating labels were adapted to reflect severity rather than frequency. In addition, a remark concerning trauma-relatedness was added. Timepoints: T1, T2, T3 and T4.

3.6.5. Posttraumatic cognitions inventory (PTCI)

The PTCI (Foa et al., 1999) is a 36-item self-report questionnaire that assesses negative cognitions about self and others, and self-blame. Items are rated on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*), with higher scores indicating greater endorsement of posttraumatic cognitions. Timepoints: T1, T2, T3 and T4.

3.6.6. PTSD checklist for DSM-5 (PCL-5)

The PCL-5 (Weathers et al., 2013) is a 20-item self-report questionnaire that assesses the presence and severity of PTSD symptoms according to DSM-5 over the past week. Items are rated on a 5-point Likert scale ranging from 1 (*not at all*) to 5 (*extremely*), with a higher score indicating a higher PTSD symptom severity. Timepoints: T1, T2, T3 and T4.

3.6.7. Social comparison scale (SCS)

The SCS (Allan & Gilbert, 1995) is an 11-item self-report rating scale that measures self-perceptions of social rank and relative social standing using bipolar constructs. Participants are required to make a global comparison of themselves in relation to other people along a 10-point scale, with low scores pointing to feelings of inferiority and generally low rank self-perceptions. For this study we adjusted the original instrument to specifically examine empowerment in terms of comparison with the perpetrator. In addition, we added 7 constructs: guilty vs innocent; paralyzed or frozen vs not paralyzed or frozen; ashamed vs not ashamed; anxious vs relaxed; I continuously think about the perpetrator vs I don't think about the perpetrator anymore; It is my fault vs The perpetrator is to blame; and I can't let it go vs I can let it go. Timepoints: T1, T2, T3 and T4.

3.6.8. Transgression related interpersonal motivation scale (TRIM)

The TRIM (McCullough et al., 1998) is a 20-item self-report instrument that measures forgiveness in terms of interpersonal motivation for revenge, avoidance,

and benevolence in response to direct acts of aggression. Items are answered on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*), with higher scores indicating a greater motivation. Timepoints: T1, T2, T3 and T4.

3.6.8. Qualitative interview

At T3 a semi-structured interview is administered to evaluate participants' experiences with the intervention. Topics include reasons to participate in the study, aims of participation, evaluation of what the participant said and heard at the confrontation, effect of the intervention, any adverse effects, and suggestions for improvement of the intervention.

3.6.9. Harms

Harms are defined as any undesirable experience occurring to a participant during the study, whether or not considered related to the deepfake intervention. These specifically concern the worsening of posttraumatic stress disorder, depressive disorder, suicidal ideations or behaviour, and automutilation. Harms are checked at T3 through an interview question as well as at study conclusion.

3.7. Procedure

3.7.1. Recruitment

Recruitment takes place at PSYTREC, ARQ Centrum'45, and Landelijk Psychotraumacentrum voor Kinderen en Jongeren UMC Utrecht. In addition, patients may be alerted to the study by their therapists who are informed of the study through LinkedIn. Patients receive a participant information form through their therapist and may consequently contact the principal investigator. A phone call is planned to answer any questions and provide additional information. The patient then has one week to consider participating. If the patient agrees to participate, an online appointment is made to assess inclusion and exclusion criteria. At the inclusion session the participant provides written informed consent.

3.7.2. Assignment of baseline length

After inclusion, participants are randomly assigned to one of three baseline lengths. Before the first inclusion, the allocation sequence is generated through an online tool by an independent researcher, assigning 10 days to four participants, 15 days to three participants, and 20 days to three participants in random order. Each allocation is written on a piece of paper and inserted in a numbered envelope which is opened in order of inclusion, ensuring allocation concealment. After randomization, neither participants nor researchers are blind to baseline length.

3.7.3. Participant timeline

For a participant timeline, see Table 3. Depending on baseline length, total study participation takes 32–42 days.

3.8. Data collection and management

Daily assessments are conducted at 8 PM, using a secure app (m-Path). Main assessments are conducted through a secure assessment program (Embloom). A research assistant monitors the assessments and contacts the participant if they fail to complete the assessment on the set day. The following personal data are collected: age, sex, education level, employment status, marital status, exposure to sexual violence, medication, and psychological treatment. Data files are coded, generated by a data manager, and stored for 15 years after completion of the study. All data belong to ARQ Centrum'45 but may be shared upon reasonable request.

3.8.1. Monitoring

Quality monitoring is performed before, during and after data collection to determine whether the research is conducted properly and whether any systematic errors during data collection are made. A clinical research associate establishes a monitoring plan and verifies (1) essential research data in source documents, (2) the reporting of serious adverse events, (3) the documentation of informed consents, and (4) study conduct in accordance with the research protocol.

3.9. Statistical methods

Outcomes are analysed on an individual level (primary) as well as a group level (secondary). Bayesian statistics are used to compare differences in symptom severity between different phases (baseline phase, posttreatment phase) and timepoints (T1, T2, T3, T4). Bayesian evaluation of SCEDs uses the following steps: (1) formulation of hypotheses regarding the development of outcome measures over time; (2) specification of a statistical model; (3) formalizing of the hypotheses; and (4) evaluation of the hypotheses using the Bayes Factor and/or Posterior Model Probabilities. The hypotheses are evaluated simultaneously for all outcome variables.

3.9.1. Individual outcomes

The daily measures of the three main symptoms for each individual patient are analysed using the following hypotheses: there is (1) no difference, (2) an improvement, (3) a deterioration, or (4) something else, during (i) baseline and (ii) posttreatment. In addition, symptom change during the baseline and

posttreatment phases are compared using hypotheses regarding (1) no difference, (2) superiority of symptom change in one phase over the other, or (3) something else. A multivariate regression analysis with three dependent variables (main symptoms) and one predictor (time) is conducted. The hypotheses are evaluated using Posterior Model Probabilities.

3.9.2. Group outcomes

Bayesian statistics are used to compare differences in symptom severity between different timepoints: T1 and T2 (baseline), T2 and T3 (pre-treatment to post-treatment), and T3 and T4 (post-treatment to follow-up). The following hypotheses are specified for each comparison: there is (1) no difference, (2) an improvement, (3) a deterioration, or (4) something else, in secondary outcome measures between (i) T1 and T2, (ii) T2 and T3, and (iii) T3 and T4. In addition, symptom change during the three phases (baseline, pre- to post-treatment, and post-treatment to follow-up) is compared using hypotheses regarding (1) no difference, (2) superiority of symptom change in one phase over the other, or (3) something else. A multivariate regression analysis with five dependent variables (posttraumatic cognitions, posttraumatic stress, guilt and shame, forgiveness, and empowerment) and one predictor (time) is conducted. The hypotheses are evaluated using Posterior Model Probabilities.

3.9.3. Other study parameters

Study parameters that focus on acceptability (treatment credibility and expectancy, treatment satisfaction) are analysed in SPSS. Outcomes of the semi-structured interviews are analysed using MAXQDA.

3.10. Ethics

The study design was approved by the Medical Ethics Review Board of the Amsterdam University Medical Center and preregistered at OSF. Protocol amendments are submitted for major changes. All participants provide informed consent. Personal information is kept during study participation on the ARQ Centrum'45 servers. After completing the follow-up assessment, participants evaluate their participation with the principal investigator and if necessary, appropriate clinical action is taken.

4. Discussion

We developed an intervention protocol for a perpetrator conversation using deepfake technology, based on interviews, scientific and clinical literature, and trial runs, and primarily aimed at reducing residual symptoms in patients with sexual abuse-related PTSD. The efficacy of this intervention is evaluated in a multiple baseline study. Given the innovative nature of the intervention, feasibility, acceptability and potential efficacy merit special reflection.

4.1. Feasibility

In a reflection on the legal aspects of deepfake therapy, Hoek and colleagues (2025) depict the use of the perpetrator's image as potential 'healing through stealing' and stress that therapeutic deepfake images should be created in compliance with applicable laws. These may differ per country, meaning that the legal admissibility of deepfake therapy across countries may not be assumed. Hoek et al. also stress that while using a deepfake image without consent by the perpetrator might be in the legitimate interest of the patient, measures should be taken to protect the depicted person's privacy. To mitigate these concerns, in our study deepfake images are seen only by the therapists and deleted immediately after the session. In a reflection on the ethics of deepfake technology, Meskys et al. (2020) stress the importance of a strong position from regulators. This means that legal developments would need to be closely followed and that the legal position of therapeutic deepfakes remains a matter of interest.

As sexual abuse often involves family members or acquaintances, relationships following the abuse may be complicated and may involve negative as well as positive emotions (e.g. Herland, 2023). Consequently, the exchange between the patient and the deepfake therapist requires a lot of clinical skill, as it is adaptive to the patients' needs and responses rather than strictly scripted. This means that the intervention might be most feasible in specialized settings.

4.2. Acceptability

The deepfake protocol was developed in close collaboration with patient representatives. As such, its development may be perceived as an example of a 'person-based' approach to assessment and enhancement of the acceptability of a new intervention (Yardley et al., 2015). It also followed principles advocated by a perception of care ethics as an interdisciplinary field, addressing the question of what would be morally good from a perspective of care, using both interdisciplinary discussions and empirical research, and involving lived experience (Leget et al., 2017).

Guerouaou and colleagues (2021) have raised concerns about the truthfulness, fairness, autonomy, and transparency of deepfakes. Our protocol addresses these concerns by ensuring patients are fully informed of the deepfake nature of the intervention, excluding those unable to understand that the image is deepfaked, aligning with the patient's goals, emphasizing patient autonomy, and prioritizing therapeutic

benefit. Hoek and colleagues (2025) additionally point to the risk of overattachment, speculating that the patient may associate the deepfake therapist with the deepfaked person or become 'addicted' to communicating with the deepfaked image. Such risks appear limited in our protocol: the regular therapist typically takes the coaching role while the deepfake therapist is involved in the deepfake intervention only, the intervention is limited to one session, and the deepfake platform cannot be accessed independently by the patient.

4.3. Potential efficacy

Deepfake technology may offer valuable opportunities for advancing health care (e.g. Becker & Laycock, 2023). Compared to traditional role play or imagination techniques, deepfake technology might bring enhanced emotional impact and realism. This might be especially helpful for individuals with mental imagery problems (King et al., 2024). In patients with PTSD, increased emotional involvement may lead to more cognitive shifts allowing for more effective trauma-processing. For patients unable to face their perpetrator, a deepfake meeting can mimic interaction in a vivid and tangible manner.

Deepfake therapeutic techniques are increasingly explored across different healthcare environments (see Kaur et al., 2024 for an overview). However, an important question is whether such interventions might also be ineffective or even harmful. While no studies have been conducted that address this question, studies have shown that some restorative justice practices may indeed be unhelpful. Choi and Severson (2009) showed that some apologies offered by offenders may lead to distrust, disappointment or high negative emotion in victims. In another study, Choi and colleagues (2013) showed that victims occasionally felt intimidated by offenders. While these findings may not necessarily generalize to deepfake perpetrator conversations, they stress the importance of responding in a way that is helpful to the patient. The deepfake intervention allows for real-time adjustment or repetition of responses during the session, which is not possible in real-life perpetrator confrontations. Both response helpfulness and possible harms need to be addressed in future studies.

4.4. Strengths and limitations

The intervention was carefully developed by a multidisciplinary team following scientific guidelines (O'Cathain et al., 2019) and is strongly rooted in evidence-based treatment approaches. A literature search and interviews led to the formulation of an intervention theory and the subsequent development of the intervention. The development resulted in an intervention protocol that is deemed technologically, legally and clinically feasible, and clinically and ethically acceptable. In addition, the development process yielded ethical and legal insights that might be useful to those who wish to develop similar protocols.

The development process also has several limitations. One concerns the limited diversity of those involved in the development, all of whom were White and identified as cisgender. Efforts to involve a more diverse group in the trial runs were unsuccessful, resulting in the exclusion of perspectives from patients with a non-Western background such as immigrants or refugees, persons of colour, or persons with a gender-diverse identity, who may be especially vulnerable to sexual abuse (e.g. Abern et al., 2022). In addition, the intervention protocol was limited to one session based on restorative justice practices. However, the trial runs suggested that one session might not suffice for all patients. Future research will have to determine the number of sessions required for specific patients.

5. Conclusion

In conclusion, we developed an intervention protocol for a deepfake perpetrator-conversation for adult patients with residual PTSD symptoms related to sexual abuse. The protocol was deemed feasible, acceptable and potentially effective. The actual efficacy is tested in a multiple-baseline study.

Acknowledgements

The authors thank the two anonymous patients for taking part in the trial run of the intervention; professor Herbert Hoijtink for advising on the proposed Bayesian analyses; and Jan Willem van den Berg, Rens Boeijen, Nathalie de la Cousine, Greetje de Haan, Anneke van Hoek, Jorien Lenting, Irma Nibbelink, Alistair Niemeijer, Marina van der Roest, Anja Rosner, Nicky Slobbe, Wyke Stommel, Marieke Tuinenburg, Arnold Twigt, Linda Verhaak and Ilonka Wolswinkel for sharing their expertise through interviews and comments on different versions of the intervention protocol.

Author contributions

CRediT: **Jackie June ter Heide:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing; **Iva Bicanic:** Conceptualization, Funding acquisition, Methodology, Writing – original draft, Writing – review & editing; **Brieneke Holtrop:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing; **Larissa Maassen:** Investigation, Writing – review & editing; **Inge van Nistelrooij:** Supervision, Writing – review & editing; **Maryëlle Hoogeveen:** Investigation, Methodology, Writing – review & editing; **Mathilde Overbeek:** Methodology, Supervision, Writing – review & editing; **Eline Voorendonk:** Investigation, Methodology, Writing –

review & editing; **Sven Zebel:** Supervision, Writing – review & editing; **Theo Gevers:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing; **Agnes van Minnen:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Disclosure statement

Theo Gevers is a co-founder of 3DUniversum, a spin-off company from the University of Amsterdam, based in Amsterdam, the Netherlands. The other authors have no conflicts of interest to disclose.

Funding

This work was supported by Stichting Achmea Slachtoffer en Samenleving (SASS).

ORCID

Jackie June ter Heide http://orcid.org/0000-0001-7593-4359
Iva Bicanic http://orcid.org/0000-0001-7535-1006
Inge van Nistelrooij http://orcid.org/0000-0002-9053-7279
Maryëlle Hoogeveen http://orcid.org/0009-0008-9445-9096
Mathilde Overbeek http://orcid.org/0000-0001-9090-8271
Eline Voorendonk http://orcid.org/0000-0001-7554-8343
Sven Zebel http://orcid.org/0000-0002-0905-5687
Theo Gevers http://orcid.org/0000-0002-1190-5492
Agnes van Minnen http://orcid.org/0000-0002-3099-8444

References

Abern, L., Diego, D., Krempasky, C., Cook, J., & Maguire, K. (2022). Prevalence of sexual assault in a cohort of transgender and gender diverse individuals. *Journal of General and Internal Medicine*, *38*(5), 1331–1333. https://doi.org/10.1007/s11606-022-07900-y

Allan, S., & Gilbert, P. (1995). A social comparison scale: Psychometric properties and relationship to psychopathology. *Personality and Individual Differences*, *19*(3), 293–299. https://doi.org/10.1016/0191-8869(95)00086-l

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

Basile, K. C., Smith, S. G., Breiding, M. J., Black, M. C., & Mahendra, R. R. (2014). *Sexual violence surveillance: Uniform definitions and recommended data elements, Version 2.0*. National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

Becker, C., & Laycock, R. (2023). Embracing deepfakes and AI-generated images in neuroscience research. *European Journal of Neuroscience*, *58*, 2657–2661. https://doi.org/10.1111/ejn.16052

Bolívar, D. (2013). For whom is restorative justice? *Restorative Justice*, *1*(2), 190–214. https://doi.org/10.5235/20504721.1.2.190

Breslau, N., Kesller, R. C., Chilcoat, H. D., Schultz, L. R., Davis, G. C., & Andreski, P. (1998). Trauma and posttraumatic stress disorder in the community. *Archives of General Psychiatry*, *55*, 626–632. https://doi.org/10.1001/archpsyc.55.7.626

Cagney, J., Spencer, C., Flor, L., Herbert, M., Khalil, M., O'Connell, E., Mullany, E., Bustreo, F., Chandan, J. S., Metheny, N., Knaul, F., & Gakidou, E. (2025). Prevalence of sexual violence against children and age at first exposure: A global analysis by location, age, and sex (1990–2023). *Lancet*, *405*, 1817–1836. https://doi.org/10.1016/S0140-6736(25)00311-3

Centraal Bureau voor de Statistiek. (2024). *Prevalentiemonitor huiselijk geweld en seksueel grensoverschrijdend gedrag* 2024 [Prevalence monitor domestic violence and sexually transgressive behaviour 2024]. https://www.cbs.nl/nl-nl/longread/rapportages/2024/prevalentiemonitor-huiselijk-geweld-en-seksueel-grensoverschrijdend-gedrag-2024/samenvattin.

Choi, J. J., & Severson, M. (2009). What! What kind of apology is this?": The nature of apology in victim offender mediation. *Children and Youth Services Review*, *31*(7), 813–820. https://doi.org/10.1016/j.childyouth.2009.03.003

Choi, J. J., Gilbert, M. J., & Green, D. L. (2013). Patterns of victim marginalization in victim-offender mediation: Some lessons learned. *Crime, Law and Social Change*, *59*, 113–132. https://doi.org/10.1007/s10611-012-9382-1

Cooper, A. A., Clifton, E. G., & Feeny, N. C. (2017). An empirical review of potential mediators and mechanisms of prolonged exposure therapy. *Clinical Psychology Review*, *56*, 106–121. https://doi.org/10.1016/j.cpr.2017.07.003

Cortoni, F., Babchishin, K. M., & Rat, C. (2017). The proportion of sexual offenders who are female is higher than thought: Meta-analysis. *Criminal Justice and Behavior*, *44*(2), 145–162. https://doi.org/10.1177/0093854816658923

Craig, T. K. J., Rus-Calafell, M., Ward, T., Leff, J. P., Huckvale, M., Howarth, E., Emsley, R., & Garety, P. A. (2018). AVATAR therapy for auditory verbal hallucinations in people with psychosis: A single-blind, randomised controlled trial. *The Lancet Psychiatry*, *5*, 31–40. https://doi.org/10.1016/S2215-0366(17)30427-3

Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry*, *31*(2), 73–86. https://doi.org/10.1016/s0005-7916(00)00012-4

Dworkin, E. R., Krahé, B., & Zinzow, H. (2021). The global prevalence of sexual assault: A systematic review of international research since 2010. *Psychology of Violence*, *11*(5), 497–508. https://doi.org/10.1037/vio0000374

Elliott, R., Watson, J. C., Goldman, R. N., & Greenberg, L. S. (2004). *Learning emotion-focused therapy: The process-experiential approach to change*. American Psychological Association.

Ferron, J., Bell, M., Hess, A. A., Rendina-Gobioff, M. R., & Hibbard, G. T. S. (2009). Making treatment effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods*, *41*(2), 372–384. https://doi.org/10.3758/BRM.41.2.372

Foa, E. B., Tolin, D. F., Ehlers, A., Clark, D. M., & Orsillo, S. M. (1999). The Posttraumatic Cognitions Inventory (PTCI): Development and validation. *Psychological Assessment*, *11*(3), 303–314. https://doi.org/10.1037/1040-3590.11.3.303

Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A description of Think Aloud method and protocol analysis. *Qualitative Health Research*, *3*, 430–441. https://doi.org/10.1177/104973239300300403

Guerouaou, N., Viava, G., & Aucouturier, J. J. (2021). The shallow of your smile: The ethics of expressive vocal deep-fakes. *Philosophical Transactions of the Royal Society B*, *377*, 20210083. https://doi.org/10.1098/rstb.2021.0083

Hansen, T., & Umbreit, M. (2018). State of knowledge: Four decades of victim-offender mediation research and practice: The evidence. *Conflict Resolution Quarterly*, *36*, 99–113. https://doi.org/10.1002/crq.21234

Herland, M. D. (2023). Narratives of child sexual abuse: Conceptualizing relational complexities. *Children & Society*, *37*, 1612–1626. https://doi.org/10.1111/chso.12706

Herstelbemiddeling, P. (2020). *Herstellen doe je samen: Jaarverslag 2019 [Recovery is what you do together: Year report 2019]*. Perspectief Herstelbemiddeling.

Hoek, S., Metselaar, S., Ploem, C., & Bak, M. (2025). Promising for patients or deeply disturbing? The ethical and legal aspects of deepfake therapy. *Journal of Medical Ethics*, *51*, 481–486. https://doi.org/10.1136/jme-2024-109985

Hoppen, T. H., Jehn, M., Holling, H., Mutz, J., Kip, A., & Morina, N. (2023). The efficacy and acceptability of psychological interventions for adult PTSD: A network and pairwise meta-analysis of randomized controlled trials. *Journal of Consulting and Clinical Psychology*, *91*(8), 445–461. https://doi.org/10.1037/ccp0000809

Hoppen, T. H., Schlechter, P., Arntz, A., Rameckers, S. A., Ehring, T., & Morina, N. (2022). A brief measure of guilt and shame: validation of the Guilt and Shame Questionnaire (GSQ-8). *European Journal of Psychotraumatology*, *13(2)*), https://doi.org/10.1080/20008066.2022.2146720

Kaur, A., Noori Hoshyar, A., Wang, X., & Xia, F. (2024, October). Beyond deception: Exploiting deepfake technology for ethical innovation in healthcare. *In Proceedings of the 1st International Workshop on Multimedia Computing for Health and Medicine*, 70–78.

King, R., Buxton, H., & Tyndall, I. (2024). Aphantasia and autism: An investigation of mental imagery vividness. *Consciousness and Cognition*, *125*, 103749. https://doi.org/10.1016/j.concog.2024.103749

Kip, A., Diele, J., Holling, H., & Morina, N. (2022). The relationship of traumarelated guilt with PTSD symptoms in adult trauma survivors: A meta-analysis. *Psychological Medicine*, *52*, 2201–2211. https://doi.org/10.1017/S0033291722001866

Kooistra, M. J., Hoeboer, C. M., Oprel, D. A. C., Schoorl, S. M. D., Van der Does, W., Ter Heide, F. J. J., Van Minnen, A., & De Kleine, R. A. (2023). Changes in trauma-related cognitions predict subsequent symptom improvement during prolonged exposure in patients with childhood abuse-related PTSD. *Behaviour Research and Therapy*, *163*, 104284. https://doi.org/10.1016/j.brat.2023.104284

Kraaijeveld, S. R., & Ivanova, D. (2026). Using deepfakes for psychotherapy: Ethical and philosophical issues. *Ai and Ethics*, *6*, 79. https://doi.org/10.1007/s43681-025-00918-y

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D., & Shadish, W. R. (2012). Single-Case Intervention Research Design Standards. *Remedial and Special Education*, *34*(1), 26–38. https://doi.org/10.1177/0741932512452794

Larsen, D. L., Attkisson, C., Hargreaves, W. A., & Nguyen, T. D. (1979). Assessment of client/patient satisfaction: Development of a general scale. *Evaluation And Program Planning*, *2*(3), 197–207. https://doi.org/10.1016/0149-7189(79)90094-6

Larsen, S. E., Eubanks Fleming, C. J., & Resick, P. A. (2019). Residual symptoms following empirically supported treatment for PTSD. *Psychological Trauma: Theory, Research, Practice, and Policy*, *11*(2), 207–215. https://doi.org/10.1037/tra0000384

Leget, C., Van Nistelrooij, I., & Visse, M. (2017). Beyond demarcation: Care ethics as an interdisciplinary field of inquiry. *Nursing Ethics*, *26*(1), 17–25. https://doi.org/10.1177/0969733017707008

Litz, B. T., Lebowitz, L., Gray, M. J., & Nash, W. P. (2016). *Adaptive disclosure: A new treatment for military trauma, loss, and moral injury*. Guilford.

Lloyd, A., & Borrill, J. (2020). Examining the effectiveness of restorative justice in reducing victims' post-traumatic stress. *Psychological Injury and Law*, *13*, 77–89. https://doi.org/10.1007/s12207-019-09363-9

McBride, O., Hyland, P., Murphy, J., & Elklit, A. (2020). Network analysis of posttraumatic stress experiences of adults see psychological treatment for childhood sexual abuse. *Journal of Traumatic Stress*, *33*, 10–18. https://10.1002jts.22381

McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington, E. L., Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal Of Personality And Social Psychology*, *75*(6), 1586–1603. https://doi.org/10.1037/0022-3514.75.6.1586

McLeod, D. A., & Dodd, M. (2022). Modernized female sex offender typologies: Intrapsychic, behavioral, and trauma related domains. *Cogent Social Sciences*, *8*, Article 2085360. https://doi.org/10.1080/23311886.2022.2085360

Meskys, E., Liaudanskas, A., Kalpokiene, J., & Jurcys, P. (2020). Regulating deep fakes: Legal and ethical considerations. *Journal of Intellectual Property Law & Practice*, *15*(1), 24–31. https://doi.org/10.1093/jiplp/jpz167

Nascimento, A. M., Andrade, J., & De Castro Rodrigues, A. (2023). The psychological impact of restorative justice practices on victims of crimes: A systematic review. *Trauma, Violence, & Abuse*, *24*(3), 1929–1947. https://doi.org/10.1177/15248380221082085

National Institute for Health and Care Excellence. (2018). NICE guideline: Post-traumatic stress disorder. www.nice.org.uk/guidance/ng116.

O'Cathain, A., Croot, L., Duncan, E., Rousseau, N., Sworn, K., Turner, K., Yardley, L., & Hoddinott, P. (2019). Guidance on how to develop complex interventions to improve health and healthcare. *BMJ Open*, *9*, 029954. https://doi.org/10.1136/bmjopen-2019-029954

Petkovski, M., Knipschild, R., Gevers, T., & Lenferink, L. I. M. (2026). Virtual rescripting after loss using deepfake technology in prolonged grief treatment: A study protocol for a multiple baseline design. *European Journal of Psychotraumatology*, *17*(1), 2643984. https://doi.org/10.1080/20008066.2026.2643984

Schouten, P. J. (2016). *Traumasexuality: Men abused as boys, men between victim and perpetrator*. Eburon.

Semmlinger, V., Leithner, C., Klöck, L. M., Ranftl, L., Ehring, T., & Schreckenbach, M. (2024). Prevalence and predictors of nonresponse to psychological treatment for PTSD: A meta-analysis. *Depression and Anxiety*, 899034. https://doi.org/10.1155/2024/9899034

Sherman, L. W., & Strang, H. (2010). Restorative justice as a psychological treatment: Healing victims, reintegrating offenders. In G. J. Towl, & D. A. Crighton (Eds.), *Forensic psychology* (pp. 398–415). Wiley.

Sherman, L. W., Strang, H., Barnes, G., Woods, D. J., Bennett, S., Inkpen, N., Newbury-Birch, D., Rossner,

M., Angel, C., Mearns, M., & Slothower, M. (2015). Twelve experiments in restorative justice: The Jerry Lee program of randomized trials of restorative justice conferences. *Journal of Experimental Criminology*, *11*, 501–540. https://doi.org/10.1007/s11292-015-9247-6

Shnabel, N., & Adler, A. (2008). A needs-based model of reconciliation: Satisfying the differential emotional needs of victim and perpetrator as a key to promoting reconciliation. *Journal of Personality and Social Psychology*, *94*(1), 116–132. https://doi.org/10.1037/0022-3514.04.1.116

Shnabel, N., & Nadler, A. (2015). The role of agency and morality in reconciliation processes: The perspective of the needs-based model. *Current Directions in Psychological Science*, *24*(6), 477–483. https://doi.org/10.1177/0963721415601625

U.S. Centers for Disease Control and Prevention. (2024). Sexual violence prevention: Risk and protective factors. https://www.cdc.gov/sexual-violence/risk-factors/index.html.

Van de Weijer, S., Eichelsheim, V., Blokland, A., & Hendriks, J. (2023). *Typologieën van plegers van zedendelicten: Een umbrella review. [Typologies of offenders of sexual offences: An umbrella review]*. Nederlands Studiecentrum Criminaliteit en Rechtshandhaving (NSCR).

Van Minnen, A., Ter Heide, F. J. J., Koolstra, T., De Jongh, A., Karaoglu, S., & Gevers, T. (2022). Initial development of perpetrator confrontation using deepfake technology in victims with sexual violence-related PTSD and moral injury. *Frontiers in Psychiatry*, *13*, 882957. https://doi.org/10.3389/fpsyt.2022.882957

Weathers, F. W., Byrne, C. A., Marx, B. P., Schnurr, P. P., & Keane, T. M. (2022). *The Clinician-Administered PTSD Scale for DSM-5, Revised (CAPS-5-R) – Past Month–Initial Interview Version*. Veterans Administration.

Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013). *The PTSD Checklist for DSM-5 (PCL-5) and Life Events with extended A criterion*. National Center for Posttraumatic Stress Disorder.

Yardley, L., Ainsworth, B., Arden-Close, E., & Muller, I. (2015). The person-based approach to enhancing the acceptability and feasibility of interventions. *Pilot and Feasibility Studies*, *1*, 37. https://doi.org/10.1186/s40814-015-0033-z

Zebel, S. (2012). Een *quasi-experimentele studie naar de effecten van de Nederlandse slachtoffer-dadergesprekken* [A quasi-experimental study into the effects of the Dutch victim-offender meetings]. In I. Weijers (Ed.), *Bemiddeling na strafbare feiten: De Nederlandse slachtoffer-dadergesprekken* [Mediation after punishable offences: The Dutch victim-offender meetings] (pp. 21–44). Boom.