-- --

How to e-mental health: a guideline for researchers and practitioners using digital technology in the context of mental health

-- --

Seiferth, Caroline; Vogel, Lea; Aas, Benjamin; Brandhorst, Isabel; Carlbring, Per; Conzelmann, Annette; Esfandiari, Narges; Finkbeiner, Marlene; Hollmann, Karsten; Lautenbacher, Heinrich; Meinzinger, Edith; Newbold, Alexandra; Opitz, Ansgar; Renner, Tobias J. [**and 11 more**]

How to e-mental health: a guideline for researchers and practitioners using digital technology in the context of mental health

Caroline Seiferth [1,22], Lea Vogel [2,3,22], Benjamin Aas[3,4], Isabel Brandhorst[5,6],
Per Carlbring [7], Annette Conzelmann[5,6,8], Narges Esfandiari [9],
Marlene Finkbeiner[5,6], Karsten Hollmann[5,6], Heinrich Lautenbacher[5,6,10],
Edith Meinzinger[11], Alexandra Newbold[12], Ansgar Opitz [2], Tobias J. Renner[5,6],
Lasse Bosse Sander[13], Philip S. Santangelo[14], Ramona Schoedel[15],
Björn Schuller[16], Clemens Stachl [17], sysTelios Think Tank*, Yannik Terhorst [18],
John Torous[19], Katarzyna Wac[20], Aliza Werner-Seidler[21], Sebastian Wolf[11]
& Johanna Löchner [5,6] ✉

Despite an exponentially growing number of digital or e-mental health services, methodological guidelines for research and practical implementation are scarce. Here we aim to promote the methodological quality, evidence and long-term implementation of technical innovations in the healthcare system. This expert consensus is based on an iterative Delphi adapted process and provides an overview of the current state-of-the-art guidelines and practical recommendations on the most relevant topics in e-mental health assessment and intervention. Covering three objectives, that is, development, study specifics and intervention evaluation, 11 topics were addressed and co-reviewed by 25 international experts and a think tank in the field of e-mental health. This expert consensus provides a comprehensive essence of scientific knowledge and practical recommendations for e-mental health researchers and clinicians. This way, we aim to enhance the promise of e-mental health: low-threshold access to mental health treatment worldwide.

Mental illness is on the rise, and since the coronavirus disease 2019 (COVID-19) pandemic, prevalence rates have increased substantially[1]. At the same time, healthcare systems around the world are challenged to provide adequate psychological help owing to several (individual and social) barriers and challenges. These challenges include health disparities, insufficient infrastructure, workforce shortage, long waiting lists[2], stigmatization and low perceived need[3]. The COVID-19 outbreak also resulted in an unexpected acceleration of digitalization in different fields[4], as well as in the increased efforts in prevention, treatment and rehabilitation of mental health disorders. There are numerous advantages that e-mental health may offer: low-threshold access, geographic independence, constant availability and potentially lower cost[5].

Despite the increase and numerous advantages of e-mental health solutions, several shortcomings affect the development and delivery of e-mental health interventions. First, therapists are still sceptical of prescribing digital mental healthcare because they perceive themselves as not sufficiently trained in this area and lack knowledge about which technologies are validated and affordable for various patient groups[6]. Furthermore, uncertainty exists regarding the impact that digital

A full list of affiliations appears at the end of the paper. *A list of authors and their affiliations appears at the end of the paper.
✉e-mail: Johanna.Loechner@med.uni-tuebingen.de

approaches might have on the therapeutic relationship[7]. Second, among other challenges (for example, not being familiar with new technological developments), users face difficulties in distinguishing scarcer, evidence-based interventions among the plethora of health and well-being offers[8]. Third, the field of research is also inconsistent regarding terminology, leading to miscommunication with users and stakeholders[9]. For a consensus statement on the problem of terminology in digital psychological interventions, see ref. 9. Finally, guidance and methodological advancements are necessary not only to improve user experience but also to raise quality standards in development and evaluation. In meta-analytical reviews, most e-mental health apps were shown to perform badly regarding data security and transparency[10], methodological quality and attrition rates[11]. However, the field is evolving, and app evaluation frameworks (AEFs)[12] and common glossaries[9] have been developed to overcome such limitations and increase methodological quality.

In this Consensus Statement, we aim to provide current practical guidelines for researchers and practitioners in the field of e-mental health to cover the most important topics of the development, deployment and evaluation of e-mental health assessments and interventions. The term e-mental health covers four types of digital service[13]: (1) information provision, (2) assessment for screening and monitoring, (3) intervention and (4) social support. More specifically, this includes digital solutions in a comprehensive way, including mobile and web-based apps, digitally delivered interactions with clients via, for example, video calls and chats, chatbots and devices for assessing and monitoring health (for example, wearables and smart-watches). Since the field is so dynamic, and constantly renews itself, we refrain from focusing on specific devices, data collection, analysis or interaction methods, but aim to provide overarching and enduring recommendations.

Method

This is an expert consensus of the work of international researchers in the field of e-mental health aiming to promote the methodological quality, evidence and longer-term implementation of technical innovations in the healthcare system. For this purpose, the original author group from Germany investigated and contacted leading experts worldwide in the field of e-mental health based on Google Scholar profiles, groundbreaking publications and achievements, and personal recommendations. Thirty-six e-mental health experts were invited to contribute with their knowledge, provide an overview of the current state-of-the-art and give practical suggestions, resulting in 25 authors and a think tank contributing actively (Fig. 1). The authors' expertise covers multiple disciplines (psychologists, psychiatrists, computer scientists and industry) with different working areas (clinical studies, (tele-)psychotherapy, mental health state assessment, development and conducting of digital interventions in the field of mental health, app development and artificial intelligence (AI)) in children, adolescents and adults around the globe. We sought diversity in terms of research seniority, culture and gender.

For finding consensus on relevant recommendations and guidelines for clinicians and for researchers, an adapted, structured Delphi procedure in nine steps based on iterative feedback and co-reviewing by the authors was implemented[14], guided by L.V., C. Seiferth and J.L. First, a list of the three most important objectives was discussed within the authors and agreed on: (1) development, (2) study specifics and (3) evaluation of e-mental health assessments and interventions. A total of 15 topics were brainstormed within the objectives (Terminology; Where to start; Content, Participatory research; Target group; Suicidality; Data protection and data security; AI in assessment and intervention; Sensing and wearables; Dropout rates and compliance; Efficacy evaluation; Ecological momentary assessment (EMA); Transfer into (clinical) practice; AEFs). Thereafter, the topic 'Dropout rates and compliance' was removed as a separate chapter, and the section 'Where

to start' and 'Terminology' as well as 'Target group' and 'Participatory research' were combined. Second, authors were grouped into teams due to their expertise and preference (two to four for each section). All author teams were asked to include the latest literature and findings, clear recommendations respecting their topic, potentially helpful links for further literature recommendations, and a list of 'dos and don'ts' for each topic (Supplementary Information). After all authors delivered their first drafts, L.V., C. Seiferth and J.L. reviewed the content and checked redundancies, and synthesized all parts into one piece on an online document, accessible and editable for all authors. Consequently, all participating authors were asked to review the whole paper and comment on each section regarding (1) discrepancies, (2) agreement with their own experience, (3) literature recommendations and (4) other comments. First authors of each section finally discussed and/ or integrated such comments with the support of C. Seiferth, L.V. and J.L., who again developed a second clean version that was then handed over to more senior researchers in the field (P.C., N.E., A.N. and T.R.) for a global check-up and proof of coherence. Minor issues (typos and references) were resolved by the first and last authors, and specific comments were fed back to the author teams and either discussed, integrated or dismissed. To achieve a final version and common 'ground truth', the last issues were syndicated, and all authors reviewed the again cleared paper and consented.

E-mental health development

Where to start? The implementation of any e-mental health project—assessment or/and intervention—is preceded by the fundamental decision about whether a digital approach is appropriate to address the specific research issue. Researchers need to identify the characteristics of the problem that allow for a digital operationalization (that is, multi-faceted, context sensitive and time sensitive) and the specificities of applying technology (that is, problem definition). Once it is established that a digital solution is the most appropriate approach, researchers can be clear about (I) their objectives, theory and hypotheses, which they aim to investigate. This may guide through several decisions that need to be taken throughout the process (Fig. 2). Furthermore, (II) the specification of the main target group (that is, demographics, mental health disorder and cultural background) and the target group involvement (that is, participatory research) is essential for following decisions, such as (III) the extent (for example, self-guided, partly guided and blended counselling) and nature (for example, on-demand, asynchronous, chat and video-based) of the delivered approach.

Digital technologies can be used to facilitate communication between practitioners and/or patients and can vary in their intensity of communication. Thus, the level of interaction between users and providers needs to be defined (for example, guided by a research team for technical support, or therapists). Furthermore, content transfer may range from passively reading a text versus clicking, and engaging more actively with the digital solution or with a coach/therapist. This, greatly depends on (IV) the chosen type of platform that is used to deliver the e-mental health service (for example, online and offline, browser or app). In this context, sensors can also be used (for example, touch, motion, pulse and gaze) to provide direct feedback about physical and emotional responses. This decision also depends on budget and collaboration with (external, potentially commissioned) technology companies, self-made toolkit supplies for e-health studies, in-house IT support and/or cooperation within a project with a technical partner.

For high-quality assessment and interventions, (V) best-practice and evidence-based components should show the foundation of digital solutions. Furthermore, the definition of (VI) the technical development process, including different disciplines, experiences, work cultures and (project) aims should not be underestimated. Ideally, an agile, iterative process in a multidisciplinary team is set up to develop and transfer psychological content into an attractive digital solution.

June 2022

> **Coordination and framework development**
> J.L., C.S. and L.V. determined scope, objectives and structure of the paper

July 2022

> **Finalization of the author team**
> J.L., C.S. and L.V. reached out to 35 researchers, clinical therapists, psychologists and computer scientists in the field of e-mental health to contribute to the paper (selection based on thematic expertise, publication activities and personal recommendation)

> Nine researchers did not want to contribute/did not respond

> **Development of the outline**
> Co-authors proposed further topics that were included in the outline. All co-authors agreed on 15 topics (that is, sections) within the three objectives

> **Coordination of the writing process**
> J.L., C.S. and L.V. proposed a key structure for each section (relevance of this topic in the context of e-mental health, state of the art, dos and don'ts and practical recommendations) that was send out to all author teams (two to three people per section)

August–September 2022

> **Finalization of first draft**
> All co-authors turned in their section by 12 September. J.L., C.S. and L.V. combined the sections, checked for redundancies and reviewed the overall structure, which resulted in sections being merged or removed as a separate section

October 2022

> **Feedback to the first draft**
> All co-authors were invited to provide feedback on the completed first draft until 21 October. J.L., C.S. and L.V. coordinated this review process and contacted authors wherever discrepancies or questions have arisen

November 2022

> **Finalization of second draft**
> J.L., C.S. and L.V. finalized the second draft and incorporated all changes

December 2022

> **Feedback to the second draft**
> P.C., N.E., A.N. and T.R. (who did not read the draft until this point) were invited to review the draft independently.

> **Finalization of the final draft**
> J.L., C.S. and L.V. incorporated the changes and finalized the final draft

Fig. 1 | Flow diagram illustrating the paper creation process. The following researchers are denoted by initials: P.C., Per Carlbring; N.E., Narges Esfandiari; J.L., Johanna Löchner; A.N., Alexandra Newbold; T.R., Tobias Renner; C.S., Caroline Seiferth; L.V., Lea Vogel.

Especially for the implementation of gamification features, interactive content and delivery logic, an interdisciplinary shoulder-to-shoulder working culture is most promising. Together with the technical experts, (VII) decisions about data flow, data storage, access and transparency need to be taken and the following procedure clearly defined. Study participants should be well informed about such details and comfy their trust in academic e-mental health research (as a quality criterion, diverging from more commercially driven supplies). Following these

I.	Objectives, theory and hypotheses
II.	Definition and involvement of target group
III.	Extent and nature of the delivered approach
IV.	Type of platform
V.	Best-practice and evidence-based components
VI.	Technical development process
VII.	Data flow, data storage, access and transparency
VIII.	Risk management
IX.	Proof of objective and hypotheses
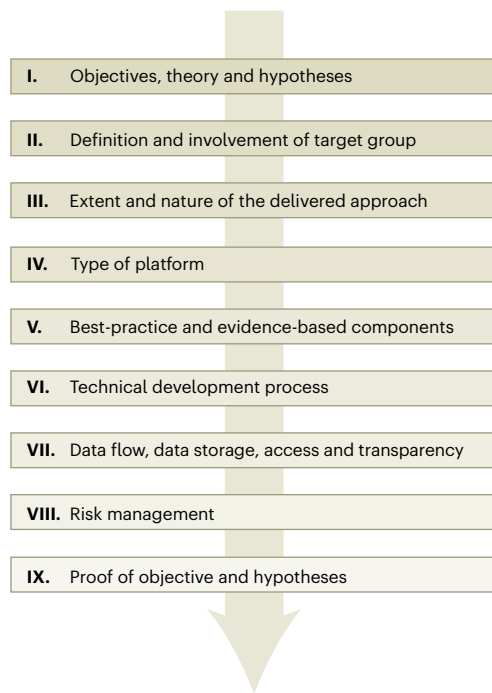
Fig. 2 | E-mental health study conceptualization process. The figure illustrates nine major stages that need to be considered when developing a study in the context of e-mental health.

steps, the (VIII) risk management strategies and dropout prevention may be defined. As a final step, the research team may determine (IX) what study design best suits the proof of objective and hypotheses. Naturally, those defined steps interact, are dynamic and need to be reconsidered during the whole process. In addition, other specific frameworks and guidelines may support researchers and clinicians in their project planning decisions[15–20].

Intervention content development. The process of content development for a multicomponent e-mental health intervention is two-staged. First, researchers need to select psychological and psychotherapeutic strategies based on existing evidence or best-practice approaches for the selected target group and intervention aim. Second, the components of the intervention need to be transferred to the digital solution. This technical translation poses a range of pitfalls and therefore requires a highly iterative and dynamic research approach that should take place within a multidisciplinary team (for example, mental health professionals, software engineers and design experts[6,21]).

A pragmatic approach comprises converting existing resources, such as applying psychological content from text-based manuals, exercises or questionnaires in agreement with the original authors. However, information displayed in digital solutions follows a rather different temporal and architectural structure and the user engages with the app with a different 'user mindset' because app use occurs at varying times, with varying intensity, in varying contexts. To consider these peculiarities of the digital environment, a substantial amount of time, financial costs, perspectives and tests must be dedicated to the process of transforming specific components of traditional health interventions. More concretely, this means that each piece of content must be condensed to the core aims and elements that are to be conveyed through the digital solution. It is important that the structure (for example, division into modules, sessions/lessons and exercises), delivery logic (for example, temporal availability of content) and complexity of content is always set against the background of the targeted group, and outcomes of the intervention. Once the crucial

elements and user needs are determined, a user experience story may be developed.

Engaging elements may enhance a positive and reinforcing environment (for example, text, audiovisual, prompts, quizzes, self-report questions and gamification features). Although forced guidance through an assessment or intervention may be needed to address the research objectives, flexibility and the personalization of features (that is, just-in-time adaptive interventions[22]) are likely to be beneficial to increase the attractiveness of an app-based or smart-phone intervention[23]. In general, the content should match the 'look and feel' of the digital format (for example, length of a video and amount of text displayed[6,24]). It is also necessary to consider which resources are realistically available to the development team and if it is possible to develop new, customized multimedia elements. Finally, content development should be specifically focused on the target group.

User-centered design and participatory approaches. The implementation of user engagement participatory research within the development of e-health interventions is currently recognized as a way to increase the ease of use as well as the likelihood to fit the users' needs. It is therefore recommended to limit common problems such as low uptake, high complexity and poor fit to the user's needs. Participatory research actively involves end users, healthcare professionals and other stakeholders in all stages of the development and research process (including the formulation of the research question and goal, planning research design, selection of research methods and outcomes, interpretation of results, and dissemination) by taking into account their views, needs, expectations and cultural background[25,26]. For a participatory approach, it is mandatory that end users also participate in the decision-making processes[27].

In the field of e-health intervention development, user-centered design (UCD) has been established in recent years[28,29]. UCD represents a systematic, iterative process with three phases during development[30]. First, an initial investigation of the users' needs should be conducted (for example, differentiating children, adolescents, adults and elderly users). The purpose of the first phase is to identify the needs of the target group, and to identify features and characteristics of the intervention that would be acceptable and preferred. For example, strategies such as personalization, gamification and including a social component have been identified as important for the users' engagement[31,32]. Focus groups or interviews with future users or individuals in their environment (for example, therapists) and/or open-ended written survey questions are suitable methods for user needs assessments. Qualitative research methods (for example, thematic analyses) are suitable for establishing UCD guidelines[33]. Second, a prototype with key features of the intervention should be created, which can be used in usability tests[30]. During the third step, usability tests, researchers observe potential users interacting with the prototype in a controlled environment, while they are simultaneously thinking aloud[34]. Researchers take notes about the participant's behaviors, comments and issues, to uncover and adapt functional and design flaws[29,30]. This phase is a balancing act between drawing evidence-informed strategies and content from the literature and combining them with ways of delivering this information in an acceptable and engaging way. Continuing to engage with the target group at this stage ensures that, when the product is finalized, the target group has been involved and has provided continual feedback and guidance throughout the process, maximizing the likelihood that the final product will meet the needs of the users. It must be noted that UCD represents a preliminary stage of participative research as participation takes on a strictly consultative role and the project's decisions are still in the control of the researchers[35]. To achieve meaningful participation, it is necessary to involve end users as early as possible in the research process and in all decision-making processes.

Focusing on the target group's specific needs is particularly important when it comes to digital interventions, with notable variability

in the aspects of technology that will appeal to different groups of users[36]. Clinical observation shows less adherence when participants expressed a wide range of needs but the digital treatment addresses a single disorder[37]. When identifying the target group, specifics that should be considered include age, gender, cultural (racial and ethnic) background, delivery context and delivery format. This information can guide the best ways to engage with the specific target groups at the outset of the project. Additional questions relate to the answer if it is recommended by the mental health professionals. In a UCD, all potential specificities must be explored together with the target group[38].

Once the specifics of the target group have been identified, the next step is to conduct an appropriate stakeholder engagement process with all parties involved in the delivery, dissemination and implementation of the intervention, as well as the end users. In addition, the examination of the usual consumer behavior by the specific target group may be helpful, for example, what kinds of health apps are used, how often and what features are more or less appealing.

Study specifics

Managing suicidality. E-mental health research is often conducted with participants recruited via the internet without any face-to-face contact throughout the entire research process. As a result, both researchers and institutional review boards express great uncertainty about how to manage participants who are experiencing severe mental health crises such as suicidal thoughts or behavior (STB)[39,40]. In common practice (not only in e-mental-health research), individuals with a history of suicidal behavior or who affirm suicide-related questionnaire items (for example, item 9 of the Patient Health Questionnaire-9 (PHQ-9)) are often excluded from trials at baseline[40]. This practice, however, results in almost no increase in safety for participants, because it overlooks that suicidality often is a highly fluctuating symptom[41] and study participants may conceal their suicidal ideation to be admitted to the study[42]. Moreover, while there is an established association between suicidal ideation and previous suicidal attempts with subsequent suicidal behavior, their practical predictive utility in differentiating individuals who are likely to exhibit suicidal behavior from those who are not is limited[43]. Indeed, most people who die by suicide do not score in commonly used suicide risk assessments[44]. Thus, excluding participants who score on suicidality items primarily reduces the external validity of study results[45], which poses potential risks to users when these interventions are implemented in real-world care.

Given the impossibility of eliminating the risk of suicidality in e-mental-health research, we propose implementing the following measures to increase participant safety during the intervention, as it has been practiced in prior randomized controlled trials (RCTs) of digital interventions specifically designed for individuals with STB[46]. The assessment of STB should be expanded, including the use of specifically validated questionnaires[47]. At any point where participants may potentially report suicidality (for example, in the intervention or questionnaires), it must be ensured that this is noticed by the study team. The study protocol should explicitly outline how to react to reports of STB. It is crucial to ensure that all designated team members receive appropriate training and supervision to effectively implement specified procedures. This reaction can, but does not necessarily need to, include a telephone or other contact by the study team. However, in case of a disclosed immediate and definite plan for suicide, the country-specific emergency services should be informed. Participants should be clear about these procedures as well as about the time frame within which their entries will be seen by a member of the study team. We recommend documenting this in the informed consent. When STB is reported, detailed and visible information on support and contact services (for example, national emergency numbers and 24-hour help lines) should be provided automatically including low-threshold click-to-call links. The use of other forms of treatment should not lead to an exclusion from the trial. Instead, individual crisis plans should be developed together with the participant. For studies with particularly vulnerable study samples, a collaboration with local emergency centres should be arranged in advance. In intervention trials for mental health disorders, optional modules that specifically target STB should be available[46,48]. In general, help options should be equally available to all participants, irrespectively of their group allocation and the type of intervention[35].

Data protection and data security. Substantial deficiencies in data protection and data security may inhibit e-mental health assessment and intervention studies[49,50]. The focus of data security is to prevent unwanted data loss and the unauthorized manipulation of data. The protection of personal data (for example, patient contact details) is of uttermost importance in e-mental health applications.

In the development of an e-mental health offering, it must be anticipated that users may unintentionally reveal their access data, lose their devices or use the devices for other (harmful) actions (for example, children visiting adult websites). To counter these problems, tools can be installed on devices that lock access to other content. Preconfigured and password-protected study smart-phones should be used. Two-factor authentication prevents mass registrations by fake users that can lead to poor data quality. In any case, users should be thoroughly informed about typical problems and dangers. This also applies to harmful software that the user captured unintentionally (for example, keyloggers and spyware spy on sensitive data).

Further challenges include incorrect programming, which can enable unauthorized access to sensitive data. Therefore, a quality-assured software development process is essential[51]. If a manufactured app is used, the data should be stored in the healthcare institution's storage facilities rather than in the manufacturer's cloud. An external data hosting service provider should be certified. No data should be stored permanently on the device of the user, and a virus and Trojan scanner should be installed. Immediate data transfer instead of data storage on the device as well as automated data backups could also ensure data quality. To prevent an attack where data traffic is intercepted, manipulated or deleted, an end-to-end encryption (via Transport Layer Security/Secure Sockets Layer (TLS/SSL)) should be used to transfer data. There should be brute-force attack protection built into the platform and all information in the database should be encrypted using a high-end algorithm with separate keys for each study.

The most effective measure is the pseudonymization of sensitive data, which makes it worthless for unauthorized persons without any additional information[52]. The process of pseudonymization and internal de-pseudonymization of the data must take place in a separate system[52] and be considered even before the selection or development of an e-mental health system. Data protection and transparency are especially relevant for the use of AI methods.

AI in assessment and intervention. AI holds great promise for e-mental health, largely owing to the advances in affective computing. The latter includes the analysis, synthesis and reaction to the affect and emotions of humans using the former. The past decade has seen major progress thanks to the rise of deep learning as an enabler in (generative) AI[53].

Likewise, great progress has been made in the recognition of emotion (for example, in categories or dimensions such as arousal and valence), depression (for example, in 'dimensions' such as depression assessment questionnaires as the Beck Depression Inventory-II (BDI-II) or PHQ-9), or other mental health disorders[54,55]. The means of assessment serve mostly audio (for example, speech), video (for example, facial expression, body posture and gait), text (written or spoken language) and physiology (for example, via heart rate and skin conductance). A series of research competitions (for example, Audio/Visual Emotion Challenge (AVEC))[56] have been benchmarking the progress of the community including tasks of the above from these modalities.

Additionally, several reports exist on successful emotion and depression analysis from phone usage data, touch and other information sources[57]. At the same time, readily available toolkits independent of the application domain and target tasks are sparsely available as 'out of the box' solutions. Usually, training these to match the target domain and target task is required. Also, robustness of real-world applications 'in-the-wild' has increased notably over the past decade[58]. However, not all free-for-research and beyond solutions include state-of-the-art de-noising, target person verification or features such as subject adaptation. In addition, while such solutions often work largely independently of the subject, most of these tools are mostly geared towards a specific culture or language, or another context, due to the data they were trained upon. For practical solutions, this usually requires re-training of such tools or 'engines' on the target data.

Emotion can also be synthesized with increasing realism by AI and recently deep learning approaches—often reaching human-level or close-to-human-level quality for speech and image or even language rendering. This led to effective virtual agents such as the 'sensitive artificial listeners' that may be implemented in clinical practice for assessment and interventions. Again, platforms are available open source and free for research, but usually require some adaptation to the target task. Most notably, the AVEC challenge series had recently hosted the first ever 'AI only' depression challenge, where interviews were conducted by an AI, and the recognition of depression severity was also conducted by AI reaching competitive results concerning human assessment considering the subjective nature of the task.

The recent past brought further breakthroughs in AI and particularly deep learning by the advent of transformer architectures, and diffusion approaches enabling a next generation of abilities in recognition of affect, and generation. This era is also coined by the 'foundation models': these extensive data pre-trained models are marked (1) by convergence, that is, rather than training 'your own model' from scratch, the trend is to use these models and fine-tune them to one's needs, which led to considerable improvements in a field, where data is continuously (too) scarce; and (2) by emergence. The latter is fascinating, as, while these models may not have been trained on tasks in affective computing or such relevant to e-mental health, they may show emergent skills in these stemming from the sheer 'big' quantities of data they were trained upon. The well-known ChatGPT (a general pre-trained transformer (GPT)) was recently shown to predict suicidal risk at competitive performance 'out of the box' levels[59]. This is without fine-tuning and training on the task, when compared with traditional and deep approaches fully trained on the task[59]. Similarly, Dall-E 2—also based on GPT—can paint emotional faces from verbal descriptions, arguably also emergent behavior from the perspective of affective computing. In short, we seemingly enter an era in which e-mental health relevant skills can emerge in AI of the present and the future, whereby these 'foundation model' AIs are often trained on internet-scale data such as all of Wikipedia, several years of speech, or millions of images. Such models could render even explicit training of tasks increasingly obsolete. In combination with the increasing power of generative AI (GenAI), interventions could be produced in a rich manner including questioning and chatty communication, potentially including the audiovisual rendering of artificial therapists, which are highly personalized and socio-emotionally empowered. Current foundation models such as GPT-4 or Metaverse as virtual space may be only a sneak preview of the oncoming power and abilities, which may help overcome the uncanny valley of such artificial therapists and help AI get to know patients better than any human depending on their data access. Accordingly, they might also soon be able to influence us in strange ways.

Potential dangers may relate to AI-driven chatbots or generative AI such as ChatGPT, which can be charismatic and appear emotionally involved due to expressing emotions (with emojis or an empathic language)[60]. This interaction may implicate an image of being a friend or human but if assumed would be highly unethical. Because individuals with mental illness are a vulnerable group, often longing for appreciation and security in social contacts, great emphasis must be put on ethical guidelines. Bot-based interaction must be recognized as non-human to minimize the possibility of manipulation and harm or even dependence on the interaction with such AI—potentially at the cost of human relationships[61]. Furthermore, such AI may find its ways of behavior, which may be even more persuasive and change the human-to-human behavior of such interacting with it in the long run. Asking participants to use sensing and wearable data collection tools can often provide supplementary data to support AI research methods.

Sensing and wearables. Historically, diagnosing mental health conditions has relied on thoroughly validated self-report questionnaires. While questionnaire-based assessments are an indispensable source of information in this context, they are based purely on introspection, can lack vital information that is systematically neglected by the patient (for example, due to self-other knowledge asymmetry[62]), are temporally constrained (that is, one-time, infrequent assessments), limited in granularity (that is, in terms of a selection of questions in the anamnesis) and suffer from floor and ceiling effects (that is, lack sensitivity to change at their scale's extremes). Foremost, it is becoming increasingly apparent that patients are likely to be not able to self-report the fine-grained and complex patterns of behaviors in various situations of daily life that characterize their physical and psychological traits, states and changes in these.

The ongoing evolution in mobile sensing and mobile computing and communication technologies ameliorates this situation. More and more sophisticated and accurate sensors in consumer electronics (for example, smart-phones or wearables) allow for the unobtrusive and automated collection of high-frequency, objective, longitudinal data on human behaviors, states and environmental conditions[63,64]. Figure 3 provides an overview of the variety of data that off-the-shelf consumer electronics sensors can provide.

On a growing scale, mobile sensing data are increasingly being used throughout the health and behavioral sciences to understand behavioral aspects of mental health through digital biomarkers[65–67] to detect health conditions and deterioration[68,69] and improve conditions through behavioral interventions[70,71].

While mobile sensing is becoming increasingly established as a method in mental health research, its standardization is challenging due to rapid and frequent changes in hardware, operating systems, and ethical and legal frameworks, among others[12,72]. Participants should be aware of often liberal data storage and access policies of companies. While this circumstance has acted as a roadblock in the past, the main mobile operating systems have started to develop standardized so-called application programming interfaces for researchers to access and use in empirical studies (for example, Android Health Connect (https://android-developers.googleblog.com/2022/05/introducing-health-connect.html), Apple SensorKit (https://developer.apple.com/documentation/sensorkit) or HealthKit (https://developer.apple.com/documentation/healthkit)).

However, the most innovative methods can be useless if they miss the mark. While offering specific new opportunities, e-mental health interventions need to be evaluated properly.

Evaluation
Efficacy evaluation, RCTs and other methods. There is no shortage of available e-mental health interventions, most of which are not well evaluated[73]. However, despite the young age of the field, high-quality evidence is needed from the start, as unreliable results can stick around in a classic canon of literature[74] and lead to low quality of developments, or even harm the patients. This section offers recommendations and ideas for how to produce this high-quality evidence. While there are some unique ways to evaluate e-mental health interventions, which
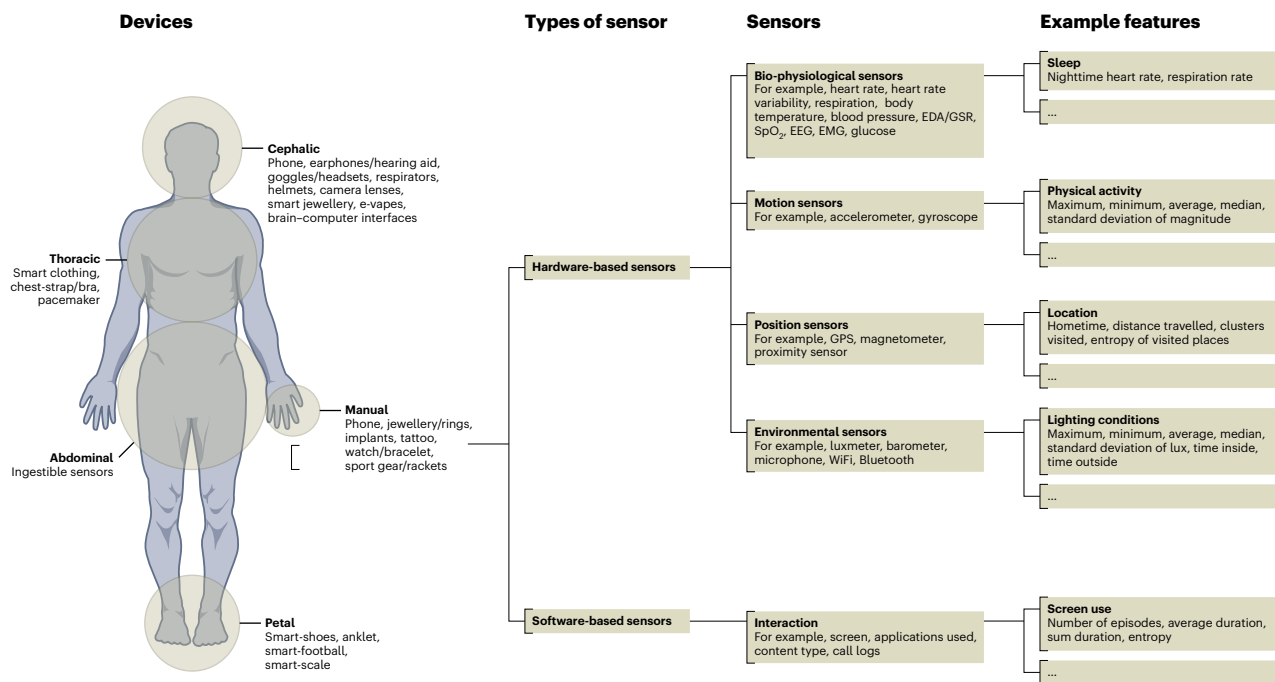
Fig. 3 | Overview of mobile sensors embedded in consumer electronics and variables they provide[117]. EEG, electroencephalography; EDA/GSR, electrodermal response/galvanic skin response; EG, electroencephalography; EMG electromyography; GPS, global positioning system; SpO$_2$, peripheral capillary oxygen saturation.

will be addressed below, a good starting point for an evaluation study are the same principles that apply to classic interventions: besides observational or case studies, the gold standard to evaluate (mental) health interventions are RCTs. First meta-analyses of e-mental health intervention RCTs show promising effects, even when compared with face-to-face treatments, but also that primary studies have been focusing on a small range of diagnoses and age groups[75]. A high variance in types of control groups and interventions further reduces the amount of knowledge that can be gained from meta-analyses. Therefore, the field would benefit from further RCTs addressing these issues.

When setting up and selecting variables for an evaluation study of an e-mental health intervention, past studies of classical interventions can serve as an example. Researchers and practitioners still need to investigate any potential adverse treatment effects[76] and the importance of mediators and moderators of treatment effects that apply in face-to-face settings (that is, symptom severity, self-efficacy, motivation, age or amount of therapist involvement). Special attention should be paid to therapist effects, which robustly explain a relevant amount of variance in classical treatment outcomes[77]. For the evaluation of e-mental health interventions, the type of the application (stand-alone, prescribed after seeing a professional, and continued blended care) can influence which therapist effects are present. There might be none if there is no therapist involved, they might be similar to classical mental care, or they might be even stronger, for example, when negative biases of a professional towards digital solutions are present. Studies should aim for an extensive and diverse pools of therapists, also because estimates of the therapist-level random slope suffer from more bias when there are very few therapists in a study[78].

Going beyond these traditional evaluation standards, evaluations of e-mental health interventions offer exciting new possibilities. The underlying technological infrastructure has the potential to extend the classical outcome-oriented designs and measures as it becomes more achievable to measure various process variables. These can focus on psychological content, such as therapeutic relationships (for example, rupture–repair[79]), sudden gains/losses[80] or personalized items and networks[81,82]. Time series data on an individual level will allow new

hypotheses to be answered. Also, by using shifting time windows, one can produce a meta-time series of, for example, dynamic variance or critical fluctuations and use their change as an outcome variable[80,83,84]. Another possibility is to evaluate individuals' network parameters (for example, networks of symptoms) and their change over time or recurrence plot quantification[85,86]. In short, the type and amount of data from e-mental health studies can change the classical approach of aggregating first (across participants) and analyze second to analyze first (on the individual level) and aggregate second. Therefore, e-mental health studies have the huge potential to expand the concept of traditional RCTs. Going beyond RCTs, further methodological approaches (for example, A/B testing and trials of principles) can be used to test small differences within an intervention or to test the efficacy of a general principle of an electronic solution (for example, self-monitoring). These approaches of agile science might contribute to the reduction of the time discrepancies between technical development and evaluation results[21], which is especially important when working with fast-changing technologies. As a specific option for evaluation, EMA will be discussed in the next section.

EMA. EMA (synonyms: ambulatory assessment, experience sampling method and real-time data capture) encompasses a range of methods that involve repeated assessments of individuals' dynamical experiences and behaviors in their natural habitat, thereby increasing both ecological validity and generalizability, while minimizing recall biases[87,88]. This method can be used in various stages of the therapeutic process (for example, diagnostic process, tracking the course of symptoms during treatment and transfer of therapeutic effects thereafter).

EMA offers the possibility to combine subjective assessments with further methods (for example, psychophysiological and physical activity assessments)[87]. EMA also allows for integrating continuous mobile sensing (that is, digital phenotyping[89]) to predict critical phases[90] and to improve the timing of EMA enquiries[87]. By providing a detailed picture of mental state and functioning, EMA promises to be more sensitive to capturing change and, thereby, improving the assessment of the therapeutic effects of interventions[91]. One of the most promising

avenues of EMA is the opportunity to extend treatment beyond the clinical setting into real life using e-mental health applications[92].

When setting up a study, the following aspects are very important: there are various sampling designs (that is, time-based, event-based, combined sampling schemes). Choose the one that fits your research question. Carefully balance the length of the questionnaire presented at each assessment, the number of assessments per day, and the assessment epoch to ensure high compliance rates[93]. Also, allow participants to delay or actively decline alarms. Choose an adequate time frame for the questions. Whereas questions referring to the present moment minimize retrospective bias, those with a specific time interval enhance representativity. When deciding on the order, group items with the same time frames, and ask transitory constructs (for example, emotions) first and questions that are not likely influenced by preceding questions (for example, context) at the end. If you must develop new items, use two, or better three, per construct, to be able to determine the items' reliability[94]. A crucial point is that the sampling strategy must fit the temporal dynamics of the underlying process; otherwise, results can be misleading[95].

Carefully determine the length of the EMA period that is needed to answer the research questions. However, balancing the lengths is key to ensuring participants' compliance[93,96]. Meta-analytic results revealed higher compliance rates in studies offering monetary incentives compared with other or no incentives[93]. Moreover, linking the incentives to a certain degree of compliance might reduce dropouts during the assessment period[96].

Transfer into (clinical) practice. To make e-health interventions feasible for real-world settings, the following criteria should be considered. (1) Research should integrate follow-up measurements to assess long-term usage since there is a lack of meta-analysis on long-term benefits of mental health apps as the handling of follow-up measurements and dropouts is inconsistent[75,97]. Indeed, reviews showed that too few studies used (long-term) follow-up measurements and many showed huge dropout rates of 47% (refs. [98,99]). (2) Researchers, developers and practitioners should consider relevant factors to improve adherence to digital health interventions in real-world contexts[100]. When looking into real-life settings, ref. [99] found in over 10,000 digital mental health apps only 11 peer-reviewed publications analyzed uptake and usage data in such real-life settings. The completion rate was between 44% and 99% in RCTs but dropped to 1–28% when looking at real-world usage. Furthermore, new (machine learning) approaches showed that a distinction into user subtypes and, therefore, personalization of interventions could lessen the effects of interventions[101]. Thus, researchers and developers should consider relevant factors to improve adherence to digital health interventions in real-world contexts. (3) Integrate mood monitoring, feedback and human/automated support to lower dropout rates[102]. For example, dropout rates decreased by 46% when therapeutic support was provided and even minimal care with only administrative support resulted in a meaningful decline in dropout rates[103]. Further, it has been shown that when specific data of EMA are fed back to clients regularly, the number of missing EMA data is low (<10%) and reduces over time[84]. Digital health is a global challenge, but the implementation of digital health interventions is based on complex national and local economic and political processes. (4) Hence, when conceptualizing and evaluating the implementation process of e-mental health interventions, researchers and practitioners should always consider the integration of all relevant stakeholders that will be involved in the final roll-out of the digital interventions, such as lived experience users and beneficiaries, companies, health insurance or other political institutions and decision-makers. We argue that for each digital intervention a unique approach for its roll-out should be considered and developed along with its scientific evaluation. Target groups, clinical scope (prevention or intervention), business models, funding strategies, long-term technical maintenance, requirements for quality management, regularity frameworks, data safety, market access and reimbursement schemes are only some examples to be considered. The exploitation of evidence-based interventions may further benefit substantially from the flexibility, variety of resources and agile methods of industrial partners. Even where the process is successful, any ongoing quality control in clinical practice is substantial and very challenging in a dynamic field of tech industries. Furthermore, potential side effects tend to be underestimated, leading to a broad supply of unapproved interventions.

AEFs. As the number and diversity of e-mental health solutions increases, so does the need to evaluate which are most effective and safe. While regulatory bodies are beginning to approach the regulation of primarily mobile and web-based apps but also other sorts of digitally delivered interventions, most efforts remain nascent[104,105]. This means clinicians and patients must rely on tools such as AEF to help them make more informed decisions. While there are also an increasing number of AEF, there are differences in their approaches with some providing scales versus ratings or subjective versus objective metrics, and others information versus databases. Each approach has a unique value depending on the use case and clinical needs.

Perhaps the largest category of app evaluation is scales or frameworks that provide guidance and information on how to consider an app. For example, the American Psychiatric Association's app evaluator framework[106] provides a four-step process with corresponding questions about privacy, efficacy, engagement and clinical utility. While this framework does not provide scores or ratings, there are other frameworks such as the Mobile Application Rating Scale[107] that do. Often these rating systems require training before they can be properly used. A study[108] reviewed popular frameworks through the lens of diversity, equity and inclusion and found that only 58% included related metrics that offer a target for future efforts and evaluation criteria, whether subjective or objective.

A related consideration in app evaluation is the use of subjective versus objective metrics. For example, questions about the aesthetics or usability are inherently subjective and will vary between users. Examples of objective metrics may include videos or music in an app. Each approach has merits and subjective evaluations, often in the form of user or expert reviews that can provide rich contextual information about an app. However, it can be challenging to keep these reviews updated and current in the rapid-paced world of apps[109]. Objective metrics may not offer such context but often provide easier-to-update approaches that may have higher inter-rater reliability by their very nature. One example of such an approach is the Mobile App Index and Navigation Database (mindapps.org), which rates apps across 105 criteria that are derived from largely objective criteria[110].

A further consideration is how users can engage with any AEF, whether it offers subjective or objective metrics, frameworks or ratings. Some approaches, such as Psyberguide, Mobile App Index and Navigation Database, and the UK National Health Service Apps Library, maintain websites that users can search, while others provide only the rating scale or related educational material. The impact of either approach remains unstudied although recent research suggests that digital literacy and health app awareness are important related factors for app use[111]. Some newer approaches such as the adapted Mobile Application Rating Scale have been proposed, with the authors suggesting the need for concomitant support from a coach or digital navigator[112].

Discussion

The COVID-19 pandemic boosted the supply of digitally delivered assessments and treatments, along with the promise of increased availability of digital, low-threshold treatment for mental illness worldwide. However, e-mental health research and practice is still in an embryonic stage of development and evidence base. This expert consensus provides a comprehensive essence of scientific knowledge and practical

recommendations for both practitioners and researchers (see 'Dos and don'ts' in Supplementary Information).

In summary, when researchers agree on using a digital approach, they should define the development and evaluation process carefully, guided by their main objective and theory. The technical transformation of psychological content requires a transdisciplinary, participative and highly iterative research process that swallows up a range of resources that need to be estimated when planning the project. Data protection and the inclusion of vulnerable groups pose challenges to the successful implementation and should be given special consideration while adhering to current standards during the development and implementation by a user-centered approach. AI holds great promises for e-mental health, largely owing to the advances in affective computing. Mobile sensing and EMA have a huge potential to enable accurate assessments of individuals' daily life states and behaviors, which can be used for diagnostics and evaluation of interventions. RCTs are one element of an evaluation strategy for e-mental health, but should expand their focus to a wider range of populations, control groups and investigation of process variables and individual data patterns. AEF offer useful heuristics to identify apps of interest, and to share data on privacy, efficacy, engagement and clinical integration but cannot replace clinical judgement. To use e-mental health interventions in the context of prevention and intervention, it should be feasible for real-world settings.

Despite all technical innovations and novel features, one should not neglect the final goal of all attempts: health that is defined as "the ability to adapt and self manage in the face of social, physical, and emotional challenges"[113] enabling an individual to work, have relationships and express oneself in a society. Well-executed e-mental health assessments and interventions have the potential to be meaningful to individuals and their care teams. Rather than replacing health professionals, digital technologies have a potential to support the human therapeutic process in a scalable format (for example, by providing interactive psychoeducation and monitoring material). With this expert consensus, we aim to provide guidance on best practice when preventing or improving the state of mental illness using e-mental health interventions. The target group with its needs and wishes should be placed in the foreground to achieve this. Of paramount importance are high ethical standards, transparency, respect for patients' rights, data protection, the special consideration of vulnerable groups, and the explicit clarification of any non-human interactions. In the dynamic field, it is also necessary to rethink the classical statistical methods of development and evaluation of digital tools in order to not to lag behind the market in practice. It may be advisable to include flexible, iterative, built-in behavioral digital assessments and process monitoring and to collaborate with companies to provide exploitation and implementation with a viable business model, as digital intervention's availability and sustainable quality of digital intervention depend on it.

We further emphasize the need for feasibility and long-lasting usage based on long-term follow-up measurements, evaluation in real-world settings, considering factors that improve adherence, and integration of relevant stakeholders in the conceptualization, implementation, exploitation and roll-out process of the digital intervention as well as the integration of comprehensive process quality evaluation based on recognized international frameworks. Suicidality should not be an exclusion criterion, but it should be included once an adequate monitoring policy is established. Another groundbreaking development is generative AI outcomes such as GPT-4 or Metaverse, as virtual space may be only a sneak preview of the oncoming power and abilities supporting mental health aid. However, such AI, if not governed, might also soon have the ability to influence the mental health domain in unknown ways. Ethical considerations must be prioritized shielding potential harm and manipulation, especially in vulnerable groups such as individuals with mental illness.

To support the development and usage of high-quality e-mental health offers, the research field must be expanded, especially regarding long-term efficacy analysis, adherence, patient satisfaction and improved uptake, cost–benefit analysis and a facilitated access for clinicians and patients to evidence-based interventions within the routine healthcare. We claim a proof of efficacy as a prerogative of transfer into clinical practice, since side effects of e-mental health interventions have not yet been studied comprehensively and may be underestimated.

This expert consensus is unsystematic, primarily based on the subjective experience of a selective group of clinical and scientific experts, mostly from Europe, the USA and Australia. This is noteworthy, since a key advantage of e-mental health is providing low-threshold access to mental health services in particular in more rural regions with a limited supply of healthcare services. Given the increasing number of smart-phone users worldwide and the expanding reach of mental health apps within high-income countries[114], providing early scalable psychoeducation and online training in a stadium of mild symptoms or for early recognition, digital assessment and intervention methods are especially promising in preventing more severe courses with low costs[115]. In addition, the geographical range and duration of treatment can be expanded, for example, by using video calls and chats to stabilize clients through follow-up sessions, preventing relapse and providing care at a low cost. Moreover, cost-effectiveness studies are scarce and especially neglect low-income countries where e-mental health services may be particularly beneficial[115,116].

References

1. World Health Organization. *Mental Health and COVID-19: Early Evidence of the Pandemic's Impact: Scientific Brief* (2022).
2. Adair, C. E. et al. Continuity of care and health outcomes among persons with severe mental illness. *Psychiatr. Serv.* **56**, 1061–1069 (2005).
3. Schulte, J., Schulz, C., Wilhelm, S. & Buhlmann, U. Treatment utilization and treatment barriers in individuals with body dysmorphic disorder. *BMC Psychiatry* **20**, 69 (2020).
4. van Daele, T. et al. Online consultations in mental healthcare: modelling determinants of use and experience based on an international survey study at the onset of the pandemic. *Internet Interv.* **30**, 100571 (2022).
5. Paganini, S., Teigelkötter, W., Buntrock, C. & Baumeister, H. Economic evaluations of internet- and mobile-based interventions for the treatment and prevention of depression: a systematic review. *J. Affect. Disord.* **225**, 733–755 (2018).
6. Mendes-Santos, C., Nunes, F., Weiderpass, E., Santana, R. & Andersson, G. Understanding mental health professionals' perspectives and practices regarding the implementation of digital mental health: qualitative study. *JMIR Form. Res.* **6**, e32558 (2022).
7. Atik, E., Schückes, M. & Apolinário-Hagen, J. Patient and therapist expectations for a blended cognitive behavioral therapy program for depression: qualitative exploratory study. *JMIR Ment. Health* **9**, e36806 (2022).
8. de Witte, N. A. J. et al. Online consultations in mental healthcare during the COVID-19 outbreak: an international survey study on professionals' motivations and perceived barriers. *Internet Interv.* **25**, 100405 (2021).
9. Smoktunowicz, E. et al. Consensus statement on the problem of terminology in psychological interventions using the internet or digital components. *Internet Interv.* **21**, 100331 (2020).
10. O'Loughlin, K., Neary, M., Adkins, E. C. & Schueller, S. M. Reviewing the data security and privacy policies of mobile apps for depression. *Internet Interv.* **15**, 110–115 (2019).
11. Hennemann, S., Farnsteiner, S. & Sander, L. Internet- and mobile-based aftercare and relapse prevention in mental disorders: a systematic review and recommendations for future research. *Internet Interv.* **14**, 1–17 (2018).

12. Torous, J. et al. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. *J. Technol. Behav. Sci.* **4**, 73–85 (2019).

13. Lal, S. & Adair, C. E. E-mental health: a rapid review of the literature. *Psychiatr. Serv.* **65**, 24–32 (2014).

14. Pill, J. The Delphi method: substance, context, a critique and an annotated bibliography. *Socio Econ. Plan. Sci.* **5**, 57–71 (1971).

15. Bartlett Ellis, R. et al. Lessons learned: beta-testing the digital health checklist for researchers prompts a call to action by behavioral scientists. *J. Med. Internet Res.* **23**, e25414 (2021).

16. Shen, F. X. et al. An ethics checklist for digital health research in psychiatry: viewpoint. *J. Med. Internet Res.* **24**, e31146 (2022).

17. World Health Organization. *Digital Implementation Investment Guide (DIIG): Integrating Digital Interventions into Health Programmes* https://apps.who.int/iris/bitstream/handle/10665/334306/9789240010567-eng.pdf (2020)/

18. Unsworth, H. et al. The NICE Evidence Standards Framework for digital health and care technologies—developing and maintaining an innovative evidence framework with global impact. *Digit. Health* **7**, 20552076211018617 (2021).

19. Sundareswaran, V. & Sarkar, A. *Chatbots RESET: A Framework for Governing Responsible Use of Conversational AI in Healthcare* https://www3.weforum.org/docs/WEF_Governance_of_Chatbots_in_Healthcare_2020.pdf (World Economic Forum, 2020).

20. Doraiswamy, P. M. et al. Empowering 8 billion minds: enabling better mental health for all via the ethical adoption of technologies. *NAM Perspect.* https://doi.org/10.31478/201910b (2019).

21. Hekler, E. B. et al. Agile science: creating useful products for behavior change in the real world. *Transl. Behav. Med.* **6**, 317–328 (2016).

22. Fiedler, J., Seiferth, C., Eckert, T., Woll, A. & Wunsch, K. A just-in-time adaptive intervention to enhance physical activity in the SMARTFAMILY2.0 trial. *Sport Exerc. Perform. Psychol.* https://doi.org/10.1037/spy0000311 (2022).

23. Chan, A. H. Y. & Honey, M. L. L. User perceptions of mobile digital apps for mental health: acceptability and usability—an integrative review. *J. Psychiatr. Ment. Health Nurs.* **29**, 147–168 (2022).

24. Vial, S. & Boudhraâ, S. in *Revolutions in Product Design for Healthcare* (eds Subburaj, K. et al.) 21–34 (Springer, 2022).

25. Narayan, S., Mok, H., Ho, K. & Kealy, D. I don't think they're as culturally sensitive: a mixed-method study exploring e-mental health use among culturally diverse populations. *J. Ment. Health* **32**, 241–247 (2022).

26. Wright, M. T., Springett, J. & Kongats, K. in *Participatory Health Research* (eds Wright, M. T. & Kongats, K.) 3–15 (Springer, 2018).

27. Wright, M. T. Partizipative Gesundheitsforschung: Ursprünge und heutiger Stand. *Bundesgesundheitsblatt Gesundheitsforsch. Gesundheitsschutz* **64**, 140–145 (2021).

28. Carman, K. L. et al. Patient and family engagement: a framework for understanding the elements and developing interventions and policies. *Health Affairs* **32**, 223–231 (2013).

29. Mummah, S. A., Robinson, T. N., King, A. C., Gardner, C. D. & Sutton, S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. *J. Med. Internet Res.* **18**, e317 (2016).

30. McCurdie, T. et al. mHealth consumer apps: the case for user-centered design. *Biomed. Instrum. Technol.* https://doi.org/10.2345/0899-8205-46.s2.49 (2012).

31. Knowles, S. E. et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. *PLoS ONE* **9**, e84323 (2014).

32. Werner-Seidler, A. et al. A smartphone app for adolescents with sleep disturbance: development of the Sleep Ninja. *JMIR Ment. Health* **4**, e28 (2017).

33. Braun, V. & Clarke, V. in *APA Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological* (eds Cooper, H. et al.) 57–71 (American Psychological Association, 2012).

34. Dopp, A. R., Parisi, K. E., Munson, S. A. & Lyon, A. R. A glossary of user-centered design strategies for implementation experts. *Transl. Behav. Med.* **9**, 1057–1064 (2019).

35. Orlowski, S. K. et al. Participatory research as one piece of the puzzle: a systematic review of consumer involvement in design of technology-based youth mental health and well-being interventions. *JMIR Hum. Factors* **2**, e12 (2015).

36. Orji, R. & Moffatt, K. Persuasive technology for health and wellness: state-of-the-art and emerging trends. *Health Inform. J.* **24**, 66–91 (2018).

37. Esfandiari, N. et al. A specific internet-based cognitive behavior therapy for adolescents with social anxiety disorder: three-armed randomized control trial. Preprint at *Research Square* https://doi.org/10.21203/rs.3.rs-2123795/v1 (2022).

38. Yardley, L., Bradbury, K. & Morrison, L. in *Qualitative Research in Psychology: Expanding Perspectives in Methodology and Design* 2nd edn (ed. Camic, P. M.) 263–282 (American Psychological Association, 2021).

39. Bailey, E. et al. Ethical issues and practical barriers in internet-based suicide prevention research: a review and investigator survey. *BMC Med. Ethics* **21**, 37 (2020).

40. Sander, L. et al. Suicide risk management in research on internet-based interventions for depression: a synthesis of the current state and recommendations for future research. *J. Affect. Disord.* **263**, 676–683 (2020).

41. Kaurin, A., Dombrovski, A. Y., Hallquist, M. N. & Wright, A. G. C. Integrating a functional view on suicide risk into idiographic statistical models. *Behav. Res. Ther.* **150**, 104012 (2022).

42. Friedlander, A., Nazem, S., Fiske, A., Nadorff, M. R. & Smith, M. D. Self-concealment and suicidal behaviors. *Suicide Life Threat Behav.* **42**, 332–340 (2012).

43. Franklin, J. C. et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol. Bull.* **143**, 187–232 (2017).

44. Steeg, S. et al. Accuracy of risk scales for predicting repeat self-harm and suicide: a multicentre, population-level cohort study using routine clinical data. *BMC Psychiatry* **18**, 113 (2018).

45. Sisti, D. A. & Joffe, S. Implications of zero suicide for suicide prevention research. *JAMA* **320**, 1633–1634 (2018).

46. Büscher, R. et al. Digital cognitive-behavioural therapy to reduce suicidal ideation and behaviours: a systematic review and meta-analysis of individual participant data. *Evid. Based Ment. Health* **25**, e8–e17 (2022).

47. Holmes, E. A. et al. The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. *Lancet Psychiatry* **5**, 237–286 (2018).

48. Torok, M. et al. Suicide prevention using self-guided digital interventions: a systematic review and meta-analysis of randomised controlled trials. *Lancet Digit. Health* **2**, e25–e36 (2020).

49. Ferreira, T. E-health applications and data protection: a comparison of selected European Union members' national legal systems. *Bioethica* **8**, 74–84 (2022).

50. Wilkowska, W. & Ziefle, M. Privacy and data security in E-health: requirements from the user's perspective. *Health Inform. J.* **18**, 191–201 (2012).

51. Albrecht, U.-V. *Chances and Risks of Mobile Health Apps (CHARISHMA)* (Medizinische Hochschule Hannover, 2016).

52. Lablans, M., Borg, A. & Ückert, F. A RESTful interface to pseudonymization services in modern web applications. *BMC Med. Inf. Decis. Making* **15**, 2 (2015).

53. Cummins, N., Schuller, B. W. & Baird, A. Speech analysis for health: current state-of-the-art and the increasing impact of deep learning. *Methods* **151**, 41–54 (2018).

54. Ghosh, S., Löchner, J., Mitra, B. & De, P. in *Quantifying Quality of Life* (eds Wac, K. & Wulfovich, S.) 209–267 (Springer, 2022).

55. Garriga, R. et al. Machine learning model to predict mental health crises from electronic health records. *Nat. Med.* **28**, 1240–1248 (2022).

56. Valstar, M. et al. AVEC 2013: the continuous audio/visual emotion and depression recognition challenge. In *Proc. 3rd ACM International Workshop on Audio/Visual Emotion Challenge* 3–10 (ACM, 2013).

57. Wang, R. et al. Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2**, 1–26 (2018).

58. Dhall, A., Goecke, R., Gedeon, T. & Sebe, N. Emotion recognition in the wild. *J. Multimodal User Interf.* **10**, 95–97 (2016).

59. Amin, M. M., Cambria, E. & Schuller, B. W. Will affective computing emerge from foundation models and general AI? A first evaluation on ChatGPT. *IEEE Intelligent Systems* **2**, 15–23 (2023).

60. Schuller, B. W. et al. Computational charisma—a brick by brick blueprint for building charismatic artificial intelligence. Preprint at https://doi.org/10.48550/arXiv.2301.00142 (2023).

61. Véliz, C. Chatbots shouldn't use emojis. *Nature* **615**, 375 (2023).

62. Vazire, S. Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *J. Person. Soc. Psychol.* **98**, 281–300 (2010).

63. Garatva, P. et al. in *Digital Phenotyping and Mobile Sensing* (eds Montag, C. & Baumeister, H.) 395–411 (Springer, 2023).

64. Torous, J., Kiang, M. V., Lorme, J. & Onnela, J.-P. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Ment. Health* **3**, e16 (2016).

65. Cornet, V. P. & Holden, R. J. Systematic review of smartphone-based passive sensing for health and wellbeing. *J. Biomed. Inform.* **77**, 120–132 (2018).

66. Moshe, I. et al. Predicting symptoms of depression and anxiety using smartphone and wearable data. *Front. Psychiatry* **12**, 625247 (2021).

67. Vasudevan, S., Saha, A., Tarver, M. E. & Patel, B. Digital biomarkers: convergence of digital health technologies and biomarkers. *npj Digit. Med.* **5**, 36 (2022).

68. Asare, O. K. et al. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. *JMIR mHealth uHealth* **9**, e26540 (2021).

69. Fried, E. I., Rieble, C. & Proppert, R. K. K. Building an early warning system for depression: rationale, objectives, and methods of the WARN-D study. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/9qcvs (2022).

70. Lattie, E. G. et al. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. *J. Med. Internet Res.* **21**, e12869 (2019).

71. Torous, J. et al. Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: a review of current progress and next steps. *Curr. Psychiatry Rep.* **20**, 51 (2018).

72. Kargl, F., van der Heijden, R. W., Erb, B. & Bösch, C. in *Digital Phenotyping and Mobile Sensing* (eds Montag. C. & Baumeister, H.) 13–23 (Springer, 2023).

73. Larsen, M. E. et al. Using science to sell apps: evaluation of mental health app store quality claims. *npj Digit. Med.* **2**, 18 (2019).

74. Klein, R. A. et al. Investigating variation in replicability. *Soc. Psychol.* **45**, 142–152 (2014).

75. Simblett, S., Birch, J., Matcham, F., Yaguez, L. & Morris, R. A systematic review and meta-analysis of e-mental health interventions to treat symptoms of posttraumatic stress. *JMIR Ment. Health* **4**, e14 (2017).

76. Magnusson, K., Andersson, G. & Carlbring, P. The consequences of ignoring therapist effects in trials with longitudinal data: a simulation study. *J. Consult. Clin. Psychol.* **86**, 711–725 (2018).

77. Johns, R. G., Barkham, M., Kellett, S. & Saxon, D. A systematic review of therapist effects: a critical narrative update and refinement to review. *Clin. Psychol. Rev.* **67**, 78–93 (2019).

78. Rozental, A., Andersson, G. & Carlbring, P. In the absence of effects: an individual patient data meta-analysis of non-response and its predictors in internet-based cognitive behavior therapy. *Front. Psychol.* **10**, 589 (2019).

79. Mander, J. et al. The therapeutic alliance in different mental disorders: a comparison of patients with depression, somatoform, and eating disorders. *Psychol. Psychother.* **90**, 649–667 (2017).

80. Mechler, J. et al. Sudden gains and large intersession improvements in internet-based psychodynamic treatment (IPDT) for depressed adolescents. *Psychother. Res.* **31**, 455–467 (2021).

81. von Klipstein, L., Riese, H., van der Veen, D. C., Servaas, M. N. & Schoevers, R. A. Using person-specific networks in psychotherapy: challenges, limitations, and how we could use them anyway. *BMC Med.* **18**, 345 (2020).

82. van den Bergh, R. et al. The content of personalised network-based case formulations. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/yan4k (2022).

83. Olthof, M. et al. Destabilization in self-ratings of the psychotherapeutic process is associated with better treatment outcome in patients with mood disorders. *Psychother. Res.* **30**, 520–531 (2020).

84. Schiepek, G. et al. Real-time monitoring of psychotherapeutic processes: concept and compliance. *Front. Psychol.* **7**, 604 (2016).

85. Hasselman, F. & Bosman, A. M. T. Studying complex adaptive systems with internal states: a recurrence network approach to the analysis of multivariate time-series data representing self-reports of human experience. *Front. Appl. Math. Stat.* https://doi.org/10.3389/fams.2020.00009 (2020).

86. Wallot, S. Recurrence quantification analysis of processes and products of discourse: a tutorial in R. *Discourse Process.* **54**, 382–405 (2017).

87. Myin-Germeys, I. & Kuppens, P. *The Open Handbook of Sampling Methodology. A Step-by-Step Guide to Designing, Conducting, and Analyzing ESM studies* (The Center for Research on Experience Sampling and Ambulatory Methods Leuven, 2021).

88. Trull, T. J. & Ebner-Priemer, U. Ambulatory assessment. *Annu. Rev. Clin. Psychol.* **9**, 151–176 (2013).

89. Insel, T. R. Digital phenotyping: a global tool for psychiatry. *World Psychiatry* **17**, 276–277 (2018).

90. Ebner-Priemer, U. W. et al. Digital phenotyping: towards replicable findings with comprehensive assessments and integrative models in bipolar disorders. *Int. J. Bipolar Disord.* **8**, 35 (2020).

91. Myin-Germeys, I. et al. Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry* **17**, 123–132 (2018).

92. Fortea, L. et al. Development and validation of a smartphone-based app for the longitudinal assessment of anxiety in daily life. *Assessment* https://doi.org/10.1177/10731911211065166 (2021).

93. Wrzus, C. & Neubauer, A. B. Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. *Assessment* https://doi.org/10.1177/10731911211067538 (2022).

94. Geldhof, G. J., Preacher, K. J. & Zyphur, M. J. Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychol. Methods* **19**, 72–91 (2014).

95. Kockler, T. D., Santangelo, P. S. & Ebner-Priemer, U. W. Investigating binge eating using ecological momentary assessment: the importance of an appropriate sampling frequency. *Nutrients* **10**, 105 (2018).

96. Ottenstein, C. & Werner, L. Compliance in ambulatory assessment studies: investigating study and sample characteristics as predictors. *Assessment* https://doi.org/10.1177/10731911211032718 (2021).

97. Lecomte, T. et al. Mobile apps for mental health issues: meta-review of meta-analyses. *JMIR mHealth uHealth* **8**, e17458 (2020).

98. Linardon, J., Cuijpers, P., Carlbring, P., Messer, M. & Fuller-Tyszkiewicz, M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. *World Psychiatry* **18**, 325–336 (2019).

99. Fleming, T. et al. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. *J. Med. Internet Res.* **20**, e199 (2018).

100. Torous, J. et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry* **20**, 318–335 (2021).

101. Chien, I. et al. A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. *JAMA Netw. Open* **3**, e2010791 (2020).

102. Torous, J., Lipschitz, J., Ng, M. & Firth, J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. *J. Affect. Disord.* **263**, 413–419 (2020).

103. Richards, D. & Richardson, T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clin. Psychol. Rev.* **32**, 329–342 (2012).

104. Alon, N., Stern, A. D. & Torous, J. Assessing the Food and Drug Administration's risk-based framework for software precertification with top health apps in the United States: quality improvement study. *JMIR mHealth uHealth* **8**, e20482 (2020).

105. Stern, A. D. et al. Advancing digital health applications: priorities for innovation in real-world evidence generation. *Lancet Digit. Health* **4**, e200–e206 (2022).

106. Lagan, S. et al. Mental health app evaluation: updating the American Psychiatric Association's framework through a stakeholder-engaged workshop. *Psychiatr. Serv.* **72**, 1095–1098 (2021).

107. Stoyanov, S. R. et al. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR mHealth uHealth* **3**, e27 (2015).

108. Ramos, G., Ponting, C., Labao, J. P. & Sobowale, K. Considerations of diversity, equity, and inclusion in mental health apps: a scoping review of evaluation frameworks. *Behav. Res. Ther.* **147**, 103990 (2021).

109. Carlo, A. D., Hosseini Ghomi, R., Renn, B. N. & Areán, P. A. By the numbers: ratings and utilization of behavioral health mobile applications. *npj Digit. Med.* **2**, 54 (2019).

110. Lagan, S. et al. Actionable health app evaluation: translating expert frameworks into objective metrics. *npj Digit. Med.* **3**, 100 (2020).

111. Szinay, D. et al. Influences on the uptake of health and well-being apps and curated app portals: think-aloud and interview study. *JMIR mHealth uHealth* **9**, e27173 (2021).

112. Roberts, A. E. et al. Evaluating the quality and safety of health-related apps and e-tools: adapting the Mobile App Rating Scale and developing a quality assurance protocol. *Internet Interv.* **24**, 100379 (2021).

113. Huber, M. et al. How should we define health? *BMJ* **343**, d4163 (2011).

114. Wasil, A. R., Gillespie, S., Shingleton, R., Wilks, C. R. & Weisz, J. R. Examining the reach of smartphone apps for depression and anxiety. *Am. J. Psychiatry* **177**, 464–465 (2020).

115. Gentili, A. et al. The cost-effectiveness of digital health interventions: a systematic review of the literature. *Front. Public Health* **10**, 787135 (2022).

116. Mitchell, L. M., Joshi, U., Patel, V., Lu, C. & Naslund, J. A. Economic evaluations of internet-based psychological interventions for anxiety disorders and depression: a systematic review. *J. Affect. Disord.* **284**, 157–182 (2021).

117. Wac, K. in *Digital Health* (eds Rivas, H. & Wac, K.) 83–108 (Springer, 2018).

Author contributions

Competing interests

Additional information

[1]Institute of Psychology, University of Bamberg, Bamberg, Germany. [2]National Centre for Early Prevention, German Youth Institute, Munich, Germany. [3]Department of Psychology, LMU Munich, Munich, Germany. [4]sysTelios Think Tank, sysTelios Health Center, Siedelsbrunn, Germany. [5]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital of Psychiatry and Psychotherapy, Tuebingen, Germany. [6]DZPG (German Center of Mental Health), Tuebingen, Germany. [7]Department of Psychology, Stockholm University, Stockholm, Sweden. [8]PFH – Private University of Applied Sciences, Department of Psychology (Clinical Psychology II), Goettingen, Germany. [9]Department of Psychology, Faculty of Education and Psychology, Shahid Beheshti University, Tehran, Iran. [10]Medical Faculty, University of Tuebingen, Tuebingen, Germany. [11]Department of Education and Health Research, Institute of Sport Science, University of Tuebingen, Tuebingen, Germany. [12]Sir Henry Wellcome Building for Mood Disorders Research, College of Life and Environmental Sciences, University of Exeter, Exeter, UK. [13]Medical Psychology and Medical Sociology, Faculty of Medicine, University of Freiburg, Freiburg, Germany. [14]Department of Behavioural and Cognitive Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg. [15]LMU Munich, Munich, Germany. [16]GLAM, Imperial College, London, UK. [17]Institute of Behavioral Science and Technology, University of St. Gallen, St. Gallen, Switzerland. [18]Department of Clinical Psychology and Psychotherapy, University of Ulm, Ulm, Germany. [19]Beth Israel Deaconess Medical Center, Department of Psychiatry, Harvard Medical School, Boston, MA, USA. [20]Quality of Life Lab, Geneva School of Economics and Management, University of Geneva, Geneva, Switzerland. [21]Black Dog Institute, UNSW, Sydney, New South Wales, Australia. [22]These authors contributed equally: Caroline Seiferth, Lea Vogel. ✉e-mail: Johanna.Loechner@med.uni-tuebingen.de

sysTelios Think Tank

Benjamin Aas[3,4]

A full list of members and their affiliations appears in the Supplementary Information.