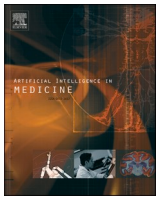
Artificial Intelligence In Medicine 160 (2025) 103052

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed



AI-enabled clinical decision support tools for mental healthcare: A product review

Anne-Kathrin Kleine [a,*], Eesha Kokje [a], Pia Hummelsberger [a], Eva Lermer [a,b], Insa Schaffernak [b], Susanne Gaube [c]

[a] LMU Munich, Germany
[b] Technical University of Applied Sciences Augsburg, Germany
[c] University College London, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Keywords
Artificial intelligence
Mental healthcare
Clinical decision support
Medical device regulation

ABSTRACT

The review seeks to promote transparency in the availability of regulated AI-enabled Clinical Decision Support Systems (AI-CDSS) for mental healthcare. From 84 potential products, seven fulfilled the inclusion criteria. The products can be categorized into three major areas: diagnosis of autism spectrum disorder (ASD) based on clinical history, behavioral, and eye-tracking data; diagnosis of multiple disorders based on conversational data; and medication selection based on clinical history and genetic data. We found five scientific articles evaluating the devices' performance and external validity. The average completeness of reporting, indicated by 52 % adherence to the Consolidated Standards of Reporting Trials Artificial Intelligence (CONSORT-AI) checklist, was modest, signaling room for improvement in reporting quality. Our findings stress the importance of obtaining regulatory approval, adhering to scientific standards, and staying up-to-date with the latest changes in the regulatory landscape. Refining regulatory guidelines and implementing effective tracking systems for AI-CDSS could enhance transparency and oversight in the field.

1. Introduction

Mental health disorders account for the majority of years lived with disability worldwide [1]. Approximately one-third of the world population experiences a mental health disorder at some point in their lifetime, and only one-third of individuals with mental health disorders receive the treatment they need [1,2]. The absence of timely diagnosis and appropriate early interventions for mental disorders can lead to a significant increase in the severity of the condition, longer recovery times, a decreased quality of life, and an overall increase in healthcare costs [3–5]. The use of artificial intelligence (AI) for mental health services comes with the promise of mitigating some of these challenges [6–8]. For instance, AI-driven algorithms can be applied on patterns identified in speech and video data to generate diagnostic and treatment suggestions [6,9,10]. The implementation of such algorithms allows for a broad spectrum of patient data to be harnessed, potentially leading to earlier and more precise detection and treatment of mental disorders [11–14].

AI-enabled Clinical Decision Support Systems (AI-CDSS) can be implemented in clinical settings to assist mental health practitioners by diagnosing or predicting the onset of conditions and the response to treatments [15–22]. Despite their potential to facilitate evidence-based practice, the integration of AI-CDSS for mental healthcare into clinical settings remains limited [8,17,19,23]. As AI-CDSS solutions for mental healthcare meet the qualifications of medical devices, they are subject to regulatory compliance prior to commercial distribution. Although regulatory oversight is designed to guarantee patient safety and the efficacy of AI-CDSS, it is recognized as a barrier to their integration into the healthcare system [24,25]. Numerous AI-CDSS for mental healthcare have been developed and patented [8,15–17,20]. However, only a limited number of tools have undergone regulatory procedures and are currently eligible for commercial sale as medical devices [26,27]. Transparency in the availability and characteristics of regulated AI-CDSS for mental healthcare enables healthcare administrators to select devices that may augment the care delivered [28]. Information about the regulatory status of available devices is vital for healthcare administrators to make informed decisions on which AI-CDSS devices can be safely and legally incorporated into their healthcare delivery systems.

* Corresponding author at: LMU Munich, Geschwister-Scholl-Platz 1, 80539 Munich, Germany.
E-mail address: anne-kathrin.kleine@psy.lmu.de (A.-K. Kleine).

https://doi.org/10.1016/j.artmed.2024.103052
Received 13 December 2023; Received in revised form 27 September 2024; Accepted 5 December 2024
Available online 6 December 2024

Policymakers need to be aware of the current regulatory landscape to identify potential gaps in regulation and develop policies that ensure the safe use of AI-CDSS while facilitating innovation in healthcare technology. Therefore, the first goal of the systematic review is to present a clear and comprehensive overview of regulated AI-CDSS in mental healthcare, including their regulatory status and characteristics to help understand the gap between research and practical application.

Regulatory approval[1] does not guarantee external validity as the extent to which the outcomes and conclusions derived from using an AI-CDSS can be generalized across different settings, patient populations, and times [29]. The absence of public information regarding the external validity of AI-CDSS makes it difficult to justify their integration into clinical practice [30–32]. Therefore, the second goal of the systematic review is to provide information about available scientific evidence addressing the external validity of the identified tools [29,33].

1.1. AI-CDSS for mental healthcare

AI-CDSS are point-of-care systems designed to assist in clinical decision-making. These systems can be used to derive more accurate diagnoses, prognoses, and treatment plans [7,20,24,34,35]. The integration of AI transforms the decision-making process by facilitating a personalized approach that takes into account the unique characteristics of each patient [10,36,37]. AI-CDSS use large amounts of patient data, encompassing medical records, genetic information, and lifestyle factors, to create predictive models specifically tailored to the individual needs of each patient [36,38]. Generally, the data is employed to train AI algorithms that are continually refined until their predictive ability is deemed satisfactory. Ultimately, the algorithms may be used to make predictions for new patient data [36].

AI-CDSS have been developed for a multitude of mental disorders and prediction aims (i.e., diagnosis, prognosis, treatment) using various data sources and providing recommendations in different output formats [18,20,21,24,38,39]. The diverse features and capabilities of AI-CDSS are also mirrored in available patent data [20]. While research and patent data yield insights into technological developments and the intellectual property landscape, they do not shed light on the characteristics of tools that are regulated and may thus be used in clinical practice [7,25]. We propose the following research question to guide the systematic examination of the characteristics of regulated AI-CDSS for mental healthcare:

> Research Question 1: What are the characteristics of regulated AI-CDSS for mental healthcare in terms of targeted disorder(s), data sources used to derive recommendations and data collection methods, prediction aims (i.e., diagnosis, prognosis, treatment recommendation), and generated outputs?

1.2. Regulation of AI-CDSS for mental healthcare

AI-CDSS fit the classification of medical devices as they are designed for the diagnosis, cure, mitigation, treatment, or prevention of disease. Therefore, these systems require regulatory approval before they can be sold commercially [40,41]. Due to the high-risk nature and potential unknown effects of employing AI-CDSS in mental healthcare, strict regulations must be adhered to.

The most recognized regulatory certifications worldwide include approval or clearance by the Food and Drug Administration (FDA) in the United States (US), the Conformité Européene (CE) mark in the European Economic Area (EEA), and the United Kingdom (UK) Conformity Assessed (UKCA[2]) mark [28,42–44]. The FDA distinguishes three risk classes, ranging from the lowest risk Class I to the highest risk Class III. The FDA's regulatory pathways include a) premarket approval (usually for Class III devices), which entails an exhaustive review that requires scientific evidence of safety and efficacy, b) the 510(k) pathway (usually for Class II devices), which entails the evaluation of the safety and efficacy of a device based on a similar, previously approved (i.e., predicate) device, and c) the less rigorous De Novo premarket review that includes checks for safety and effectiveness for new technologies without a predicate device [45–47]. Currently, most AI-CDSS are regulated through the 510(k) pathway [43,48]. For CE marking in the EEA, adherence to the Medical Device Regulation (MDR) is required. Four risk classes are distinguished: Class I (low risk), Class IIa and IIb (moderate and medium risk), and Class III (high risk), each reflecting the potential risk associated with the device to patients and users [23,49]. In vitro diagnostic devices (IVDs), including AI software that analyzes test results from patient samples, are regulated under the In Vitro Diagnostic Regulation (IVDR) [50]. Similar regulatory processes apply for UKCA marking [51]. According to Rule ii of Annex VIII in the MDR, decision-support systems generally fall under class IIa devices (moderate risk), unless they may seriously affect the patient's state of health, in which case they may fall under class IIb (medium risk) or class III (high risk) [23]. The regulation of Class II devices requires the evaluation of the conformity with the required standard by a Notified Body, an organization designated by an EU Member State to assess the conformity of certain products before being placed on the market. These conformity assessments focus on the product's safety and performance, quality system evaluations, and clinical evaluations, ensuring transparency, safety, and efficacy [52]. We propose the following research question to guide the systematic examination of the regulatory status of AI-CDSS for mental healthcare:

> Research Question 2: What is the regulatory status of regulated AI-CDSS for mental healthcare in terms of type of regulation, risk class, and regulatory pathway?

1.3. External validation of AI-CDSS for mental healthcare

Most AI-CDSS designed for mental healthcare are far from reaching generalizability [8,38,53]. For example, differences in devices used for collecting data, coding standards, digital health record systems, lab equipment, as well as differences in sample characteristics and clinical and administrative procedures may impede the generalization of AI-generated recommendations across multiple settings [53]. However, inaccurate models, or models that function well only with the data they were trained on, can lead to severely erroneous judgments if applied in clinical practice [53].

Evaluating real-world clinical performance and establishing generalizability requires structured external validation. This involves examining AI-CDSS using sufficiently large datasets distinct from those that contributed to the model training. This method ensures the representation of diverse patient demographics and different site characteristics that occur in real-world settings where the system will be deployed [53,54]. External validation of AI-CDSS for mental healthcare is still rarely implemented. Findings from a recent systematic review assessing

[1] Please note that we use the term ‘regulatory approval’ to refer to all types of regulatory authorization, including FDA approval, CE marking and UKCA marking.

[2] Following Brexit, medical devices must obtain the UKCA mark to be sold in Great Britain (GB). The UKCA mark, while essentially equivalent to the CE in terms of the standards it adheres to, is GB's own certification mark. The UKCA mark is not recognized in the EEA and, therefore, cannot replace the CE mark for products being sold in the EEA.

AI algorithms for application in mental healthcare point out that only 14 % of the 153 eligible studies that were published included external validation [38]. Even regulated AI-CDSS often lack external validation [28,30,32,55]. However, external validation is necessary to guarantee that AI-CDSS are safe and effective across different clinical settings and patient populations [30–32]. Accordingly, we propose the following research question to guide the systematic examination of available scientific evidence targeting the external validation of AI-CDSS for mental healthcare:

> Research Question 3: What is the quality and reproducibility of research targeting the external validation of regulated AI-CDSS for mental healthcare?

Beyond the conformity to CONSORT-AI standards, we seek to present insight into the performance of the identified AI-CDSS based on the sensitivity and specificity. Sensitivity and specificity are commonly used metrics for assessing the performance of an AI-CDSS. Sensitivity, or the true positive rate, assesses the ability to accurately identify individuals with a given condition. Specificity, or the true negative rate, measures the system's ability to correctly identify individuals without the condition. Together, these metrics provide a comprehensive picture of the AI-CDSS's ability to provide accurate recommendations, encompassing both its power to detect true cases and to refrain from false alarms [56,57]. We propose the following research question regarding the performance of regulated AI-CDSS according to external validation studies:

> Research Question 4: What is the sensitivity and specificity of regulated AI-CDSS for mental healthcare according to external validation studies?

2. Material and methods

Our primary aim was to provide a preliminary investigation and a comprehensive overview of the topic, focusing on the identification of regulated AI-CDSS for mental healthcare and their characteristics. Given the exploratory nature of our review, registration and a review protocol were not deemed necessary. Nonetheless, we followed the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [58] to ensure the transparency and reliability of our review process.

2.1. Product inclusion criteria

We applied four inclusion criteria for the inclusion of products into the review. First, the recommendations generated had to refer to mental disorders. Consequently, we excluded products aimed at disorders unrelated to mental health and those targeting non-clinically relevant mental health issues. For example, we excluded IntelliSpace Cognition as a digital assessment platform that supports assessing the cognition of individuals, without addressing mental health disorders. As part of this exclusion criterion, we also excluded wellness applications and devices, such as *ApolloNeuro*.

Second, the product had to include AI software in its primary functionality. We deemed technology to be AI-based if the words or phrases ‘deep learning’, ‘machine learning’, ‘neural networks’, ‘predictive modeling/algorithms’, or ‘artificial intelligence’ were used to describe the technology in official regulatory documents or on vendor websites [25,59]. For instance, we excluded *Quadrant Biosciences* because the services and devices offered for the detection of autism were not described as AI-based. Additionally, we excluded *Monsenso*, an application for tracking patient mental health issues, because all AI-based applications were indicated to be used only for research purposes, thus not being part of the commercialized product.

Third, products had to offer mental health decision-support, meaning they had to be designed to assist mental health professionals with diagnosis, prognosis, or the generation of treatment recommendations for mental disorders according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR) [60]. This led to the exclusion of pharmaceutical products (e.g., *BioXcel Therapeutics*); products that may not directly be used by the treating mental health clinician, such as *EnsoSleep*, a device for the diagnosis of sleep disorders to be used by polysomnographic technologists; and tools offering mental health treatment without providing diagnostic and treatment advice to treating mental health professionals, such as *EndeavorRx*, a video game treatment for children with attention-deficit/hyperactivity disorder (ADHD).

Finally, we only included products cleared by the FDA or by EEA/UK regulatory agencies with a CE/UKCA mark. We confirmed FDA clearance using the relevant FDA medical device database [61]; CE markings using the European Database on Medical Devices (EUDAMED) [62][3]; and UKCA marking using the Public Access Registration Database (PARD) [63]. We excluded all unregulated or not yet regulated products.

2.2. Search strategies and coding of product information

We implemented several strategies to find products online. First, we used data collected during a patent review of AI-enabled tools for mental healthcare [20]. Not all patents are commercialized as products and not all products are patented. However, there exists substantial overlap between patents and novel products [64]. In July 2023, AK examined the assignee[4] details of patents to identify relevant products. Comprehensive details about the patent search methodology and selection criteria for the patent review can be found in the patent review [20].

Second, the exhibitor lists from relevant conferences and exhibitions for mental health practitioners and corporations that market AI-CDSS for mental healthcare were searched for relevant products. In July 2023, AK, SG, PH, and EK independently reviewed the exhibitor lists of the American Psychological Association Congress, the American Psychiatric Association Congress, the Association for Behavioral and Cognitive Therapies, the Anxiety and Depression Association of America, and the German Psychotherapy Congress.

Finally, curated and regularly updated databases listing regulated AI medical devices were used to identify relevant products. In July 2023, AK reviewed the FDA Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices Database [66], the Healthskouts database [67], the Medical Futurist database [68], and the AI Watch database [69] for relevant products.

The classification of products was performed in three steps. First, product or company names were identified from the patent data, exhibitor lists, and online databases as described above (no filters apply). Second, the identified product or company names were used to search for vendor websites online. Finally, the information provided on the vendor websites was carefully scrutinized to determine whether a product met the inclusion criteria. If there were any uncertainties regarding the inclusion of products, decisions were reached through weekly consensus meetings involving all authors. Additionally, all vendors with available contact information were contacted to verify and add to the collected information. If no answer was received, we included only the data collected from the vendor website and regulatory documents. Conflicts about whether to include or exclude a product based on the inclusion criteria were resolved in regular consensus meetings. We coded the company's entry date, targeted disorders, input data, generated output, product description, product primary purpose, and information about the regulatory status, including the date of regulatory

[3] Please note EUDAMED is still in transition so its usage is not compulsory.

[4] The assignee is the individual or entity, such as a company, to whom the inventor, or their legal representative, has transferred ownership of the patent. The assignee possesses the legal rights to the patent, including control over its use, sale, or licensing [65].

approval, the FDA/CE risk class, the regulatory pathway, and information on predicate devices (if available). In April 2024, we conducted a follow-up search of products that met all inclusion criteria but were not regulated during the initial round of coding.

2.3. Inclusion criteria, search strategies, and coding of scientific evidence targeting external validation

We collected published scientific evidence targeting the external validation of the included products in two ways. First, we searched the PubMed database by vendor and product name for papers published until July 2023. Queries and search results can be accessed through the Online appendix (https://osf.io/vrfm7/?view_only=e027649e3a42428a8975f7385659b55c). Second, we conducted a manual search by reviewing all vendors' websites for related papers in July 2023. A follow-up search on PubMed and vendor websites was conducted in April 2024.

Original papers were included if they mentioned the product name (the device of interest) or a former product name, were peer-reviewed, and in English. Furthermore, the evidence had to target external validation, i.e., be reported using an independent dataset on which the algorithm was not trained [29]. Since Ada Assess may also be used for decision-support for issues unrelated to mental health and Limbic Access may also be used as an e-triage tool, we additionally excluded publications that did not focus on decision-support for mental disorders. Papers were evaluated by two of the authors (AK and PH) who independently screened the title, abstract, and complete paper for inclusion criteria. Any disagreements were resolved in consensus meetings.

To assess the quality and reproducibility of the study results, we utilized the CONSORT-AI (Consolidated Standards of Reporting Trials Artificial Intelligence) checklist [70,71]. Comprising of 51 criteria, the CONSORT-AI checklist is designed to standardize the presentation of study results and may be used to evaluate the quality and reproducibility of clinical studies and study results [70,71]. In compliance with prior research, an item was scored '1' if all information for that criterion was reported, and '0' if information for that item was missing or incomplete [72,73]. The initial scoring was independently completed by two of the authors (AK and PH). Each scoring was discussed and agreed upon in consensus meetings. An independent coding was subsequently performed by a trained research assistant, yielding an interrater agreement score of 96.86 % (247/255). Finally, AK collected the performance information reported in the research papers. If no information in sensitivity and specificity was available, other performance indicators are reported.

3. Results

The results section is organized as follows: First, we present an overview of the inclusion and exclusion of products based on the inclusion criteria. Second, we present information about the

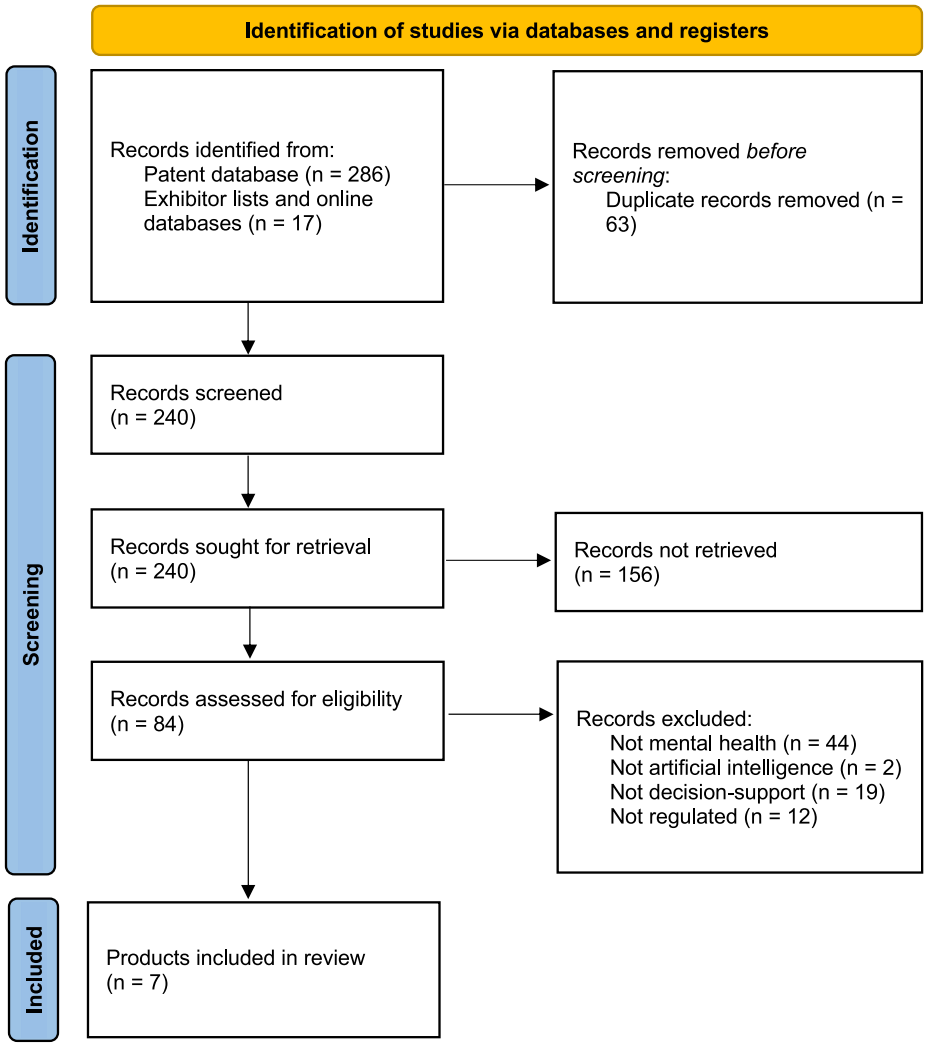


Fig. 1. PRISMA flow diagram.

characteristics and regulatory status of regulated products (Research Questions 1 and 2). Finally, we present information on the quality of research targeting the external validation of regulated products and reported sensitivity and specificity (Research Questions 3 and 4).

3.1. Inclusion and exclusion of products

The PRISMA diagram in Fig. 1 shows the process from the initial identification of products to their inclusion in the review. Of the initial 240 product names identified from the patent database, exhibitor lists, and other online databases, 84 product websites (35 %) could be found. Of the 84 products found on vendor websites, seven (8 %) met our inclusion criteria. In total, we contacted 29 vendors during the initial round of coding to obtain additional information (e.g., the CE risk class), of which six (21 %) responded and three (10 %) provided the requested information. No information was missing for the decision whether to include a product based on the inclusion criteria. However, the presentation of results is limited to the information we obtained from vendor websites, regulatory documents, and through the few email exchanges.

3.2. Product characteristics and regulation of AI-CDSS for mental healthcare

Fig. 2 displays the dates of company foundation, product certification, and the publication of scientific evidence. As can be seen from the figure, the total time passed between firm entry and certification decreased over time, indicating that firms that were founded later needed less time until their products were cleared. Scientific evidence is available for two of the seven products.

Table 1 provides an overview of the product characteristics. We identified three groups of products that may be characterized by common targeted disorders, data sources, and data collection methods, namely the diagnosis of autism spectrum disorder based on clinical history, behavioral, and eye-tracking data, the diagnosis of multiple disorders based on conversational data, and medication selection based on clinical and genetic data.

3.2.1. Diagnosis of autism spectrum disorder based on clinical history, behavioral, and eye-tracking data

Two devices were designed for use by healthcare providers to assist in the diagnosis of autism spectrum disorder (ASD) in children, namely the *Cognoa ASD Diagnosis Aid* by *Cognoa Medical Affairs* and the *EarliPoint System* by *EarliTec*. Both devices received FDA Class II approval. The *Cognoa ASD Diagnosis Aid* was cleared through the De Novo pathway in June 2021. It helps healthcare providers in diagnosing ASD in children aged 18 to 72 months who are perceived to be at risk for developmental delay. The *Cognoa ASD Diagnosis Aid* uses a machine learning algorithm to predict ASD based on clinical and video data. Caregivers upload videos of the patient and answer questions on demographic characteristics and the child's behavior via a mobile app. Trained analysts respond to questions about the patient's behavior based on the uploaded video data. The mobile app includes a portal for the healthcare provider, with questions about typical developmental behaviors for the child's age group. AI is used to analyze the combined input data and generate a value that is compared to predefined thresholds to decide the ASD diagnosis (positive, negative, or indeterminate if a score cannot be determined) [26].

The *EarliPoint System* was cleared via the 510(k) pathway in June 2022, with the *Cognoa ASD Diagnosis Aid* serving as its predicate device. The *EarliPoint System* aids healthcare providers in diagnosing ASD in children aged 16 to 30 months who are at risk for developmental delay. The *EarliPoint System* uses AI to diagnose ASD based on eye-tracking data. The data is collected by tracking the patient's visual reactions to social information shown in videos. The *EarliPoint* WebPortal is used for entering and accessing patient information and assessment results. The *EarliPoint System* delivers a positive or negative ASD diagnosis without an indeterminate option [74].

3.2.2. Diagnosis of multiple disorders based on conversational data

Two devices were designed for general diagnostic purposes using data collected through chatbots, namely *Ada Assess* by *AdaHealth* and *Limbic Access* by *Limbic Limited*. *Ada Assess* is *AdaHealth*'s CE-certified medical device. *Ada Assess* has been CE-certified and UKCA-recognized in December 2022 and *Limbic Access* has been UKCA-recognized in January 2023. *Ada Assess* is designed as a mobile application that guides users through questions about their demographic background and health conditions and symptoms, including but not limited to mental health issues. The application interacts with the user, prompting them to input information about their symptoms. AI is used to evaluate symptoms against known disorders to propose potential clinical diagnoses. The information generated by *Ada Assess* can be integrated into a provider report [75–77].

Limbic Access also uses conversational data collected through a mobile application. The process begins with the user providing demographic information and describing their symptoms. These inputs trigger more detailed questions to create a thorough symptom profile for each user. *Limbic Access* was primarily designed as an e-triage tool that aids the referral process for new patients, including automatically redirecting patients who are ineligible for psychiatric treatment to other

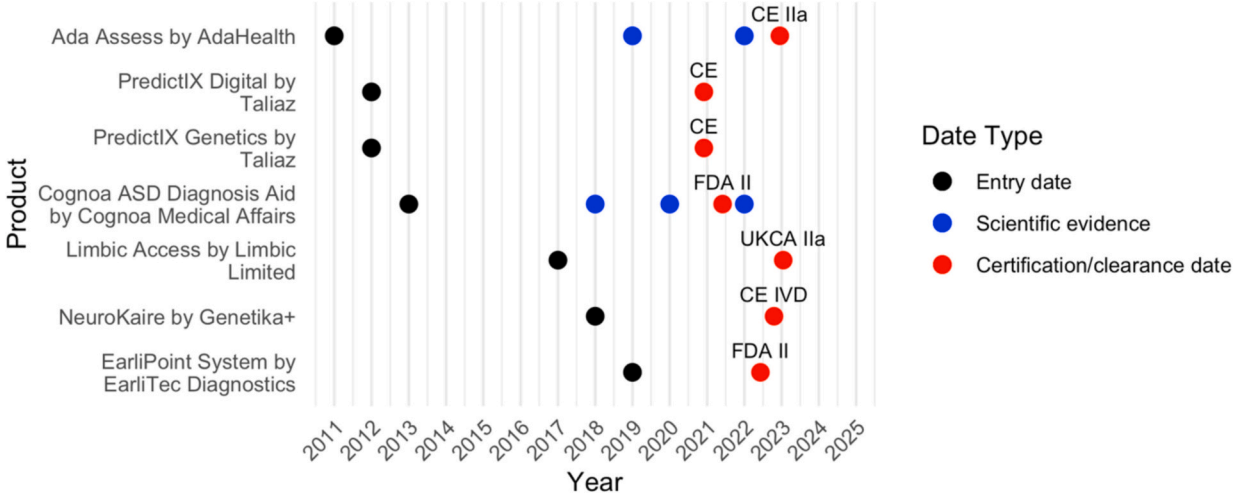


Fig. 2. The figure shows firm entry dates, dates at which scientific evidence was published, and certification/clearance dates of the reviewed AI-CDSS.

Table 1
The characteristics of regulated AI-CDSS included in the review.

Company	Product	Certification	Database confirmed	Risk class	Pathway	Disorder	Data sources	Data collection	Prediction aims	Output
AdaHealth	Ada Assess	CE	EUDAMED	IIa	–	Multiple	Conversational data	Mobile application (chatbot)	Diagnosis, treatment	Digital report displaying potential diagnoses and suggested treatments
Taliaz	PredictIX Digital	CE	No	NA	–	Multiple, focus on depression	Clinical history	Digital questionnaire	Treatment	Digital report ranking psychiatric drugs by predicted likelihood of patient's response and details associated side effects
Taliaz	PredictIX Genetics	CE	No	NA	–	Multiple, focus on depression	Clinical history, genetic	Digital questionnaire, swab sample	Treatment	Digital report ranking psychiatric drugs by predicted likelihood of patient's response and details associated side effects and recommended dosages
Cognoa Medical Affairs	Cognoa ASD Diagnosis Aid	FDA	FDA 513(f)(2) (De Novo)	II	De Novo	ASD	Clinical history, behavioral	Digital questionnaire, video upload portal	Diagnosis	Digital report indicating the ASD diagnosis ("Positive", "Negative", or "No Result" when a score cannot be determined)
Limbic Limited	Limbic Access	UKCA	PARD	IIa	–	Multiple	Conversational data	Mobile application (chatbot)	Diagnosis, treatment	Digital report indicating potential diagnoses and automatic tagging for high-risk patients
Genetika+	NeuroKaire	CE-IVD	No	NA	–	Depression	Clinical history, blood sample	Digital questionnaire, blood draw	Treatment	Digital report indicating the optimal antidepressant
EarliTec Diagnostics	EarliPoint System	FDA	FDA 510(k) Premarket Notification	II	510(k)	ASD	Clinical history, eye-tracking	Digital questionnaire, video recording	Diagnosis	Digital report indicating the ASD diagnosis ("Positive" or "Negative")

services and recommending prioritized assessment for those who may be at risk. AI is used to identify most likely diagnoses. The generated insight is presented to the treating clinician as a digital report to aid them with diagnostic and treatment decision-making [78,79]. *Limbic Access* has been integrated within the National Health Service (NHS) in the UK [80].

3.2.3. Medication selection based on clinical history and genetic data

NeuroKaire by *Genetika+* and both *PredictIX Digital* and *PredictIX Genetic* by *Taliaz* offer treatment recommendations regarding the optimal antidepressant medication. *NeuroKaire* by *Genetika+* obtained CE IVD certification in October 2020. *PredictIX Digital* and *PredictIX Genetic* received CE certification in December 2022. Remarkably, there was no available information on the risk class of *NeuroKaire* and the two *PredictIX* devices. None of the three tools could be found on EUDAMED and we received no additional information from vendors.

NeuroKaire uses clinical history data gathered through a digital questionnaire and a blood sample. The combined data is used to generate tailored treatment recommendations that predict which medications patients are most likely to respond to. The treating practitioner can access the report through a mobile application or a website [81,82]. *PredictIX Digital* and *PredictIX Genetics* use clinical history data collected through an online questionnaire to generate recommendations regarding the optimal antidepressant medication. *PredictIX Genetics* additionally uses genetic data. For *PredictIX Genetics*, a DNA sample is collected with a cheek swab. Both tools employ AI algorithms to generate a personalized patient report for practitioners that ranks psychiatric drugs according to the predicted likelihood of a patient's response. The report also provides insights into how patients might metabolize certain medications and predicts associated side effects and recommended dosages [83–85].

3.3. External validation of AI-CDSS for mental healthcare

The PRISMA flow diagram of the search and inclusion process may be accessed through the Online appendix (https://osf.io/vrfm7/?view_only=e027649e3a42428a8975f7385659b55c). In total, 149 publications were screened, of which five met the inclusion criteria. The overall mean CONSORT-AI score of the included trials was 52 %. Three articles provided evidence for the effectiveness of the *Cognoa ASD Diagnosis Aid* (19 of 51 (37 %) [12], 29 of 51 (57 %) [86], and 34 of 51 (67 %) [14]) and two for *Ada Assess* (29 of 51 (57 %) [76], and 21 of 51 (41 %) [77]).

An overview of the compliance of the included studies with each of the individual CONSORT-AI criteria ("Overview CONSORT-AI") can be found in the Online appendix (https://osf.io/vrfm7/?view_only=e027649e3a42428a8975f7385659b55c). Since none of the studies were randomized controlled trials, none of the CONSORT-AI criteria related to randomization were met (items 1a, 8a, 8b, 9, 10, and 11b). Additionally, none of the trials reported changes to trial outcomes after the trial began or conducted interim analyses (6b and 7b). None of the studies provided a rationale for the chosen sample size (7a) or reasons for the trial termination or cessation (14b). Important harms or unintended effects were not reported in any of the studies (19). The following criteria were only met in a single study: description of significant changes to methods after the trial started (3b), statement of inclusion and exclusion criteria at the level of the input data (4a-ii), specification of the AI algorithm version (5-i), description of how poor quality or unavailable input data were assessed and handled (5-iii), descriptions of methods and results for additional analyses (12b and 18), and results of any analysis of performance errors (19-i). Only one study provided the trial registration number and location of the trial protocol (23 and 24), and stated whether and how the AI intervention or its code can be accessed (25-i).

Three studies examined the external validity of different versions of the *Cognoa ASD Diagnosis Aid* [12,14,86]. The first study examined a

previous version of the *ASD Diagnosis Aid* using two data inputs, namely a digital caregiver questionnaire and a digital video analyst questionnaire [12]. The study included 230 children from three clinics in the US. Participants completed the screening prior to their clinic visit. Diagnoses by specialist clinicians based on the results of screening instruments and DSM-5 criteria were used as gold standard diagnoses. The *ASD Diagnosis Aid* showed a specificity of .62 and a sensitivity of .75 in detecting ASD. Only specificity surpassed that of other screening instruments [12,87].

The second study examined a previous version of the *ASD Diagnosis Aid* using three data inputs, namely a caregiver questionnaire, a video analyst questionnaire, and a healthcare provider questionnaire [86]. The study used data from 375 children collected in 2016 and 2017 at three tertiary care centers in the US. The children, aged 18 to 72 months, were referred for suspected ASD. The study compared the outputs of the *ASD Diagnosis Aid* and other screening instruments. Diagnoses by specialist clinicians were used as gold standard diagnoses. The *ASD Diagnosis Aid* showed a specificity of .75 and a sensitivity of .80 in detecting ASD [86,87].

Finally, building on previous work, the device inputs and algorithm were refined to arrive at the device version that was validated in a third prospective double-blinded comparator study [14]. The final was conducted to support the FDA Market Authorization of the most recent product version. The study included 425 children from 14 care centers aged 18 to 72 months with concerns for developmental delay but not yet diagnosed with ASD. The study was conducted between August 2019 and June 2020. The *ASD Diagnosis Aid* output was compared with diagnoses made by specialist clinicians and validated by one or more blinded reviewing specialists. The sensitivity and specificity for a determinate output were .98 and .79, respectively. The system's performance did not differ across childrens' sex, race or ethnicity, income, or education level. However, the rate of a determinate diagnosis was higher for children under three years, and specificity was higher for those over three years [14].

Two studies investigated the external validity of *Ada Assess* [76,77]. Both studies use the Ada self-help application without providing information on how the generated insight would be integrated into a practitioner report. Additionally, it remains unclear if the studies used the same Ada Health App version because the version number was only provided in one study [76]. The first study involved two psychotherapists, two psychology students, and two laypersons [77]. The six participants used Ada to diagnose 20 case vignettes describing various mental disorders. The app's diagnosis was then compared with the textbook diagnosis to assess interrater reliability. The results indicate a moderate agreement (kappa = .64) in diagnosing adults with mental disorders, but a lower agreement (kappa = .40) for children and adolescents. Diagnostic accuracy was comparatively higher among psychotherapists than in the other two groups. On average, the app required 34 questions and approximately seven minutes per case.

The second study included 49 adult patients from one outpatient psychotherapy clinic [76]. In this study, the self-assessment version of the app was evaluated. The app's performance was assessed by comparing the condition suggestions provided by the app with diagnoses made by therapists based on structured clinical interviews. Sensitivity was highest for affective disorders (.65 to .71) and lowest for somatoform and associated disorders (.22 to .29), while specificity was highest for somatoform and associated disorders (.85 to .95) and lowest for affective disorders (.50 to .67).

4. Discussion

4.1. Product characteristics

The included products fell into three main categories, namely the diagnosis of ASD among children based on clinical history, behavioral, and eye-tracking data; diagnosis of multiple disorders using conversational data; and medication selection based on clinical history and genetic data. The ASD diagnoses generated by the *Cognoa ASD Diagnosis Aid* and the *EarliTec System* rely on video and eye-tracking data. Video data provide rich insights into a child's behavior and interactions, which are crucial factors in the diagnosis of ASD [88–90]. Early diagnosis of ASD allows for timely intervention, which has been shown to improve cognitive and language abilities in children with ASD [91]. The average age for which the *ASD Diagnosis Aid* provided a positive result was 2.81 years, which is 1.5 years earlier than the average age children receive an ASD diagnosis [14]. Additionally, traditional diagnostic methods often require in-person assessments, which can be challenging for some families to access [92]. Including remotely recorded video data in the diagnostic process allows for more convenience for families and may reduce wait times during the diagnosis process [90]. Cost and reimbursement data, as well as appropriate practitioner training to guarantee safe and effective use are needed to determine if devices like the *ASD Diagnosis Aid* or *EarliTec* could be widely adopted [14].

Ada Assess and *Limbic Access* use data collected through conversational agents to generate diagnosis suggestions. Chatbots have been shown to facilitate interactions with individuals reluctant to seek mental health advice due to stigmatization [93]. While some studies found conversational agents to reliably predict mental disorders [94], others caution that they are not yet accurate enough to provide formal diagnoses [77,95,96]. Recommendations generated based on chatbot data can support decision-making, but the final diagnosis must come from a qualified mental health professional, emphasizing the role of conversational agents as complementary tools rather than replacements for human professionals [97].

NeuroKaire, *PredictIX Digital*, and *PredictIX Genetics* use clinical and genetic data to aid in selecting appropriate medication. The variability in primary care management of depression and the limited effectiveness of first-line antidepressant drug treatment underscore the need for more personalized approaches [98]. Research findings indicate that biomarkers and genetic variants might be useful in predicting treatment outcomes and guiding medication selection [99]. It is widely acknowledged that genomic factors contribute to the development of disorders and the response to medication together with environmental conditions, such as socioeconomic background, lifestyle and nutrition, and adverse life events [37,100,101]. Therefore, nongenetic information should be included in prediction models to improve the predictive value of genomic information [37]. Because the algorithms and data used in the products are proprietary, we are unaware of how genetic and other types of data are combined to make decisions. While all three devices received CE certification and meet the definition of medical devices based on descriptions from vendor websites, we were unable to find specific information about their risk class on EUDAMED or anywhere else online. Unfortunately, none of our emails were answered by the vendors.

4.2. Regulation of AI-CDSS for mental healthcare

The time lag between company founding and product certification has decreased over the years. This might be because newer companies take advantage of work predecessors have completed [43]. For instance, the *Cognoa ASD Diagnosis Aid* was regulated through the De Novo pathway and served as a predicate device for the *EarliPoint System*, which was regulated through the 510(k) pathway. The De Novo pathway generally takes longer than the 510(k) pathway due to the devices' novelty [45,46]. The mental health AI market is expected to grow at a compound annual growth rate of 38.87 % from 2023 to 2031 [102]. As more AI-CDSS are cleared, more predicate devices will become available, subsequently accelerating market growth, as shown for other AI-enabled medical devices [43]. The FDA takes a proactive approach to innovating regulatory requirements for AI-CDSS. In the newest white paper published in March 2024 the FDA summarizes several areas they want to focus on for the regulation of AI-enabled medical devices in the future [103]. These areas include fostering collaboration between regulatory bodies, healthcare professionals, patients, researchers, and

industry; advancing innovation, for example, through a final guidance on Predetermined Change Control Plans (PCCP) that outline approval strategies for adaptive AI-based medical devices; harmonizing current standards; and promoting research related to the evaluation and monitoring of AI performance [103].

Previous research has shown that the process of receiving CE certification has generally been quicker than obtaining FDA approval or clearance [42]. This may change with the strict enforcement of the MDR, which mandates reassessments of Notified Bodies and more rigorous clinical device testing [104,105]. The MDR will take full effect when the transition period ends in 2027 for Class IIb and III devices, and in 2028 for Class I and IIa devices [62,104]. In addition to the relatively strict requirements of the MDR, AI-enabled medical devices must adhere to the General Data Protection Regulation (GDPR) and, in the near future, the stipulations in the Proposal for a Regulatory Framework for Artificial Intelligence (EU AI Act) in order to be marketed in the EEA [106,107]. The cross-sectoral European regulatory approach has been criticized for its inefficiency in regulating AI-CDSS. Notably, Notified Bodies have expressed concerns about significant delays in regulation due to the additional procedures introduced by the EU AI Act [107]. Furthermore, the European regulatory approach has been criticized for its ambiguity regarding post-certification modifications to AI-CDSS, potentially resulting in unequal access to innovative and effective healthcare technology [108]. In summary, while regulatory adjustments aim to ensure the safety and efficacy of AI-CDSS, they also impose significant challenges on their path to market. Developers and vendors of AI-CDSS for mental healthcare will need to develop strategies to address the evolving challenges of new regulatory requirements.

4.3. External validation of AI-CDSS for mental healthcare

None of the identified papers adhered to all CONSORT-AI criteria, indicating that research quality and reporting detail are still subpar. A similar lack of scientific rigor has been reported for AI devices used in other healthcare domains [72,73]. Only one study mentioned the AI algorithm version that was used [76], making comparison of results from studies investigating the same device's efficacy challenging. Moreover, only one study identified inclusion and exclusion criteria at the level of the input data and reported how poor quality or unavailable input data was handled and assessed [14]. This lack of reporting casts doubt on the applicability of devices in real-world settings, where noisy and missing data are common [73]. Only one of the articles provided information on the trial registration number, where to find the trial protocol, and important method changes after the trial commenced [14]. The lack of reporting on these criteria poses a potential source of bias because changes could have been implemented after results were already obtained [72]. None of the articles provided information on the sample size rationale. An underpowered study might not result in stable and precise sensitivity and specificity estimates, creating uncertainty in the accuracy and robustness of the reported results [109]. Most of the studies did not provide any information on interim analyses, trial outcome changes, and additional exploratory analyses, such as subgroup or adjusted analyses, making it difficult for other researchers or practitioners to accurately interpret the results [110]. Finally, due to the lack of full reporting of potential harms or unexpected effects, it becomes challenging for practitioners to evaluate the risks and benefits of the tools [111]. Additionally, failing to present potential harms could lead to unrealistic expectations about how an AI-CDSS performs in real-world settings [71]. On the other hand, it should be emphasized positively that four of the five studies tested the product's performance under clinically relevant conditions and on a sample similar to the intended user population. In addition, the product's recommendations were compared to valid comparators, such as commonly used diagnostic tools [12,86] or diagnoses made by clinicians/textbook diagnoses [14,76,77].

The results of the studies indicate a wide variability in performance based on sensitivity and specificity values. The *Cognoa ASD Diagnosis Aid* exhibited a sensitivity ranging from .75 in the first study [12] to .98 in the third study [14], indicating a broad performance range in identifying true positive cases. Specificity varied from .62 [12] to .79 [14], indicating a broad range in correctly identifying true negatives. Notably, performance improved throughout the sequence of studies, which likely reflects the technological enhancements implemented in the different versions of the tool. For *Ada Assess*, sensitivity ranging from a high of .65 to .71 for affective disorders to a low of .22 to .29 for somatoform and associated disorders. Specificity was highest for somatoform and associated disorders (.85 to .95) and lowest for affective disorders (.50 to .67), suggesting a need for improvement in generalizing the diagnostic efficacy across disorders [76].

The inability to retrieve any scientific peer-reviewed papers for five of the seven products suggests a lack of research or potential publication bias against studies with negative results. The observed lack of published research findings is not restricted to AI devices in mental healthcare but affects AI medical devices generally [32,112]. One reason for a lack of research could be the high costs and resources associated with conducting rigorous clinical trials. Another reason for the scarcity of external validation studies may be a lack of common standards for the clinical evaluation of high-risk AI applications and transparency of their evidence and performance [106]. Current initiatives such as the Coordinating Research and Evidence for Medical Devices (CORE-MD) initiative are actively developing standards for assessing evidence pertaining to high-risk medical devices within the EEA. These efforts may stimulate convergence among stakeholders and the adherence to reporting standards for clinical trials addressing the external validation of AI-CDSS [106,113].

4.4. Limitations

The aim of this review was to provide detailed insight into regulated AI-CDSS for mental healthcare, though the relatively strict inclusion criteria limit the generalizability of the obtained insight to a very specific scope of products. Specifically, we focused solely on decision-support tools, excluding AI-assisted interventions [114]. Additionally, we had to establish a threshold for what technology should be regarded as AI-based. The terms we selected might have been too narrow to identify all AI-CDSS for mental healthcare. Finally, similar to other reviews, we only included AI-CDSS that received FDA clearance or CE/UKCA certification [25,28], thus disregarding unregulated technologies or technologies regulated elsewhere.

The insight only provides a preliminary review of the landscape of AI-CDSS for mental healthcare available up to April 2024. All products were cleared or certified within the past four years, indicating that the market is still emerging and new developments may not be considered in the review. Given the market's infancy, future studies are crucial to keep pace with the potentially rapid development of AI-CDSS for mental healthcare.

In our search for scientific evidence of identified AI-CDSS, we only considered publications available in English, consistent with previous approaches [28]. All scientific evidence provided by firms on their respective websites was published in English. This suggests that most of the evidence addressing the external validation of AI-CDSS regulated in the US or the EEA will be published in English. However, we cannot rule out the possibility that we may have missed some external validation data due to this language limitation. Additionally, while we used the CONSORT-AI checklist to evaluate the quality and reproducibility of external validation studies, we acknowledge that other reporting guidelines, such as the Minimum Information for Medical AI Reporting (MINIMAR) [115], the Developmental and Exploratory Clinical Investigations of Decision support systems driven by Artificial Intelligence (DECIDE-AI) [116], or the Standard Protocol Items: Recommendations for Interventional Trials-Artificial Intelligence (SPIRIT-AI) [117] may also be appropriate. Future studies may consider using alternative or multiple reporting guidelines to capture a broader spectrum of reporting

standards. Finally, the provided insight is limited by restrictions in the availability of data. Data collection posed a considerable challenge, as many vendors did not respond or chose not to share certain requested information. Using data collected from vendor websites may not accurately reflect the devices' actual capabilities, limitations, or clinical applications.

5. Conclusions

Out of a broad initial pool (n = 84), only seven products met the inclusion criteria for AI-CDSS and received FDA clearance or CE/UKCA certification. Developers and vendors of AI-CDSS for mental healthcare need to familiarize themselves with up-to-date regulatory requirements and guidelines that determine how tools should be designed. The FDA, Health Canada, and the UK's Medicines and Healthcare products Regulatory Agency (MHRA) have jointly identified ten guiding principles that can inform the development of Good Machine Learning Practice (GMLP). These include leveraging multi-disciplinary expertise throughout the total product lifecycle, using representative samples in clinical studies, employing independent training and test data sets, and placing focus on the performance of the human-AI team, just to name a few [118]. The regulatory landscape is changing quickly and tool developers and firms must ensure they stay up-to-date with the latest national and international developments, such as additional regulatory requirements to adhere to as part of the EU AI Act. Achieving regulatory approval does not only ensure the safety and efficacy of AI-CDSS but also increases the chance of widespread adoption, as can be seen with *Limbic Access* that has been adopted by the NHS in the UK [80].

The scarcity of available external validation studies raises doubts about the clinical usefulness of the tools when applied in different clinical settings. In addition, the few studies we found did not meet all reporting requirements for clinical studies described in the CONSORT-AI. Tool developers and vendors need to ensure they consider when and how external validation of their tool will be performed before submitting it for regulatory approval. When conducting clinical studies, researchers need to adhere to good practice guidelines, such as those outlined in the CONSORT-AI, to guarantee the external validity of the generated recommendations. We expect additional useful guidelines and recommendations to be released soon as part of dedicated initiatives, such as CORE-MD [106,113].

The findings of this review underscore the challenges that manufacturers encounter when striving to obtain regulatory approval for their devices. At present, numerous opinion articles and recommendations for modifying the regulation of AI-enabled medical devices exist [107,108,119,120]. Essentially, regulatory agencies and policymakers need to ensure meticulous oversight while also fostering an environment for innovation. A positive step would be aligning regulatory and scientific standards as suggested by the CORE-MD initiative [106]. Adhering to rigorous reporting guidelines like CONSORT-AI as part of the regulatory criteria may improve the regulatory approval process by providing a structured framework that ensures comprehensive evaluation. This approach would help guarantee both the safety and efficacy of AI-CDSS while addressing issues often overlooked in the regulatory approval process, such as external validation and reporting on the usefulness of the solution in real-life settings. Additionally, improved coordination among national and international regulatory bodies and the establishment of global regulatory convergence would help to minimize the risk of blocks on innovation or divergent quality standards [106]. Regularly updated registers of AI-enabled medical devices, such as the *Health AI Register* for radiology, could be utilized for mental healthcare to ensure transparency about the characteristics, regulatory status, and availability of scientific evidence for AI-CDSS for mental healthcare [121].

CRediT authorship contribution statement

Anne-Kathrin Kleine: Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Eesha Kokje:** Writing – review & editing, Validation, Methodology, Investigation, Data curation. **Pia Hummelsberger:** Validation, Investigation, Data curation. **Eva Lermer:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Insa Schaffernak:** Writing – review & editing, Validation, Investigation, Data curation. **Susanne Gaube:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT (GPT4) in order to check the logic of the Introduction and Discussion sections and to revise some of the paragraphs. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research was funded by the VW Foundation (“Human-AI-Interaction in Healthcare: Identifying Factors Contributing to Clinical Utility”; Grant #: 98525). The sponsor was not involved in the choice of the study design, data collection, data analysis, and interpretation of results.

References

[1] Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016;3:171–8. https://doi.org/10.1016/S2215-0366(15)00505-2.
[2] Henderson C, Evans-Lacko S, Thornicroft G. Mental illness stigma, help seeking, and public health programs. Am J Public Health 2013;103:777–80. https://doi.org/10.2105/AJPH.2012.301056.
[3] Sowers KM, Dulmus CN, Linn BK. Mental illness: worldwide. In: Encycl soc work; 2019.
[4] Thornicroft G. Most people with mental illness are not treated. Lancet Lond Engl 2007;370:807–8.
[5] Wang PS, Aguilar-Gaxiola S, Alonso MD, Angermeyer MC, Borges G, Bromet EJ, et al. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. Lancet Lond Engl 2007;370:841–50.
[6] Graham SA, Depp CA, Lee EE, Nebeker C, Tu XM, Kim H-C, et al. Artificial intelligence for mental health and mental illnesses: an overview. Curr Psychiatry Rep 2019;21.
[7] Koutsouleris N, Hauser TU, Skvortsova V, Choudhury MD. From promise to practice: towards the realisation of AI-informed mental health care. Lancet Digit Health 2022;4:e829–40. https://doi.org/10.1016/S2589-7500(22)00153-4.
[8] Lee EE, Torous J, De Choudhury M, Depp CA, Graham SA, Kim H-C, et al. Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. Biol Psychiatry Cogn Neurosci Neuroimaging 2021;6:856–64. https://doi.org/10.1016/j.bpsc.2021.02.001.
[9] Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: a systematic review. Laryngoscope Investig Otolaryngol 2020;5:96–116. https://doi.org/10.1002/lio2.354.
[10] Washington P, Park N, Srivastava P, Voss C, Kline A, Varma M, et al. Data-driven diagnostics and the potential of mobile artificial intelligence for digital therapeutic phenotyping in computational psychiatry. Biol Psychiatry Cogn Neurosci Neuroimaging 2020;5:759–69. https://doi.org/10.1016/j.bpsc.2019.11.015.
[11] Chhabra C, Chaudhary S, Malik A, Bhushan B. Mental disorder detection using machine learning. In: Big Data anal. green comput. CRC Press; 2021.

[12] Kanne SM, Carpenter LA, Warren Z. Screening in toddlers and preschoolers at risk for autism spectrum disorder: evaluating a novel mobile-health screening tool. Autism Res 2018;11:1038–49. https://doi.org/10.1002/aur.1959.

[13] Lin Y, Liyanage BN, Sun Y, Lu T, Zhu Z, Liao Y, et al. A deep learning-based model for detecting depression in senior population. Front. Psychiatry 2022:13. https://doi.org/10.3389/fpsyt.2022.1016676.

[14] Megerian JT, Dey S, Melmed RD, Coury DL, Lerner M, Nicholls CJ, et al. Evaluation of an artificial intelligence-based medical device for diagnosis of autism spectrum disorder. Npj Digit Med 2022;5:1–11. https://doi.org/10.1038/s41746-022-00598-6.

[15] Aafjes-van Doorn K, Kamsteeg C, Bate J, Aafjes M. A scoping review of machine learning in psychotherapy research. Psychother Res 2021;31:92–116.

[16] Aleem S, Huda N ul, Amin R, Khalid S, Alshamrani SS, Alshehri A. Machine learning algorithms for depression: diagnosis, insights, and research directions. Electronicsweek 2022;11:1111.

[17] Chekroud AM, Bondar J, Delgadillo J, Doherty G, Wasil A, Fokkema M, et al. The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry Off J World Psychiatr Assoc WPA 2021;20:154–70.

[18] Cho G, Yim J, Choi Y, Ko J, Lee S-H. Review of machine learning algorithms for diagnosing mental illness. Psychiatry Investig 2019;16:262–9.

[19] Golden G, Popescu C, Israel S, Perlman K, Armstrong C, Fratila R, et al. Applying artificial intelligence to clinical decision support in mental health: what have we learned? Health Policy Technol 2024:100844. https://doi.org/10.1016/j.hlpt.2024.100844.

[20] Kleine A-K, Lermer E, Cecil J, Heinrich A, Gaube S. Advancing mental health care with AI-enabled precision psychiatry tools: a patent review. Comput Hum Behav Rep 2023;12:100322. https://doi.org/10.1016/j.chbr.2023.100322.

[21] Liu G-D, Li Y-C, Zhang W, Zhang L. A brief review of artificial intelligence applications and algorithms for psychiatric disorders. Engineering 2020;6:462–7. https://doi.org/10.1016/j.eng.2019.06.008.

[22] Popescu C, Golden G, Benrimoh D, Tanguay-Sela M, Slowey D, Lundrigan E, et al. Evaluating the clinical feasibility of an artificial intelligence–powered, web-based clinical decision support system for the treatment of depression in adults: longitudinal feasibility study. JMIR Form Res 2021;5:e31862. https://doi.org/10.2196/31862.

[23] Lagioia F, Contissa G. The strange case of Dr. Watson: liability implications of AI evidence-based decision support systems in health care. Eur. J Leg Stud 2020;12:245–90.

[24] Baldwin H, Loebel-Davidsohn L, Oliver D, Salazar de Pablo G, Stahl D, Riper H, et al. Real-world implementation of precision psychiatry: a systematic review of barriers and facilitators. Brain Sci 2022;12:934.

[25] Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med 2020;3:1–8.

[26] U.S. Food and Drug Administration. Cognoa ASD diagnosis aid [de novo classification]. 2021.

[27] U.S. Food and Drug Administration. EarliPoint [510(k) premarket notification]. 2023.

[28] Leeuwen KG van, Schalekamp S, Rutten MJCM, Ginneken B van, Rooij M de. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur Radiol 2021;31:3797–804. doi:https://doi.org/10.1007/s00330-021-07892-z.

[29] Andrade C. Internal, external, and ecological validity in research design, conduct, and evaluation. Indian J Psychol Med 2018;40:498–9. https://doi.org/10.4103/IJPSYM.IJPSYM_334_18.

[30] Ebrahimian S, Kalra MK, Agarwal S, Bizzo BC, Elkholy M, Wald C, et al. FDA-regulated AI algorithms: trends, strengths, and gaps of validation studies. Acad Radiol 2022;29:559–66.

[31] Harada T, Miyagami T, Kunitomo K, Shimizu T. Clinical decision support systems for diagnosis in primary care: a scoping review. Int J Environ Res Public Health 2021;18:8435.

[32] Yearley AG, Goedmakers CMG, Panahi A, Doucette J, Rana A, Ranganathan K, et al. FDA-approved machine learning algorithms in neuroradiology: A systematic review of the current evidence for approval. Artif Intell Med 2023;143:102607.

[33] Lesko LJ, Zineh I, Huang S-M. What is clinical utility and why should we care? Clin Pharmacol Ther 2010;88:729–33. https://doi.org/10.1038/clpt.2010.229.

[34] Andrew J, Rudra M, Eunice J, Belfin RV. Artificial intelligence in adolescents mental health disorder diagnosis, prognosis, and treatment. Front. Public Health 2023:11. https://doi.org/10.3389/fpubh.2023.1110088.

[35] Salazar de Pablo G, Studerus E, Vaquerizo-Serrano J, Irving J, Catalan A, Oliver D, et al. Implementing precision psychiatry: a systematic review of individualized prediction models for clinical practice. Schizophr Bull 2021;47:284–97. https://doi.org/10.1093/schbul/sbaa120.

[36] Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 2018;3:223–30. https://doi.org/10.1016/j.bpsc.2017.11.007.

[37] Koch E, Pardiñas AF, O'Connell KS, Selvaggi P, Camacho Collados J, Babic A, et al. How real-world data can facilitate the development of precision medicine treatment in psychiatry. Biol Psychiatry 2024. https://doi.org/10.1016/j.biopsych.2024.01.001.

[38] Tornero-Costa R, Martinez-Millana A, Azzopardi-Muscat N, Lazeri L, Traver V, Novillo-Ortiz D. Methodological and quality flaws in the use of artificial intelligence in mental health research: systematic review. JMIR Ment Health 2023;10:e42045. https://doi.org/10.2196/42045.

[39] Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med 2019;49:1426–48.

[40] European Commission. Medical devices. 2022.

[41] U.S. Food and Drug Administration. Overview of regulatory requirements: medical devices - transcript. FDA; 2018.

[42] Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and europe (2015–20): a comparative analysis. Lancet Digit Health 2021;3:e195–203. https://doi.org/10.1016/S2589-7500(20)30292-2.

[43] Muehlematter UJ, Bluethgen C, Vokinger KN. FDA-cleared artificial intelligence and machine learning-based medical devices and their 510(k) predicate networks. Lancet Digit Health 2023;5:e618–26. https://doi.org/10.1016/S2589-7500(23)00126-7.

[44] Zaki M. A review of international medical device regulations: contact lenses and lens care solutions. Cont Lens Anterior Eye 2019;42:136–46.

[45] Sherkow JS, Aboy M. The FDA de novo medical device pathway, patents and anticompetition. Nat Biotechnol 2020;38:1028–9.

[46] U.S. Food and Drug Administration. De novo classification request. FDA; 2023.

[47] U.S. Food and Drug Administration. Premarket notification 510(k). FDA; 2023.

[48] Lee JT, Moffett AT, Maliha G, Faraji Z, Kanter GP, Weissman GE. Analysis of devices authorized by the FDA for clinical decision support in critical care. JAMA Intern Med 2023;183:1399–401. https://doi.org/10.1001/jamainternmed.2023.5002.

[49] European Parliament. Regulation (EU) 2017/745 of the European Parliament and of the Council of on medical devices. 2023.

[50] European Parliament. Regulation (EU) 2017/746 of the European Parliament and of the Council on in vitro diagnostic medical devices. 2023.

[51] Han JED, Ibrahim H, Aiyegbusi OL, Liu X, Marston E, Denniston AK, et al. Opportunities and risks of UK medical device reform. Ther Innov Regul Sci 2022;56:596–606.

[52] Schröttner J, Baumgartner C. The notified body: the conformity assessment body for medical devices in Europe. In: Baumgartner C, Harer J, Schröttner J, editors. Med. devices vitro diagn. requir. Eur., Cham: Springer International Publishing; 2023. p. 39–61. https://doi.org/10.1007/978-3-031-22091-3_1.

[53] Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019;17:195. https://doi.org/10.1186/s12916-019-1426-2.

[54] Siontis GCM, Sweda R, Noseworthy PA, Friedman PA, Siontis KC, Patel CJ. Development and validation pathways of artificial intelligence tools evaluated in randomised clinical trials. BMJ Health Care Inform 2021;28:e100466. https://doi.org/10.1136/bmjhci-2021-100466.

[55] Park SH, Han K, Jang HY, Park JE, Lee J-G, Kim DW, et al. Methods for clinical evaluation of artificial intelligence algorithms for medical diagnosis. Radiology 2023;306:20–31. https://doi.org/10.1148/radiol.220182.

[56] Freeman T. Using performance indicators to improve health care quality in the public sector: a review of the literature. Health Serv Manage Res 2002;15:126–37. https://doi.org/10.1258/0951484021912897.

[57] Šimundić A-M. Measures of diagnostic accuracy: basic definitions. EJIFCC 2009;19:203–11.

[58] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med 2009;151:W-65. doi:https://doi.org/10.7326/0003-4819-151-4-200908180-00136.

[59] Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. Artif Intell Healthc 2020:25–60. https://doi.org/10.1016/B978-0-12-818438-7.00002-2.

[60] American Psychiatric Association. Diagnostic and statistical manual of mental disorders, text revision (DSM-5-TR). Text revis. 2022.

[61] U.S. Food and Drug Administration. Medical device databases. FDA; 2023.

[62] European Commission. EUDAMED - European database on medical devices. 2023.

[63] Medicines and Healthcare Products Regulatory Agency. Public Access Registration Database (PARD). 2023.

[64] Woo S, Jang P, Kim Y. Effects of intellectual property rights and patented knowledge in innovation and industry value added: a multinational empirical analysis of different industries. Technovation 2015;43–44:49–63.

[65] U.S. Patent and Trademark Office. Ownership/assignability of patents and applications. 2023.

[66] U.S. Food and Drug Administration. Artificial intelligence and machine learning (AI/ML)-enabled medical devices. 2023.

[67] Certified health apps library – healthskouts digital health advisory company. https://www.healthskouts.com/certified-apps/; 2023.

[68] FDA-approved A.I.-based algorithms. Med Futur 2023. https://medicalfuturist.com/fda-approved-ai-based-algorithms//.

[69] Gómez-González E, Gómez Gutiérrez E. Artificial intelligence in medicine and healthcare: applications, availability and societal impact. Publications Office of the European Union; 2020. https://doi.org/10.2760/047666.

[70] Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, Ashrafian H, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Lancet Digit Health 2020;2:e537–48. https://doi.org/10.1016/S2589-7500(20)30218-1.

[71] Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10:28–55. https://doi.org/10.1016/j.ijsu.2011.10.001.

[72] Pattathil N, Zhao JZL, Sam-Oyerinde O, Felfeli T. Adherence of randomised controlled trials using artificial intelligence in ophthalmology to CONSORT-AI

guidelines: a systematic review and critical appraisal. BMJ Health Care Inform 2023;30:e100757. https://doi.org/10.1136/bmjhci-2023-100757.
[73] Shahzad R, Ayub B, Siddiqui MAR. Quality of reporting of randomised controlled trials of artificial intelligence in healthcare: a systematic review. BMJ Open 2022; 12:e061519.
[74] U.S. Food and Drug Administration. EarliPoint system [510(k) premarket notification]. 2022.
[75] Ada Assess. AdaHealth. https://ada.com/help/what-is-ada-assess/. [Accessed 11 April 2024].
[76] Hennemann S, Kuhn S, Witthöft M, Jungmann SM. Diagnostic performance of an app-based symptom checker in mental disorders: comparative study in psychotherapy outpatients. JMIR Ment Health 2022;9:e32832. https://doi.org/10.2196/32832.
[77] Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a chatbot (ada) in the diagnosis of mental disorders: comparative case study with lay and expert users. JMIR Form Res 2019;3:e13863. https://doi.org/10.2196/13863.
[78] Limbic Access. Limbic Ltd. https://limbic.ai/access; 2023.
[79] Rollwage M, Habicht J, Juchems K, Carrington B, Stylianou M, Hauser T, et al. Using conversational AI to facilitate mental health assessmentand improve clinical efficiencies in psychotherapy services inlarge real-world dataset. JMIR AI 2023;2:e44358. https://doi.org/10.2196/44358.
[80] Habicht J, Viswanathan S, Carrington B, Hauser TU, Harper R, Rollwage M. Closing the accessibility gap to mental health treatment with a personalized self-referral chatbot. Nat Med 2024;30:595–602. https://doi.org/10.1038/s41591-023-02766-x.
[81] Albeldas C, Cohen N, Natan S, Askari N, Laifenfeld D, Solal TC. Genetika+: precision medicine solutions for mental health. Pharmacogenomics 2022;23: 571–4.
[82] NeuroKaire. Genetika+. https://www.neurokaire.com/physicians-patients. [Accessed 11 April 2024].
[83] PredictIX Digital. Taliaz. https://predictix.ai/predictix-digital/; 2023.
[84] PredictIX Genetics. Taliaz. https://predictix.ai/predictix-genetics/; 2023.
[85] Taliaz D, Souery D. A new characterization of mental health disorders using digital behavioral data: evidence from major depressive disorder. J Clin Med 2021;10:3109.
[86] Abbas H, Garberson F, Liu-Mayo S, Glover E, Wall DP. Multi-modular ai approach to streamline autism diagnosis in young children. Sci Rep 2020;10:1–8.
[87] Wall DP, Liu-Mayo S, Salomon C, Shannon J, Taraman S. Optimizing a de novo artificial intelligence-based medical device under a predetermined change control plan: improved ability to detect or rule out pediatric autism. Intell-Based Med 2023;8:100102. https://doi.org/10.1016/j.ibmed.2023.100102.
[88] Nabil MA, Akram A, Fathalla KM. Applying machine learning on home videos for remote autism diagnosis: further study and analysis. Health Informatics J 2021; 27:1460458221991882. https://doi.org/10.1177/1460458221991882.
[89] Tariq Q, Daniels J, Schwartz J, Washington P, Kalantarian H, Wall DP, et al. PLoS Med 2018:15.
[90] Wu C, Helvaci H, Cheung S, Chuah C-N, Ozoff S, Young G. Machine learning based autism spectrum disorder detection from videos. In: Heal. int. conf. e-health netw. appl. serv; 2021. p. 1–6.
[91] Daniels A, Mandell DS. Explaining differences in age at autism spectrum disorder diagnosis: a critical review. Autism Int J Res Pract 2013;18:583–97.
[92] Okoye C, Obialo-Ibeawuchi CM, Obajeun OA, Sarwar S, Tawfik C, Waleed MS, et al. Early diagnosis of autism spectrum disorder: a review and analysis of the risks and benefits. Curēus 2023;15:e43226. https://doi.org/10.7759/cureus.43226.
[93] Abd-Alrazaq A, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. J Med Internet Res 2021;23.
[94] Vaidyam A, Linggonegoro DW, Torous J. Changes to the psychiatric chatbot landscape: a systematic review of conversational agents in serious mental illness: Changements du paysage psychiatrique des chatbots: Une revue systématique des agents conversationnels dans la maladie mentale sérieuse. Can J Psychiatry 2020; 66.
[95] Bhattacharyya R, Chakraborty K, Neogi R. ChatGPT and its application in the field of mental health. J SAARC Psychiatr Fed 2023;1:6.
[96] Farhat F. ChatGPT as a complementary mental health resource: a boon or a bane. Ann Biomed Eng 2023;In press.
[97] Balcombe L. AI chatbots in digital mental health. Informatics 2023;10:82.
[98] Scala JJ, Ganz AB, Snyder MP. Precision medicine approaches to mental health care. Physiology 2023;38:82–98.
[99] Kambeitz-Ilankovic L, Koutsouleris N, Upthegrove R. The potential of precision psychiatry: what is in reach? Br J Psychiatry 2022;220:175–8. https://doi.org/10.1192/bjp.2022.23.
[100] Franks PW, Melén E, Friedman M, Sundström J, Kockum I, Klareskog L, et al. Technological readiness and implementation of genomic-driven precision medicine for complex diseases. J Intern Med 2021;290:602–20. https://doi.org/10.1111/joim.13330.
[101] Johansson Å, Andreassen OA, Brunak S, Franks PW, Hedman H, Loos RJF, et al. Precision medicine in complex diseases—molecular subgrouping for improved prediction and treatment stratification. J Intern Med 2023;294:378–96. https://doi.org/10.1111/joim.13640.
[102] insightaceanalytic. AI in mental health market share, size, growth and forecast to 2031 2023. https://www.insightaceanalytic.com/report/global-ai-in-mental-health-market-/1272.
[103] U.S. Food and Drug Administration. Artificial intelligence and medical products: how CBER, CDER, CDRH, and OCP are working together. 2024.
[104] Bretthauer M, Gerke S, Hassan C, Ahmad OF, Mori Y. The new european medical device regulation: balancing innovation and patient safety. Ann Intern Med 2023; 176:844–8.
[105] Van Laere S, Muylle KM, Cornu P. Clinical decision support and new regulatory frameworks for medical devices: are we ready for it? - a viewpoint paper. Int J Health Policy Manag 2021;11:3159–63. https://doi.org/10.34172/ijhpm.2021.144.
[106] Fraser AG, Biasin E, Bijnens B, Bruining N, Caiani EG, Cobbaert K, et al. Artificial intelligence in medical device software and high-risk medical devices – a review of definitions, expert recommendations and regulatory initiatives. Expert Rev Med Devices 2023;20:467–91. https://doi.org/10.1080/17434440.2023.2184685.
[107] Gilbert S, Anderson S, Daumer M, Li P, Melvin T, Williams R. Learning from experience and finding the right balance in the governance of artificial intelligence and digital health technologies. J Med Internet Res 2023;25:e43682. https://doi.org/10.2196/43682.
[108] Gilbert S, Fenech M, Hirsch M, Upadhyay S, Biasiucci A, Starlinger J. Algorithm change protocols in the regulation of adaptive machine learning-based medical devices. J Med Internet Res 2021;23:e30545. https://doi.org/10.2196/30545.
[109] Bujang MA, Adnan TH. Requirements for minimum sample size for sensitivity and specificity analysis. J Clin Diagn Res JCDR 2016;10:YE01–6. doi:https://doi.org/10.7860/JCDR/2016/18129.8744.
[110] Riveros C, Dechartres A, Perrodeau E, Haneef R, Boutron I, Ravaud P. Timing and completeness of trial results posted at ClinicalTrials.gov and published in journals. PLoS Med 2013;10:e1001566.
[111] Ioannidis JPA, Evans SJW, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004;141:781–8. https://doi.org/10.7326/0003-4819-141-10-200411160-00009.
[112] Vokinger KN, Feuerriegel S, Kesselheim AS. Continual learning in medical devices: FDA's action plan and beyond. Lancet Digit Health 2021;3:e337–8.
[113] Fraser AG, Nelissen RGHH, Kjærsgaard-Andersen P, Szymański P, Melvin T, Piscoi P. Improved clinical investigation and evaluation of high-risk medical devices: the rationale and objectives of CORE–MD (Coordinating Research and Evidence for Medical Devices). EFORT Open Rev 2021;6:839–49. https://doi.org/10.1302/2058-5241.6.210081.
[114] Jaliaawala MS, Khan RA. Can autism be catered with artificial intelligence-assisted intervention technology? A comprehensive survey. Artif Intell Rev 2020; 53:1039–69. https://doi.org/10.1007/s10462-019-09686-8.
[115] Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. J Am Med Inform Assoc 2020;27:2011–5. https://doi.org/10.1093/jamia/ocaa088.
[116] Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ 2022;377:e070904. https://doi.org/10.1136/bmj-2022-070904.
[117] Rivera SC, Liu X, Chan A-W, Denniston AK, Calvert MJ, Ashrafian H, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health 2020;2:e549–60. https://doi.org/10.1016/S2589-7500(20)30219-3.
[118] U.S. Food and Drug Administration. Good machine learning practice for medical device development guiding principles. 2021.
[119] Gilbert S, Pimenta A, Stratton-Powell A, Welzel C, Melvin T. Continuous improvement of digital health applications linked to real-world performance monitoring: safe moving targets? Mayo Clin Proc Digit Health 2023;1:276–87. https://doi.org/10.1016/j.mcpdig.2023.05.010.
[120] Li P, Williams R, Gilbert S, Anderson S. Regulating artificial intelligence and machine learning-enabled medical devices in Europe and the United Kingdom Symposium: regulatory futures and medical devices. Law Technol Hum 2023;5: 94–113.
[121] Health AI. Register. https://healthairegister.com/. [Accessed 23 April 2024].