Factors Influencing the Effectiveness of AI-Assisted Decision-Making in Medicine: A Scoping Review

Nicholas J Jackson, BS[1], Katherine E. Brown, PhD[2], Rachael Miller, PharmD[1], Matthew Murrow, PhD[1] Michael R Cauley, PhD, DMin[2], Benjamin Collins MD, MA, MS[2,3], Laurie L Novak, PhD, MHSA[2], Natalie C Benda, PhD[4], Jessica S Ancker, PhD, MPH[2]

1. Department of Biomedical Informatics, Vanderbilt University, Nashville, USA
2. Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, USA
3. Department of Medicine, Vanderbilt University Medical Center, Nashville, USA
4. School of Nursing, Columbia University, New York City, USA

Contact author and address:

Nicholas J Jackson

Email: nicholas.jackson@vanderbilt.edu

Address: 2525 West End Ave, Nashville, TN, 37203, 14118E

Keywords: human-computer interaction, artificial intelligence, review, medical decision-making,

Word Count (excluding captions & headings): 3,951

Abstract Word Count: 249

ABSTRACT

Objective: Research on artificial intelligence-based clinical decision-support (AI-CDS) systems has returned mixed results. Sometimes providing AI-CDS to a clinician will improve decision-making performance, sometimes it will not, and it is not always clear why. This scoping review seeks to clarify existing evidence by identifying clinician-level and technology design factors that impact the effectiveness of AI-assisted decision-making in medicine.

Materials and Methods: We searched MEDLINE, Web of Science, and Embase for peer-reviewed papers that studied factors impacting the effectiveness of AI-CDS. We identified the factors studied and their impact on three outcomes: clinicians' attitudes toward AI, their decisions (e.g., acceptance rate of AI recommendations), and their performance when utilizing AI-CDS.

Results: We retrieved 5,850 articles and included 45. Four clinician-level and technology design factors were commonly studied. Expert clinicians may benefit less from AI-CDS than non-experts, with some mixed results. Explainable AI increased clinicians' trust, but could also increase trust in incorrect AI recommendations, potentially harming human-AI collaborative performance. Clinicians' baseline attitudes toward AI predict their acceptance rates of AI recommendations. Of the three outcomes of interest, human-AI collaborative performance was most commonly assessed.

Discussion and Conclusion: Few factors have been studied for their impact on the effectiveness of AI-CDS. Due to conflicting outcomes between studies, we recommend future work should leverage the concept of 'appropriate trust' to facilitate more robust research on AI-CDS, aiming

not to increase overall trust in or acceptance of AI but to ensure that clinicians accept AI recommendations only when trust in AI is warranted.

INTRODUCTION

Given the increasing effectiveness of artificial intelligence (AI) for diagnostic and clinical reasoning tasks[1–5], many studies have evaluated the potential for these systems to augment clinical decision-making via the use of AI-based clinical decision support (AI-CDS), with promising results[6–9]. However, recent studies have found that AI-CDS may not improve clinicians' decision-making performance, even when the AI-CDS performs well (e.g., has high diagnostic accuracy). For example, Goh et al.[10] found that providing physicians with access to a large language model (LLM) chatbot assistant did not improve physicians' performance on a diagnostic reasoning task, even though the chatbot performed better on this task than physicians did. Similarly, Yu et al.[11] found that providing AI-CDS for chest X-ray diagnosis did not improve on clinicians' diagnostic performance. Similar results have been observed in real-world AI-CDS deployment: when AI-CDS recommendations conflict with clinicians' initial judgment, clinicians reject these recommendations, leading to no change in patient outcomes[12,13]. These studies demonstrate the lack of clarity in this field: sometimes providing AI-CDS to a clinician will improve their decision-making performance, sometimes it will not, and it is not yet clear why. The benefits of AI-CDS for patient outcomes cannot be realized without a more comprehensive understanding of how the design features of the AI-CDS, characteristics of the clinician, and aspects of the clinical context influence the success or failure of AI-assisted decision-making.

Although insights are available from the existing research on non-AI (i.e., traditional) CDS systems,[14–17] they are not sufficient to resolve the questions about AI-CDS. This is largely because the complexity and opacity of AI-CDS are significantly higher than those of traditional CDS and, as a result, users may be less willing to trust or accept the recommendations of AI-CDS[14,18].

Additionally, an emerging body of literature attempts to improve AI-assisted decision-making by modifying how users interact with AI systems through new computational and design approaches[23–25]. Promising approaches include providing explanations for AI decisions[26] (explainable AI) or disclosing AI uncertainty to the decision-maker[27,28] (uncertainty quantification). Despite the promise of these approaches, relatively little of this research has studied medical decision-making, which carries unique ethical, legal, and social implications[15,16,23–25,29]. As a result, it is not yet clear how these findings apply to medical AI.

Given the rapid growth of medical AI and the mixed results of AI-CDS in both experimental and real-world settings, we conducted a scoping literature review to identify technology design and clinician-level factors that influence the effectiveness of human-AI collaborative decision-making in medicine. We focused on three important outcomes relevant to effectiveness of decision support: clinicians' attitudes toward AI, rate of acceptance of AI-CDS recommendations, and human-AI collaborative performance on a task of interest (e.g., differential diagnosis).

METHODS

Review Methodology

We conducted our review in accordance with the PRISMA-ScR guidelines[30] (Supplementary Materials, Table 1). Informed by consultation with an experienced medical librarian, we searched MEDLINE, Web of Science, and Embase for peer-reviewed papers published in English between June 1st 2013 (the end date of a related review)[14] and May 1st 2025 (the date the search was conducted). Because this research area does not use consistent terminology[15,16,23,24], we developed our search query by manually identifying 20 relevant articles that met our inclusion

criteria and then adjusted our query until it identified all 20. This query ensured that the title or abstracts of studies included: a term about AI (e.g., artificial intelligence, machine learning, etc.), a term about medicine (e.g., diagnose, patient, medicine, etc.), and a term about assisted decision-making (e.g., decision-support, recommendation, human-AI, etc.). Full search queries are available in the Supplementary Materials and an Open Science Framework repository (https://osf.io/un32b/?view_only=efbeaf42a9ef4aaab45a2cde69450265).

We leveraged the literature review tool Covidence[31] to screen articles, perform full-text review, and extract data. Titles and abstracts were screened by one reviewer, who maximized recall with loosened inclusion criteria (i.e., including any study where the title or abstract mentioned any human interaction with AI in a medical context). Full-text articles were then screened by 2 reviewers according to these inclusion criteria:

1. Studies must have conducted an experiment or observed an actual implementation of AI-CDS and collected data on participant use of AI-CDS.
2. Study participants must be healthcare professionals with clinical knowledge relevant to the decision-making task under study.
3. Studies must have assessed either: an objective measure of a user's actions (e.g., the decision to accept or reject a recommendation from AI-CDS); a subjective measure of a user's attitudes towards the AI-CDS (e.g., their trust in the AI-CDS); or their performance with AI-CDS (e.g., accuracy or area under the receiver operating characteristics curve).
4. Studies must have measured or manipulated an additional variable to determine its effect on either the clinicians' attitudes towards the AI-CDS, their decisions/actions when using the AI-CDS, or their performance when using the AI-CDS.
5. Studies must have included at least 30 decision-makers.

All disagreements between reviewers were settled via consensus.

We extracted information on study design and participants, which a decision-making or performance outcome was studied (i.e., attitudes, actions, and performance), how these outcome variables were operationalized, and what factors were studied for their impact on these outcomes. When measuring performance, we used the primary performance measure reported by the authors of each study as each application of AI-CDS has its own unique objectives and performance criteria. Additionally, to enable more granular analysis, we differentiated the performance of clinicians alone from clinicians using AI-CDS (hereafter, human-AI collaborative performance) and from the AI-CDS alone. Extracted information was processed via a custom Python script to clean and map free-text inputs onto discrete categories. The protocol for this study was registered with the Open Science Framework. Extracted information, code, and citations are provided in an Open Science Framework repository (https://osf.io/un32b/?view_only=efbeaf42a9ef4aaab45a2cde69450265).

RESULTS

Screening

We identified 10,027 articles, 4,177 of which were removed as duplicates. The title and abstract screening excluded 5,090 articles, and the full-text screening phase removed 715 articles, leaving 45 articles for inclusion in our study (Figure 1). The most common exclusion reasons at the full-text stage were due to insufficient numbers of participants (N=222 studies) and the lack of an independent variable that was assessed for its impact on human-AI interaction (N=209). The full list of included studies is included in Table 1.

Authors and Year	Title	Sample Size	Factors Studied	Outcomes Measured
Adam et al. 2022	Mitigating the impact of biased artificial intelligence in emergency decision-making.	954	communication style	actions
Bond et al. 2018	Automation bias in medicine: The influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms.	30	correctness of AI and expertise	attitudes
Cabitza et al. 2019	Biases Affecting Human Decision Making in AI-Supported Second Opinion Settings	75	baseline attitudes and expertise	attitudes
Cabitza et al. 2023	Rams, hounds and white boxes: Investigating human-AI collaboration protocols in medical diagnosis.	56	expertise, explainable AI, and order of information	performance
Calisto et al. 2022	BreastScreening-AI: Evaluating medical intelligent agents for human-AI interactions.	45	expertise and explainable AI	attitudes and actions
Calisto et al. 2025	Personalized explanations for clinician-AI interaction in breast imaging diagnosis by adapting communication to expertise levels	52	communication style	attitudes and performance
Carmichael et al. 2024	Diagnostic decisions of specialist optometrists exposed to ambiguous deep-learning outputs.	30	expertise and explainable AI	attitudes, actions, and performance
Chen et al. 2024	Trust in Machine Learning Driven Clinical Decision Support Tools Among Otolaryngologists.	45	baseline attitudes, expertise, and explainable AI	actions
deOliveira et al. 2025	Effect of Explainable Artificial Intelligence on Trust of Mental Health Professionals in an AI-Based System for Suicide Prevention	78	AI knowledge, correctness of AI, expertise, and explainable AI	attitudes
Dorr et al. 2020	COVID-19 pneumonia accurately detected on chest radiographs with artificial intelligence.	54	expertise	performance
Festor et al. 2025	Safety of human-AI cooperative decision-making within intensive care: A physical simulation study	38	correctness of AI	actions
Fritz et al. 2024	Effect of machine learning models on clinician prediction of postoperative complications: the Perioperative ORACLE randomised clinical trial	89	correctness of AI and expertise	performance
Gaube et al. 2021	Do as AI say: susceptibility in deployment of clinical decision-aids.	265	baseline attitudes, correctness of AI, and expertise	attitudes and performance
Gaube et al. 2023	Non-task expert physicians benefit from correct explainable AI advice when reviewing X-rays.	223	baseline attitudes, expertise, and explainable AI	attitudes, actions, and performance
Goel et al. 2022	The effect of machine learning explanations on user trust for automated diagnosis of COVID-19.	30	explainable AI	attitudes
Goh et al. 2024	Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial	50	AI knowledge and expertise	performance
Gombolay et al. 2024	Effects of explainable artificial intelligence in neurology decision support.	365	expertise and explainable AI	attitudes and performance

Gomez et al. 2024	Explainable AI decision support improves accuracy during telehealth strep throat screening	121	expertise and explainable AI	attitudes, actions, and performance
Groh et al. 2024	Deep learning-aided decision support for diagnosis of skin disease across skin tones.	1118	correctness of AI and expertise	performance
Guo et al. 2024	Deep Learning for Chest X-ray Diagnosis: Competition Between Radiologists with or Without Artificial Intelligence Assistance.	111	expertise	performance
Jabbour et al. 2023	Measuring the Impact of AI in the Diagnosis of Hospitalized Patients: A Randomized Clinical Vignette Survey Study.	418	correctness of AI, expertise, and explainable AI	performance
Jacobs et al. 2021	How machine-learning recommendations influence clinician treatment selections: the example of the antidepressant selection.	220	AI knowledge, correctness of AI, and explainable AI	attitudes and performance
Jain et al. 2021	Development and Assessment of an Artificial Intelligence-Based Tool for Skin Condition Diagnosis by Primary Care Physicians and Nurse Practitioners in Teledermatology Practices.	40	expertise	attitudes and performance
Jin et al. 2024	Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task.	35	AI knowledge and explainable AI	attitudes, actions, and performance
Küper et al. 2025	Psychological Factors Influencing Appropriate Reliance on AI-enabled Clinical Decision Support Systems: Experimental Web-Based Study Among Dermatologists	223	baseline attitudes, correctness of AI, and expertise	attitudes, actions, and performance
Knoery et al. 2019	SPICED-ACS: Study of the potential impact of a computer-generated ECG diagnostic algorithmic certainty index in STEMI diagnosis: Towards transparent AI	91	expertise and uncertainty quantification	attitudes and performance
LancasterFarrell et al. 2022	Explainability does not improve biochemistry staff trust in artificial intelligence-based decision support.	159	explainable AI	actions and performance
Laxar et al. 2023	The influence of explainable vs non-explainable clinical decision support systems on rapid triage decisions: a mixed methods study.	32	baseline attitudes, expertise, and explainable AI	attitudes and actions
Lee et al. 2023	Effect of Human-AI Interaction on Detection of Malignant Lung Nodules on Chest Radiographs.	30	expertise	actions and performance
Li et al. 2022	How does the artificial intelligence-based image-assisted technique help physicians in diagnosis of pulmonary adenocarcinoma? A randomized controlled experiment of multicenter physicians in China.	104	available time and baseline attitudes	performance
Li et al. 2025	Understanding physicians' noncompliance use of AI-aided diagnosis-A mixed-methods approach	160	baseline attitudes	actions
Luan et al. 2025	Machine Learning-Aided Diagnosis Enhances Human Detection of Perilunate Dislocations	137	expertise	performance
Maehara et al. 2025	Artificial intelligence support improves diagnosis accuracy in anterior segment eye diseases	40	expertise	performance
Micocci et al.	Attitudes towards Trusting Artificial	50	baseline attitudes	actions and

2021	Intelligence Insights and Factors to Prevent the Passive Adherence of GPs: A Pilot Study.		and correctness of AI	performance
Nagendran et al. 2023	Quantifying the impact of AI recommendations with explanations on prescription decision making.	86	baseline attitudes, expertise, and explainable AI	attitudes and actions
Prinster et al. 2024	Care to Explain? AI Explanation Types Differentially Impact Chest Radiograph Diagnostic Performance and Physician Trust in AI	220	correctness of AI, expertise, explainable AI, and uncertainty quantification	attitudes and performance
Rainey et al. 2023	An experimental machine learning study investigating the decision-making process of students and qualified radiographers when interpreting radiographic images.	106	expertise	attitudes
Senoner et al. 2024	Explainable AI improves task performance in human-AI collaboration	113	explainable AI	actions and performance
Tschandl et al. 2020	Human-computer collaboration for skin cancer recognition.	456	baseline attitudes, correctness of AI, expertise, and explainable AI	actions and performance
Wang et al. 2023	Artificial intelligence suppression as a strategy to mitigate artificial intelligence automation bias.	40	expertise	actions and performance
Wang et al. 2024	A Deep Learning Model Enhances Clinicians' Diagnostic Accuracy to More Than 96% for Anterior Cruciate Ligament Ruptures on Magnetic Resonance Imaging.	38	expertise and task difficulty	performance
Williams et al. 2024	Artificial Intelligence Assisted Surgical Scene Recognition: A Comparative Study Amongst Healthcare Professionals	348	expertise	performance
Yoon et al. 2023	Developing and Evaluating an AI-Based Computer-Aided Diagnosis System for Retinal Disease: Diagnostic Study for Central Serous Chorioretinopathy.	66	expertise and explainable AI	performance
Yu et al. 2024	Heterogeneity and predictors of the effects of AI assistance on radiologists.	140	correctness of AI and expertise	performance
Zhang et al. 2025	Evaluating the effectiveness of a clinical decision support system (AI-Antidelirium) to improve Nurses' adherence to delirium guidelines in the intensive care unit	80	expertise and longitudinality	performance

Table 1: Overview of included studies

Study Designs & Sample Sizes

The number of decision-making participants across studies was skewed with a mean of 161.4 participants and a median of 86.0, largely due to a small number of studies that had over 250 participants[7,34–40] (Figure 2). Most studies used image or video-based modalities (N=29) with the

rest using clinical vignettes (N=12) or electrocardiograms (N=4). Studies primarily recruited attendings (N=42) or physician trainees such as interns, residents, or fellows (N=24). A smaller number of studies recruited nurses (N=3) and advanced practice providers (N=3) with registered nurse anesthetists, mental health professionals, ambulance staff, biochemistry staff, and medical students each being recruited in one study.

Outcomes Assessed in the Included Studies

Participants' performance with AI-CDS was the most studied outcome (N=33, 73% of studies), followed by participant attitudes (N=20, 44%) and actions (N=18, 40%). Importantly, few studies measured all three of these outcomes (N=5, 11%). Participants' performance was primarily measured via accuracy (e.g., percentage of decisions that were correct, N=26/33), followed by area under the receiver operating characteristic curve (N=4/33), sensitivity (N=1/3), rate of adherence to clinical guidelines (N=1/33), and a custom score provided by subject-matter experts (N=1/33).

Participants' actions were most commonly measured via the agreement between the clinician and the AI-CDS (N=8/18). The next most common measure of users' actions was decision-switching (N=7/18) which was broken down into: (1) the rate at which participants changed a binary decision (e.g., a diagnosis) after receiving advice from AI-CDS (N=5/18) and (2) weight-on-advice[41], a weighted measure of how much a participant changed their decision about a continuous quantity (e.g., selecting a medication dose), (N=2/18). Additionally, two studies measured the participants' level of agreement with AI-CDS when the AI-CDS provided incorrect advice, which was referred to as 'over-reliance' or 'automation bias'.

In the 20 studies assessing clinician attitudes towards AI, trust was the most commonly studied clinician attitude (N=11/20). This was followed by a series of indirect measures of clinicians' attitudes toward AI-CDS. Namely, clinicians' confidence either in the AI-CDS or their AI-assisted decisions (N=8/20), clinicians' perception of the AI-CDS' utility or quality (N=7/20), and clinicians' self-reported understanding of the AI-CDS (N=4/20).

Factors Impacting the Effectiveness of AI-Assisted Decision-Making

Expertise*:*

The most frequently studied variable was expertise (i.e., whether the clinician using the AI system was an expert in the clinical domain under study) (N=34). Which individuals were considered experts differed in each study as these studies covered many different clinical domains. Typically, experts were distinguished from non-experts via their years of experience or their level of clinical training (e.g., subspecialist vs attending vs resident vs medical student). Nine studies found that AI assistance improved performance for experts and non-experts to a similar degree,[7,10,11,35,42–46] two found that AI improved performance more for experts,[40,47] and six found that AI improved performance more for non-experts[39,48–52].

In the 11 studies that assessed the impact of expertise on clinicians' actions, the effect was mixed. Four studies found that experts agreed with or aligned their decisions to AI-CDS recommendations less often than novices[39,44,48,53], but seven studies found no clear relationship between participant expertise and their actions with AI[54–56].

Of the 5 studies measuring the impact of expertise on clinicians' attitudes toward AI, 4 found that experts perceived AI as less helpful or exhibited lower trust in it than non-experts[53,57–59]. However, an additional study suggested no relationship between expertise and trust[60].

Several studies additionally evaluated how expertise interacted with other variables such as explainable AI (xAI) or incorrect advice. Specifically, in one study experts rated xAI as lower quality than non-experts[61]. Another study found that those experts who did find xAI useful (i.e., explainable[62]) exhibited decreased performance[37] while two studies reported that experts were less likely to follow incorrect advice than non-experts[36,48].

Explainable AI:

The next most studied factor was the use of xAI to better communicate the AI-CDS' decision-making process to clinicians (N=19). We defined xAI broadly as any AI system containing a component that "aims to increase the transparency, trustworthiness and accountability of the AI system"[63,64]. Therefore, any approach that the authors described as intending to explain, support, or clarify AI-CDS decisions was considered xAI for the purposes of this study. Notably, there were mixed effects of xAI on performance. Of the 13 such studies, five studies found that xAI improved performance[52,58,61,65,66], but six identified no effect[37,38,49,67–69], and two showed that xAI worsened human-AI collaborative performance compared to human-AI performance with non-explainable AI[39,57].

This heterogeneous effect of xAI on performance was mirrored by the effect of xAI on clinicians' attitudes toward AI-CDS. We identified that xAI could increase[55,57,61,67], decrease[58,60,70], or have no effect[37,66,69] on trust-related attitudes. However, these results were often observed under different circumstances. For example, xAI decreased clinicians' trust and understanding of the AI when it 'explained' incorrect advice[60,70], which is a desirable outcome as it potentially decreases acceptance of incorrect AI advice. Similarly, two studies found that xAI increased clinicians' acceptance of correct AI advice, thereby increasing human-AI collaborative performance[61,65]. However, these beneficial effects of xAI were not observed in all studies. In

three studies, xAI was more likely to decrease participants' performance when the AI-CDS was wrong (i.e., it convinced users to accept incorrect AI recommendations more frequently than with non-explainable AI)[65,66,69].

The variation among these results is potentially explained by differences in the type of xAI used. For example, when comparing local explanations vs global explanations (i.e., explanations about the model's architecture and overall performance), local explanations (i.e., decision-specific or patient-specific) were more effective, increasing performance[66] and agreement with AI advice[54]. Two additional studies compared different local xAI approaches, finding no differences in performance[37,58].

AI Correctness:

Thirteen studies stratified their analyses on whether the advice from the AI-CDS was correct or incorrect. Unsurprisingly, these studies often found that incorrect AI-CDS advice decreased diagnostic performance[11,35,36,38,39,66,69,71,72]. However, participants showed some resilience to this as incorrect advice decreased users' trust[60] in the AI-CDS and decreased their confidence in their decisions[66,72] (with one exception[69]). Moreover, participants rated correct advice as being more useful[36,66] than incorrect advice and agreed with incorrect AI recommendations less frequently than correct ones[44,73]. In particular, both experts[36,48] and those with higher self-confidence[71] were less susceptible to following incorrect AI advice.

Baseline Attitudes Towards AI:

Eleven studies assessed users' baseline attitudes about the AI-CDS or AI in general before they interacted with the AI-CDS. Participants with negative baseline attitudes towards AI were less likely to agree with AI recommendations[44,74]. These negative attitudes caused participants to

accept advice more frequently when told that it came from another human[53,61]. Conversely, those with higher trust in technology agreed with AI-CDS more frequently[44]. Lastly, only one study found that baseline attitudes toward AI did not impact decision-making[56].

Emerging Evidence:

A few factors were assessed by only one or two studies each. For example, two studies assessed the impact of conveying the AI-CDS' level of certainty to participants (e.g., the probability of the AI-CDS' decision being correct). One such study found that this increased participants' diagnostic performance; however, clinicians' performance decreased when the AI-CDS conveyed that it was highly certain, possibly indicating automation bias[75]. The other such study observed different decision-making patterns between experts and non-experts when AI was uncertain. Specifically, with low-certainty AI recommendations, experts benefitted from xAI but non-experts did not[66]. Moreover, several studies assessed participants' knowledge of AI, finding that educational interventions did not impact participants' trust in AI[60], but sharing performance metrics of the AI-CDS increased clinicians' trust[67]. Another study found that users who were more familiar with AI rated AI-CDS as having higher utility[69]. However, these studies found no effect[10] or a detrimental effect[69] of AI knowledge on decision-making performance.

Two studies assessed AI-CDS communication styles. Specifically, one study found that providing prescriptive advice (i.e., instructions on what to do) increased agreement with AI compared to descriptive advice (i.e., neutrally conveying information)[34]. Another study found that a similar strategy to prescriptive advice (which the authors refer to as assertive advice) had a positive effect on some measures of clinicians' attitudes towards the AI-CDS[76]. One study assessed the order in which AI recommendations were provided to participants (i.e., before vs after making their own decision about the patient). They observed that human-AI collaborative performance

was higher when AI recommendations were presented before the clinician made their own assessment of the patient[49].

Participants' actions with AI were found to evolve over time; one study observed that clinicians' adherence to clinical guidelines increased continuously over a two-week AI-CDS intervention period[77]. Lastly, applying time restrictions to participants was found to hinder their diagnostic performance without AI, but when AI-CDS was available, participants' reliance on the AI-CDS offset this impairment and increased performance to the level of their unaffected (control arm) counterparts[78]. Moreover, when given recommendations from the AI-CDS that differed from their initial decision, participants with lower confidence were more likely to switch their decision to that of the AI-CDS[39,54].

DISCUSSION

This study sought to clarify existing research on AI-CDS by identifying factors that influence the effectiveness of AI-assisted decision-making in medical settings. Of the three outcomes relevant to effectiveness that we looked at, clinician performance with AI (e.g., diagnostic accuracy) was much more commonly measured than clinician attitudes (e.g., trust) or clinician decisions (e.g., rates of agreement with AI-CDS recommendations). We note that measuring human-AI collaborative performance without assessing clinician-level decisions or attitudes such as trust leaves some uncertainty about reasons why collaborative performance improved or failed to do so, yet only five studies measured all three outcomes. Four factors were commonly studied for their impact on these outcomes: clinician baseline attitude toward AI, clinician expertise, AI

explainability, and AI correctness. Of these, baseline attitude and AI correctness appeared to have the most consistent impacts.

Clinicians' attitudes toward AI (e.g., their trust in AI) guided their tendency to accept the AI-CDS' recommendations and when clinicians were given incorrect advice from AI-CDS, their performance decreased. Consequently, we identified that when AI-CDS provides incorrect advice often enough and when clinicians' trust in this AI-CDS is high, it has the potential to decrease clinicians' performance relative to their baseline without AI assistance[57]. Because clinicians (of all experience levels) may not be able to identify incorrect AI advice, blindly increasing clinicians' trust in AI recommendations can increase how often they defer to such advice and potentially decrease their decision-making performance. This observation suggests the need for a paradigm shift; instead of aiming simply to increase trust in AI, a focus of much prior work[79–81], researchers should aim for trust that is "just right," promoting a level of trust that allows clinicians and health systems to benefit from AI-CDS, while avoiding the pitfalls of automation bias and over-reliance on AI[15,16,72]. This concept has been well-studied for non-AI technologies[82], under the name of "appropriate trust"[83] and has shown utility in medical and non-medical applications[14]. Under the concept of appropriate trust, clinicians learn the optimal level of trust they should place in AI-CDS relative to their own level of decision-making performance. In doing so, they aim to strike an ideal balance between under-utilization of AI and over-reliance on AI.

We found substantial heterogeneity in how clinicians' level of expertise impacted AI-assisted decision-making performance. However, the framework of appropriate trust explains several of these findings. Specifically, while expertise did not reliably predict which clinicians' performance would increase from AI-CDS assistance, we found that more experienced clinicians

(experts) were generally less trusting of AI-CDS and were less likely to accept AI recommendations[39,44,48,53]. Consequently, they were less likely to benefit from AI-CDS when it correctly provided advice that contradicted their own assessment. However, this also made experts more resistant to over-reliance (i.e., accepting incorrect AI advice) than their junior counterparts[36,48]. The synthesis of these observations is that expertise most directly predicts a clinicians' tendency to accept AI recommendations. However, whether this difference in acceptance rates will increase clinicians' decision-making performance is dependent on how well these clinicians perform relative to the AI-CDS; if the clinician performs well, then they likely only stand to decrease their performance by accepting AI advice. Similarly, if the clinician performs poorly on their own (as might be expected for non-experts), then higher trust in AI is beneficial as they have more room to improve from AI assistance. In the framework of appropriate trust, this is referred to as "trust calibration,"[82] the concept that trust should be calibrated relative to the level of benefit that the AI-CDS stands to bring. In this respect, clinicians' trust in AI can be considered well-calibrated with respect to clinical expertise as those who stand to benefit the least from AI-CDS also trust AI-CDS the least. In our review, this was highlighted most directly by Tschandl et al.,[39] who identified that "if experts have high confidence in their initial diagnosis, they should ignore AI-based support or not use it at all".

The concept of appropriate trust also enables us to analyze the heterogeneous results observed with respect to xAI. Specifically, our literature review shows that xAI often increases clinicians' trust in AI-CDS but does not necessarily make this trust more appropriate. That is, xAI can increase clinicians' acceptance of incorrect advice from the AI-CDS[65,66,69] and when it does so often enough, this can harm human-AI collaborative performance[39,57]. This finding has been corroborated by recent non-medical research[84–87]. Moreover, a recent review of xAI identified

that xAI can exacerbate certain cognitive biases, leading to decreases in decision-making accuracy[88]. However, we also identified conflicting evidence where xAI led to more appropriate trust. Specifically, xAI increased clinician acceptance of high-performing AI-CDS[61,65] and decreased their trust in the AI-CDS when it provided incorrect recommendations[60,70]. Recent systematic reviews point out that these inconsistent findings may result from the lack of standard evaluation procedures for xAI[89,90]. In this context, our results highlight the potential for xAI to increase the appropriateness of clinician trust in AI and support the role of appropriate trust as a component in xAI evaluation frameworks.

Recommendations:

To make sense of the heterogeneous results observed in existing AI-CDS research, we recommend that future research into AI-CDS should leverage the concept of appropriate trust[82,83] to guide the design and evaluation of AI-CDS systems. Specifically, researchers should measure: (1) clinicians' attitudes towards the AI system (e.g., trust), (2) clinicians' actions when presented with the AI-CDS (e.g., their acceptance or non-acceptance of AI-CDS recommendations), and (3) performance of the clinicians alone, the AI alone, and the human-AI collaboration. All three of these components are necessary to understand the complete picture around AI-assisted decision-making in medicine, but relatively little current research assesses all of these outcomes. Consider a study that assesses only trust and AI-assisted performance. This study may identify that users trust the AI-CDS, but upon deployment of the AI-CDS in clinical settings, it may fail to identify barriers that prevent clinicians from accessing or accepting its advice. Similarly, if a study measures performance and clinicians' acceptance of AI recommendations without first asking clinicians if they trust the AI-CDS, they may find that clinicians' lack of trust[10,91] or over-trust[15,32] in AI leads to inappropriate usage of the AI-CDS. Lastly, because the appropriateness of

an individual clinicians' trust in AI depends on the relative performance of the clinician compared to the AI system[83], studies must evaluate the performance of the clinician and of the AI alone to assess the appropriateness of human-AI collaborative decision-making.

Limitations:

Our review has several limitations. First, as with all literature reviews, some literature may have been missed during our search. Second, the current literature on AI-CDS focused largely on simple decisions and AI-CDS without much interactive capability. Because copilot-style AI systems[6,10] bring unique interactive capabilities beyond the AI systems studied in current research, future work must be done to understand human-AI collaboration with these more dynamic AI systems. Third, our review required that users directly interact with an AI system in an experimental or in a real-world clinical setting, as this is the only way decision-making could be reliably studied. However, this omitted a substantial body of qualitative research about the design of AI interfaces, users' experiences with the system, and sociotechnical issues potentially arising from the AI system that may impact clinicians' attitudes towards AI. Fourth, we excluded a considerable number of studies with small sample sizes as estimates derived from very small sample sizes are unlikely to be robust, and generalizability may also be limited by the limited range of participants. Nevertheless, this decision may have excluded some relevant research.

Conclusion:

This scoping review identified factors that influence AI-assisted decision-making. The most robust findings in our study were that clinicians' attitudes towards AI (e.g., trust) consistently predicted their acceptance of AI recommendations and that when AI provided incorrect recommendations, clinicians' decision-making performance decreased substantially. We

integrated these findings with theoretical frameworks and used the concept of appropriate trust to provide recommendations for more robust research into AI-assisted medical decision-making. We used this framework to analyze why the most commonly studied factors in our review - explainable AI and clinicians' level of expertise - had heterogeneous impacts on decision-making performance. In this context, we found signals that clinicians' trust in AI may be naturally appropriate – at a broad cohort level more experienced (typically higher-performing) clinicians were less trusting of AI-CDS than their junior counterparts, who stand to benefit more from AI assistance. However, trust may not be appropriate at the individual level as clinicians often could not reliably distinguish between correct recommendations which should be trusted and incorrect ones that should not. Similarly, explainable AI showed promise as it often increased clinicians' trust in highly effective AI systems and reduced trust in error-prone ones. However, the benefits of xAI were not universally observed; xAI occasionally increased acceptance of incorrect advice. We conclude that the concept of appropriate trust is an essential tool for understanding the complex interactions between clinicians and AI-CDS systems and that future research should leverage this concept to facilitate more robust study of AI-assisted decision-making in medicine.

Acknowledgements

This study was supported by the National Library of Medicine (T15LM007450) and the Agency for Healthcare Research and Quality (T32HS026122). Dr. Benda is supported by the National Institute on Minority Health and Health Disparities (R00MD015781). The funders played no role in the study design, data collection, analysis and interpretation of data, or the writing of the manuscript.

Competing Interests

All authors declare no financial or non-financial competing interests.

Data Availability

The minimum necessary materials to generate results presented in this paper (i.e., included studies and relevant data attributes) are contained in the Supplementary Materials. The full set of studies, both included and excluded, as well as all extracted information, and the code used to generate figures and complete the analysis are available in an Open Science Framework repository (https://osf.io/un32b/?view_only=efbeaf42a9ef4aaab45a2cde69450265).

Code Availability

The code necessary to reproduce the analysis in this paper and generate figures, along with a readme file is available in an Open Science Framework repository (https://osf.io/un32b/?view_only=efbeaf42a9ef4aaab45a2cde69450265).

Author Contributions

NJJ, KEB, MRC, JSA contributed to the study design. NJJ conducted the literature search, and NJJ, KEB, RM, MM, and BC screened articles. NJJ and KEB processed the data. NJJ drafted the initial manuscript. NJJ, KEB, MRC, BC, LLN, NCB, JSA contributed to data analysis and interpretation, and reviewed the manuscript. NJJ and JSA were responsible for integrating revisions and finalizing the manuscript.

References:

1. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. Npj Digit Med. 2021 Apr 7;4(1):65.

2. Xiaoxuan Liu, Liu X, Livia Faes, Faes L, Aditya U Kale, Kale A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. 2019 Oct 1;1(6).

3. Brodeur PG, Buckley TA, Kanjee Z, Goh E, Ling EB, Jain P, et al. Superhuman performance of a large language model on the reasoning tasks of a physician [Internet]. arXiv; 2025 [cited 2025 Jul 23]. Available from: http://arxiv.org/abs/2412.10849

4. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med. 2019 Jan;25(1):65–9.

5. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017 Feb 2;542(7639):115–8.

6. Huang Z, Yang E, Shen J, Gratzinger D, Eyerer F, Liang B, et al. A pathologist–AI collaboration framework for enhancing diagnostic accuracies and efficiencies. Nat Biomed Eng [Internet]. 2024 Jun 19 [cited 2024 Jul 10]; Available from: https://www.nature.com/articles/s41551-024-01223-5

7. Groh M, Badri O, Daneshjou R, Koochek A, Harris C, Soenksen LR, et al. Deep learning-aided decision support for diagnosis of skin disease across skin tones. Nat Med. 2024 Feb;30(2):573–83.

8. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. Npj Digit Med. 2020 Feb 6;3(1):17.

9. Krakowski I, Kim J, Cai ZR, Daneshjou R, Lapins J, Eriksson H, et al. Human-AI interaction in skin cancer diagnosis: a systematic review and meta-analysis. Npj Digit Med. 2024 Apr 9;7(1):78.

10. Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, et al. Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial. JAMA Netw Open. 2024 Oct 28;7(10):e2440969.

11. Yu F, Moehring A, Banerjee O, Salz T, Agarwal N, Rajpurkar P. Heterogeneity and predictors of the effects of AI assistance on radiologists. Nat Med. 2024 Mar;30(3):837–49.

12. Walker SC, French B, Moore RP, Domenico HJ, Wanderer JP, Mixon AS, et al. Model-Guided Decision-Making for Thromboprophylaxis and Hospital-Acquired Thromboembolic Events Among Hospitalized Children and Adolescents: The CLOT Randomized Clinical Trial. JAMA Netw Open. 2023 Oct 13;6(10):e2337789.

13. Taylor RA, Chmura C, Hinson J, Steinhart B, Sangal R, Venkatesh AK, et al. Impact of Artificial Intelligence–Based Triage Decision Support on Emergency Department Care. NEJM AI [Internet]. 2025 Feb 27 [cited 2025 Jul 22];2(3). Available from: https://ai.nejm.org/doi/10.1056/AIoa2400296

14. Hoff KA, Bashir M. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. Hum Factors J Hum Factors Ergon Soc. 2015 May;57(3):407–34.

15. David Lyell, Lyell D, Enrico Coiera, Coiera E. Automation bias and verification complexity: a systematic review. J Am Med Inform Assoc. 2017 Mar 1;24(2):423–31.

16. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc JAMIA. 2012;19(1):121–7.

17. Newton N, Bamgboje-Ayodele A, Forsyth R, Tariq A, Baysari MT. A systematic review of clinicians' acceptance and use of clinical decision support systems over time. Npj Digit Med. 2025 May 26;8(1):309.

18. Kundu S. AI in medicine must be explainable. Nat Med. 2021 Aug;27(8):1328–1328.

19. Jones C, Thornton J, Wyatt JC. Artificial intelligence and clinical decision support: clinicians' perspectives on trust, trustworthiness, and liability. Med Law Rev. 2023 Nov 27;31(4):501–20.

20. Buck C, Doctor E, Hennrich J, Jöhnk J, Eymann T. General Practitioners' Attitudes Toward Artificial Intelligence–Enabled Systems: Interview Study. J Med Internet Res. 2022 Jan 27;24(1):e28916.

21. Kosch T, Welsch R, Chuang L, Schmidt A. The Placebo Effect of Artificial Intelligence in Human–Computer Interaction. ACM Trans Comput-Hum Interact. 2022 Dec 31;29(6):1–32.

22. Scott IA, Carter SM, Coiera E. Exploring stakeholder attitudes towards AI in clinical practice. BMJ Health Care Inform. 2021 Dec;28(1):e100450.

23. Ueno T, Sawa Y, Kim Y, Urakami J, Oura H, Seaborn K. Trust in Human-AI Interaction: Scoping Out Models, Measures, and Methods. In: CHI Conference on Human Factors in Computing Systems Extended Abstracts [Internet]. New Orleans LA USA: ACM; 2022 [cited 2024 Jun 2]. p. 1–7. Available from: https://dl.acm.org/doi/10.1145/3491101.3519772

24. Lai V, Chen C, Smith-Renner A, Liao QV, Tan C. Towards a Science of Human-AI Decision Making: An Overview of Design Space in Empirical Human-Subject Studies. In: 2023 ACM Conference on Fairness, Accountability, and Transparency [Internet]. Chicago IL USA: ACM; 2023 [cited 2024 May 21]. p. 1369–85. Available from: https://dl.acm.org/doi/10.1145/3593013.3594087

25. Vereschak O, Bailly G, Caramiaux B. How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. Proc ACM Hum-Comput Interact. 2021 Oct 13;5(CSCW2):1–39.

26. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform. 2021 Jan;113:103655.

27. Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. Npj Digit Med. 2021 Jan 5;4(1):1–6.

28. Begoli E, Bhattacharya T, Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. Nat Mach Intell. 2019 Jan 7;1(1):20–3.

29. Bach TA, Khan A, Hallock H, Beltrão G, Sousa S. A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective. Int J Human–Computer Interact. 2024 Mar 3;40(5):1251–66.

30. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018 Oct 2;169(7):467–73.

31. Covidence systematic review software [Internet]. Melbourne, Australia: Veritas Health Innovation; Available from: www.covidence.org

32. Chen S, Guevara M, Moningi S, Hoebers F, Elhalawani H, Kann BH, et al. The effect of using a large language model to respond to patient messages. Lancet Digit Health. 2024 Jun;6(6):e379–81.

33. Asan O, Bayrak AE, Choudhury A. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. J Med Internet Res. 2020 Jun 19;22(6):e15154.

34. Adam H, Balagopalan A, Alsentzer E, Christia F, Ghassemi M. Mitigating the impact of biased artificial intelligence in emergency decision-making. Commun Med. 2022 Nov 21;2(1):149.

35. Fritz BA, King CR, Abdelhack M, Chen Y, Kronzer A, Abraham J, et al. Effect of machine learning models on clinician prediction of postoperative complications: the Perioperative ORACLE randomised clinical trial. Br J Anaesth. 2024;133(5):1042–50.

36. Gaube S, Suresh H, Raue M, Merritt A, Berkowitz SJ, Lermer E, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. Npj Digit Med. 2021 Feb 19;4(1):1–8.

37. Gombolay GY, Silva A, Schrum M, Gopalan N, Hallman Cooper J, Dutt M, et al. Effects of explainable artificial intelligence in neurology decision support. Ann Clin Transl Neurol. 2024 May;11(5):1224–35.

38. Jabbour S, Fouhey D, Shepard S, Valley TS, Kazerooni EA, Banovic N, et al. Measuring the Impact of AI in the Diagnosis of Hospitalized Patients: A Randomized Clinical Vignette Survey Study. JAMA. 2023 Dec 19;330(23):2275–84.

39. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella NCF, Halpern AC, et al. Human–computer collaboration for skin cancer recognition. Nat Med. 2020 Jun 22;26(8):1229–34.

40. Williams SC, Zhou J, Muirhead WR, Khan DZ, Koh CH, Ahmed R, et al. Artificial Intelligence Assisted Surgical Scene Recognition: A Comparative Study Amongst Healthcare Professionals. Ann Surg [Internet]. 2024;((Williams S.C., simon.williams32@nhs.net; Zhou J., zjf@umich.edu; Muirhead W.R., w.muirhead@ucl.ac.uk; Khan D.Z., d.khan@ucl.ac.uk; Koh C.H., austin.koh@ucl.ac.uk; Ahmed R., razna.ahmed.21@alumni.ucl.ac.uk; Funnell J.P., jonathan.funnell.13@ucl.ac.uk; Han). Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2035484291&from=export

41. Bailey PE, Leon T, Ebner NC, Moustafa AA, Weidemann G. A meta-analysis of the weight of advice in decision-making. Curr Psychol. 2023 Oct;42(28):24516–41.

42. Dorr F, Chaves H, Serra MM, Ramirez A, Costa ME, Seia J, et al. COVID-19 pneumonia accurately detected on chest radiographs with artificial intelligence. Intell Based Med. 2020;3:100014.

43. Jain A, Way D, Gupta V, Gao Y, de Oliveira Marinho G, Hartford J, et al. Development and Assessment of an Artificial Intelligence-Based Tool for Skin Condition Diagnosis by Primary Care Physicians and Nurse Practitioners in Teledermatology Practices. JAMA Netw Open. 2021;4(4):e217249.

44. Küper A, Lodde GC, Livingstone E, Schadendorf D, Krämer N. Psychological Factors Influencing Appropriate Reliance on AI-enabled Clinical Decision Support Systems: Experimental Web-Based Study Among Dermatologists. J Med Internet Res [Internet]. 2025;27((Küper A., alisa.kueper@uni-due.de; Krämer N.) Social Psychology: Media and Communication, University of Duisburg-Essen, Duisburg, Germany(Lodde G.C.; Livingstone E.; Schadendorf D.) Department of Dermatology, University Hospital Essen, Essen, Germany). Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2038219509&from=export

45. Lee JH, Hong H, Nam G, Hwang EJ, Park CM. Effect of Human-AI Interaction on Detection of Malignant Lung Nodules on Chest Radiographs. Radiology. 2023;307(5):e222976.

46. Maehara H, Ueno Y, Yamaguchi T, Kitaguchi Y, Miyazaki D, Nejima R, et al. Artificial intelligence support improves diagnosis accuracy in anterior segment eye diseases. Sci Rep. 2025;15(1).

47. Guo L, Zhou C, Xu J, Huang C, Yu Y, Lu G. Deep Learning for Chest X-ray Diagnosis: Competition Between Radiologists with or Without Artificial Intelligence Assistance. J Imaging Inf Med. 2024;

48. Wang DY, Ding J, Sun AL, Liu SG, Jiang D, Li N, et al. Artificial intelligence suppression as a strategy to mitigate artificial intelligence automation bias. J Am Med Inf Assoc. 2023;30(10):1684–92.

49. Cabitza F, Campagner A, Ronzio L, Cameli M, Mandoli GE, Pastore MC, et al. Rams, hounds and white boxes: Investigating human-AI collaboration protocols in medical diagnosis. Artif Intell Med. 2023;138:102506.

50. Luan A, von Rabenau L, Serebrakian A, Crowe C, Do B, Eberlin K, et al. Machine Learning-Aided Diagnosis Enhances Human Detection of Perilunate Dislocations. HAND-Am Assoc HAND Surg. 2025;

51. Wang DY, Liu SG, Ding J, Sun AL, Jiang D, Jiang J, et al. A Deep Learning Model Enhances Clinicians' Diagnostic Accuracy to More Than 96% for Anterior Cruciate Ligament Ruptures on Magnetic Resonance Imaging. Arthroscopy. 2024;40(4):1197–205.

52. Yoon J, Han J, Ko J, Choi S, Park JI, Hwang JS, et al. Developing and Evaluating an AI-Based Computer-Aided Diagnosis System for Retinal Disease: Diagnostic Study for Central Serous Chorioretinopathy. J Med Internet Res. 2023;25:e48142.

53. Cabitza F. Biases Affecting Human Decision Making in AI-Supported Second Opinion Settings. In 2019. p. 283–94.

54. Chen H, Ma X, Rives H, Serpedin A, Yao P, Rameau A. Trust in Machine Learning Driven Clinical Decision Support Tools Among Otolaryngologists. The Laryngoscope. 2024 Jun;134(6):2799–804.

55. Laxar D, Eitenberger M, Maleczek M, Kaider A, Hammerle FP, Kimberger O. The influence of explainable vs non-explainable clinical decision support systems on rapid triage decisions: a mixed methods study. BMC Med. 2023;21(1):359.

56. Nagendran M, Festor P, Komorowski M, Gordon AC, Faisal AA. Quantifying the impact of AI recommendations with explanations on prescription decision making. Npj Digit Med. 2023 Nov 7;6(1):206.

57. Carmichael J, Costanza E, Blandford A, Struyven R, Keane PA, Balaskas K. Diagnostic decisions of specialist optometrists exposed to ambiguous deep-learning outputs. Sci Rep. 2024 Mar 21;14(1):6775.

58. Gomez C, Smith B, Zayas A, Unberath M, Canares T. Explainable AI decision support improves accuracy during telehealth strep throat screening. Commun Med. 2024;4(1).

59. Rainey C, Villikudathil AT, McConnell J, Hughes C, Bond R, McFadden S. An experimental machine learning study investigating the decision-making process of students and qualified radiographers when interpreting radiographic images. PLOS Digit Health. 2023;2(10):e0000229.

60. de Oliveira A, Azevedo J, Ruback L, Moreira R, Teixeira S, Teles A. Effect of Explainable Artificial Intelligence on Trust of Mental Health Professionals in an AI-Based System for Suicide Prevention. IEEE ACCESS. 2025;13:60987–1005.

61. Gaube S, Suresh H, Raue M, Lermer E, Koch TK, Hudecek MFC, et al. Non-task expert physicians benefit from correct explainable AI advice when reviewing X-rays. Sci Rep. 2023 Jan 25;13(1):1383.

62. Silva A, Schrum M, Hedlund-Botti E, Gopalan N, Gombolay M. Explainable Artificial Intelligence: Evaluating the Objective and Subjective Impacts of xAI on Human-Agent Interaction. Int J Human–Computer Interact. 2023 Apr 21;39(7):1390–404.

63. A. S, R. S. A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. Decis Anal J. 2023 Jun;7:100230.

64. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020 Jun;58:82–115.

65. Senoner J, Schallmoser S, Kratzwald B, Feuerriegel S, Netland T. Explainable AI improves task performance in human-AI collaboration. Sci Rep. 2024;14(1).

66. Prinster D, Mahmood A, Saria S, Jeudy J, Lin CT, Yi PH, et al. Care to Explain? AI Explanation Types Differentially Impact Chest Radiograph Diagnostic Performance and Physician Trust in AI. Radiology. 2024 Nov 1;313(2):e233261.

67. Jin W, Fatehi M, Guo R, Hamarneh G. Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task. Artif Intell Med. 2024 Feb;148:102751.

68. Lancaster Farrell CJ. Explainability does not improve biochemistry staff trust in artificial intelligence-based decision support. Ann Clin Biochem Int J Lab Med. 2022 Nov;59(6):447–9.

69. Jacobs M, Pradier MF, McCoy TH, Perlis RH, Doshi-Velez F, Gajos KZ. How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection. Transl Psychiatry. 2021 Feb 4;11(1):108.

70. Goel K, Sindhgatta R, Kalra S, Goel R, Mutreja P. The effect of machine learning explanations on user trust for automated diagnosis of COVID-19. Comput Biol Med. 2022;146:105587.

71. Micocci M, Borsci S, Thakerar V, Walne S, Manshadi Y, Edridge F, et al. Attitudes towards Trusting Artificial Intelligence Insights and Factors to Prevent the Passive Adherence of GPs: A Pilot Study. J Clin Med. 2021;10(14).

72. Bond RR, Novotny T, Andrsova I, Koc L, Sisakova M, Finlay D, et al. Automation bias in medicine: The influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms. J Electrocardiol. 2018;51(6S):S6–11.

73. Festor P, Nagendran M, Gordon A, Faisal A, Komorowski M. Safety of human-AI cooperative decision-making within intensive care: A physical simulation study. PLOS Digit Health. 2025;4(2).

74. Li J, Li X, Zhang C. Understanding physicians' noncompliance use of AI-aided diagnosis-A mixed-methods approach. Decis SUPPORT Syst. 2025;191.

75. Knoery C, Bond R, Iftikhar A, Rjoob K, McGilligan V, Peace A, et al. SPICED-ACS: Study of the potential impact of a computer-generated ECG diagnostic algorithmic certainty index in STEMI diagnosis: Towards transparent AI. J Electrocardiol. 2019 Nov 1;57.

76. Calisto F, Abrantes J, Santiago C, Nunes N, Nascimento J. Personalized explanations for clinician-AI interaction in breast imaging diagnosis by adapting communication to expertise levels. Int J Hum-Comput Stud. 2025;197.

77. Zhang S, Ding S, Cui W, Li X, Wei J, Wu Y. Evaluating the effectiveness of a clinical decision support system (AI-Antidelirium) to improve Nurses' adherence to delirium guidelines in the intensive care unit. INTENSIVE Crit CARE Nurs. 2025;87.

78. Li J, Zhou L, Zhan Y, Xu H, Zhang C, Shan F, et al. How does the artificial intelligence-based image-assisted technique help physicians in diagnosis of pulmonary adenocarcinoma? A randomized controlled experiment of multicenter physicians in China. J Am Med Inform Assoc. 2022 Nov 14;29(12):2041–9.

79. Tonekaboni S, Joshi S, McCradden MD, Goldenberg A. What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use [Internet]. arXiv; 2019 [cited 2025 Aug 25]. Available from: http://arxiv.org/abs/1905.05134

80. Nazar M, Alam MM, Yafi E, Su'ud MM. A Systematic Review of Human–Computer Interaction and Explainable Artificial Intelligence in Healthcare With Artificial Intelligence Techniques. IEEE Access. 2021;9:153316–48.

81. Giuste F, Shi W, Zhu Y, Naren T, Isgut M, Sha Y, et al. Explainable Artificial Intelligence Methods in Combating Pandemics: A Systematic Review. IEEE Rev Biomed Eng. 2023;16:5–21.

82. John D. Lee, Katrina A. See. Trust in Automation: Designing for Appropriate Reliance. Hum Factors. 2004 Mar 1;46(1):50–80.

83. Benda NC, Novak LL, Reale C, Ancker JS. Trust in AI: why we should be designing for APPROPRIATE reliance. J Am Med Inform Assoc. 2021 Dec 28;29(1):207–12.

84. Spitzer P, Holstein J, Morrison K, Holstein K, Satzger G, Kühl N. Don't be Fooled: The Misinformation Effect of Explanations in Human-AI Collaboration [Internet]. arXiv; 2024 [cited 2024 Sep 27]. Available from: http://arxiv.org/abs/2409.12809

85. Morrison K, Spitzer P, Turri V, Feng M, Kühl N, Perer A. The Impact of Imperfect XAI on Human-AI Decision-Making. Proc ACM Hum-Comput Interact. 2024 Apr 17;8(CSCW1):1–39.

86. Cabitza F, Fregosi C, Campagner A, Natali C. Explanations Considered Harmful: The Impact of Misleading Explanations on Accuracy in Hybrid Human-AI Decision Making. In: Longo L, Lapuschkin S, Seifert C, editors. Explainable Artificial Intelligence [Internet]. Cham: Springer Nature Switzerland; 2024 [cited 2024 Sep 10]. p. 255–69. (Communications in Computer and Information Science; vol. 2156). Available from: https://link.springer.com/10.1007/978-3-031-63803-9_14

87. Alufaisan Y, Marusich LR, Bakdash JZ, Zhou Y, Kantarcioglu M. Does Explainable Artificial Intelligence Improve Human Decision-Making? [Internet]. 2020 [cited 2024 Aug 17]. Available from: https://osf.io/d4r9t

88. Bertrand A, Belloum R, Eagan JR, Maxwell W. How Cognitive Biases Affect XAI-assisted Decision-making: A Systematic Review. In: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society [Internet]. Oxford United Kingdom: ACM; 2022 [cited 2024 Sep 21]. p. 78–91. Available from: https://dl.acm.org/doi/10.1145/3514094.3534164

89. Bauer JM, Michalowski M. Human-centered explainability evaluation in clinical decision-making: a critical review of the literature. J Am Med Inform Assoc. 2025 Sep 1;32(9):1477–84.

90. Chen H, Gomez C, Huang CM, Unberath M. Explainable medical imaging AI needs human-centered design: guidelines and evidence from a systematic review. Npj Digit Med. 2022 Oct 19;5(1):156.

91. Adjekum A, Blasimme A, Vayena E. Elements of Trust in Digital Health Systems: Scoping Review. J Med Internet Res. 2018 Dec 13;20(12):e11254.

Figures:

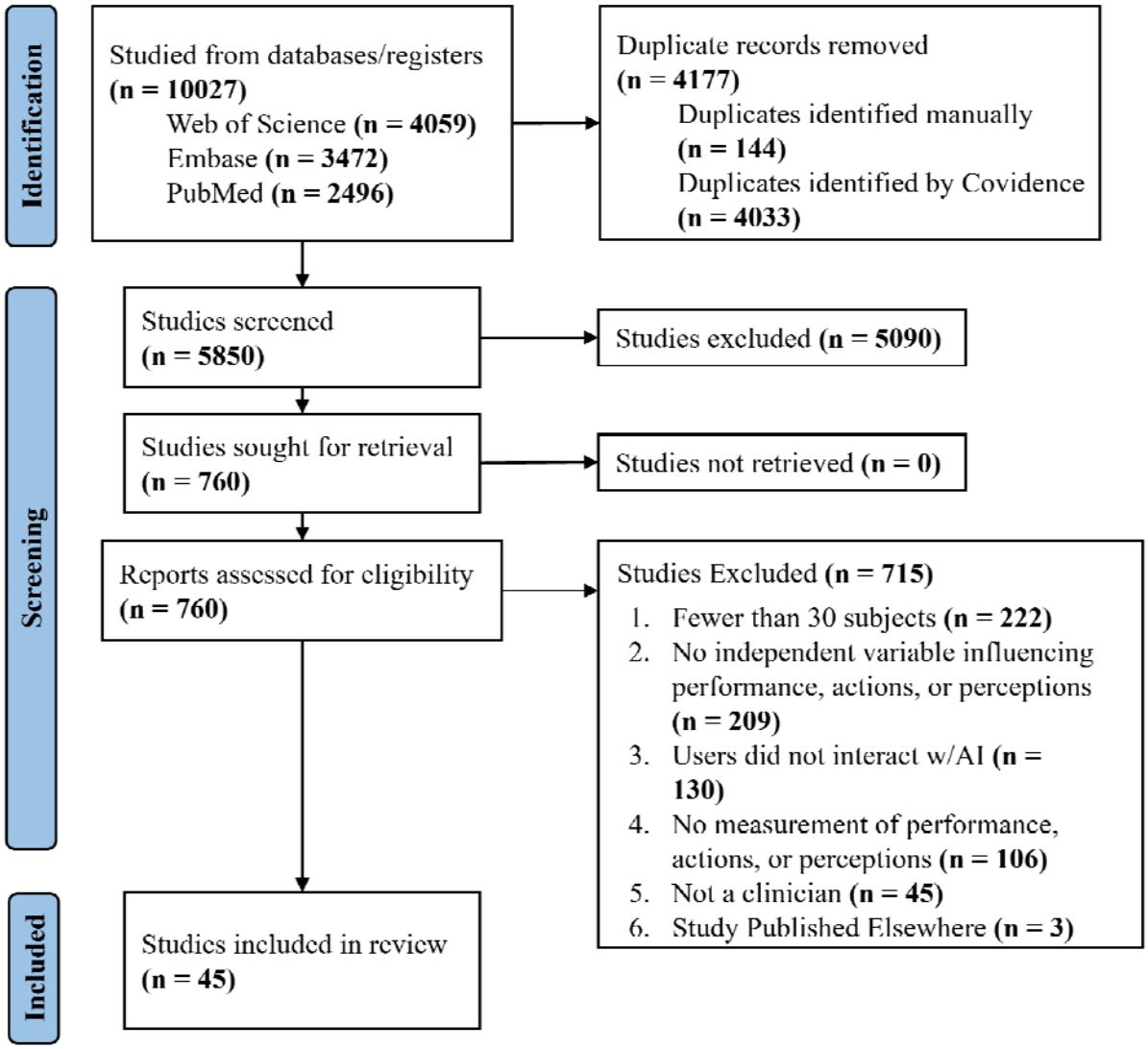

Figure 1: PRISMA diagram for study inclusion.

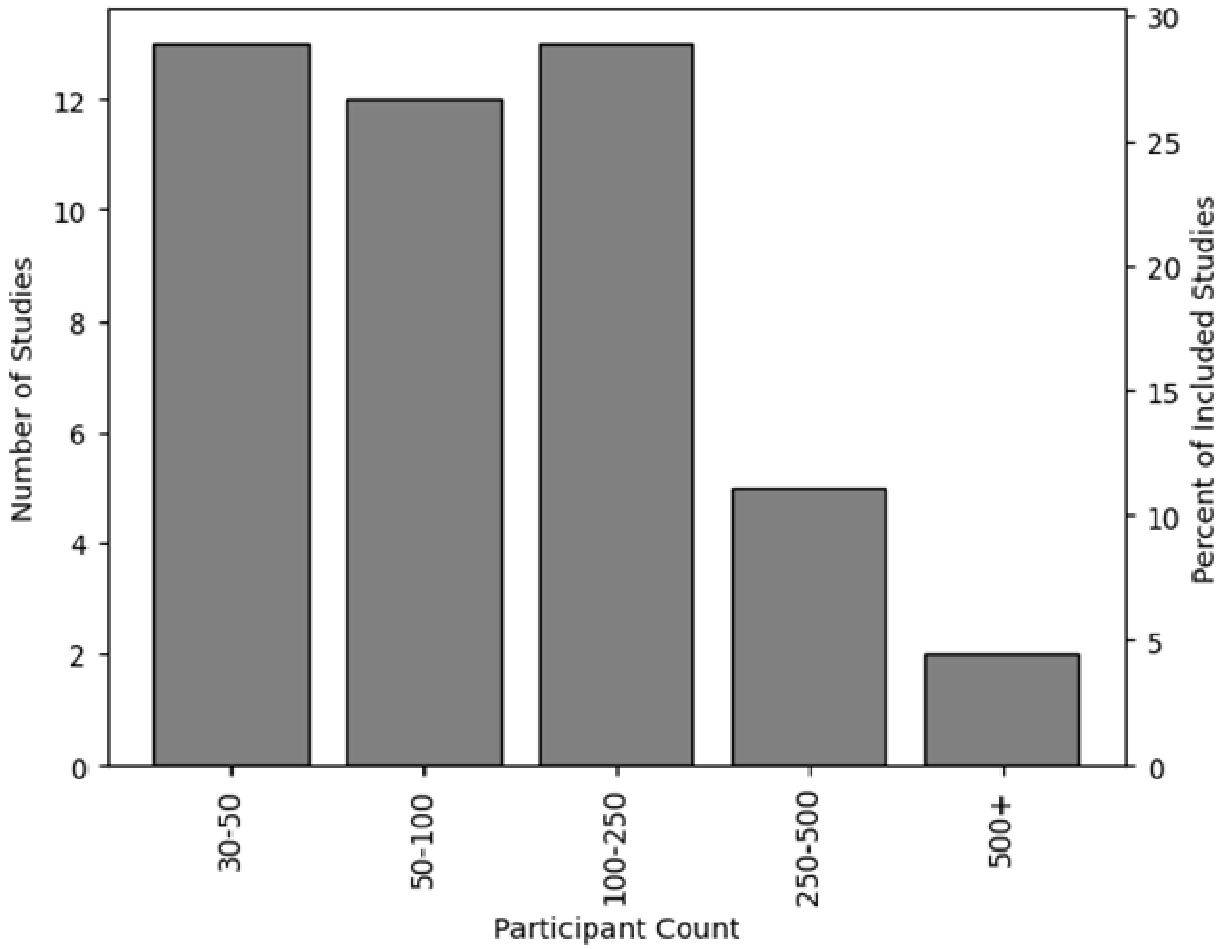

Figure 2: Participant count in the included studies.