Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Trust in AI Clinical Decision Support Systems: A Scoping Review of Measurement and Behavior

Qingyang Li
University of South Carolina
Abdul-Hanan Inusah
University of South Carolina
Yao Tang
University of South Carolina
Xiaoming Li
University of South Carolina
Shan Qiao
shanqiao@mailbox.sc.edu
University of South Carolina

Systematic Review

Keywords: Artificial intelligence, Clinical decision support systems, Trust, Trust measurement, Human–AI interaction, Reliance, Scoping review

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11557716/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Trust in artificial intelligence clinical decision support systems is critical to appropriate reliance, yet its measurement and behavioral implications remain unclear. We conducted a scoping review of quantitative studies published between January 2010 and June 2026, searching PubMed, PsycINFO, Web of Science, Scopus, and Embase. Thirty-three studies met inclusion criteria. Fourteen (42.4%) explicitly defined trust, and 16 (48.5%) assessed anticipated or hypothetical trust without direct system interaction. Twenty-seven studies (81.8%) examined factors associated with trust, emphasizing human–AI interaction and system characteristics over organizational and social contexts. Trust patterns varied across healthcare professionals, patients, and general populations, although direct comparisons were rare. Twenty-seven studies assessed separate AI-related behavioral outcomes, but only 11 formally tested trust–behavior relationships. Reported trust did not consistently correspond to preferences or decisions when AI and clinician recommendations conflicted. More comparable measures and observed behavioral outcomes are needed to evaluate appropriately calibrated trust in clinical AI.

Introduction

As Artificial Intelligence (AI) advances, it becomes increasingly integrated into different aspects of healthcare, including medical imaging & diagnostics, medical research & drug delivery, rehabilitation, virtual patient care, patient engagement & compliance, and administrative applications [1]. Among these, AI-based clinical decision support systems (AI-CDSS) represent the most impactful integrations of AI and healthcare. AI-CDSS are used in diverse clinical scenarios, including diagnosis, risk prediction, and treatment planning [2]. By predicting patient outcomes and recommending potential interventions, these systems may support personalized medicine and improve the efficiency of healthcare professionals' work [3,4].

Despite their potential benefits, AI-CDSS also raise important concerns. Inaccurate chatbot responses, diagnostic errors, and the potential to exacerbate health inequities may undermine confidence in these systems [5,6]. Healthcare professionals may also be unable to fully understand how an AI-CDSS reaches a particular recommendation because of the "black box" nature of many algorithms [7]. In addition, the biased training data can compromise the quality, fairness, and applicability of AI-CDSS generated recommendations.

These concerns can negatively affect users' trust in AI-CDSS. Trust shapes whether such systems are accepted, implemented, interpreted, and relied upon in clinical practice [8]. Without sufficient trust, technically capable AI systems may fail to be adopted or used effectively [9]. At the same time, excessive trust may result in inappropriate reliance on AI recommendations. Trust in AI-CDSS is therefore critical to the safe, effective, and equitable integration of AI into healthcare.

To better understand trust in AI-CDSS, it is important to consider how trust has been conceptualized. However, the definition of trust in healthcare AI is complex. Existing studies variously conceptualize it as trust in technology, a psychological disposition or judgement, or a form of trust situated within the broader healthcare system [10]. Trust in healthcare AI systems exists on a spectrum rather than as a binary concept [11]. Piehl et al. describe trust in health AI as a dynamic negotiation revealing power relations, uncertainties, and institutional dependencies of AI rather than a stable condition [8]. Such conceptual variation complicates comparisons across studies and highlights the need to examine how trust is defined and measured.

Trust in AI-CDSS is not merely a perception of algorithmic accuracy; it is a socially embedded judgment shaped by technology performance, healthcare relationships, institutional legitimacy, and individual experiences. Several reviews have attempted to synthesize existing evidence on trust in AI-CDSS, focusing on how trust is conceptualized and the factors that influence it. M Astobiza et al.[11] identified three categories of factors influencing trust in healthcare AI: technology-related factors, healthcare context factors, and individual user factors [11]. Mertz, et al. [10] explored factors influencing patients' and clinicians' trust in healthcare AI, categorizing them into AI literacy, AI psychology, and AI utility; Tun, et al. [12] synthesized key factors that influence health care workers' trust in AI-CDSS and highlighted the need for explainable AI models, comprehensive training, stakeholder involvement, and human-centered design. Ratta, et al. [13] extended the Unified Theory of Acceptance and Use of Technology (UTAUT) to examine the role of trust and other factors, including performance expectancy, social influence, perceived risk, and self-efficacy, in healthcare professionals' willingness to adopt AI-CDSS; Aquilino, et al. [14] distinguished cognitive trust in an AI-CDSS's reliability and functionality from affective trust that the system acts in users' interests; Piehl, et al. [8] synthesized the conceptualizations of trust in healthcare AI and characterized it as a dynamic negotiation that reveals power relations, uncertainties, and institutional dependencies of AI rather than a stable condition. Collectively, these reviews have explored the conceptualization of trust in healthcare AI and identified a broad range of factors associated with trust. However, quantitative evidence remains fragmented regarding how trust is operationalized and measured, how trust-related factors differ across sociotechnical contexts and population groups, and whether reported trust corresponds to AI-related behavioral outcomes.

Although the importance of trust in AI-CDSS is increasingly recognized, current studies mostly focus on clinicians, individual users, and technology acceptance. Evidence remains fragmented regarding how healthcare providers, patients, and general population as potential users of healthcare services, perceive and act upon trust in AI-CDSS. It is also unclear how trust is shaped through the relationship among AI-CDSS, healthcare providers, patients, and the broader healthcare and social system. This review aims to synthesize this evidence by examining how trust is defined, what socio-technical factors shape it, how it varies across population groups, and how it influences AI-related behavior.

Methods

2.1 Study Design

We conducted a scoping review to examine how trust in AI-CDSS has been conceptualized, measured, and empirically investigated. The review focused on four areas: (1) the conceptualization and measurement of trust, (2) factors associated with trust, (3) trust patterns across healthcare professionals (HCPs),

patients, and the general population, and (4) AI-related behavioral outcomes and trust–behavior relationships. The review process and study selection were reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR).

2.2 Literature Sources and Search Strategy

A comprehensive literature search was conducted in five databases, including PubMed, PsycINFO, Web of Science, Scopus, and Embase for studies published between January 1, 2010 and June 24, 2026.

The search strategy combined three groups of terms: AI-CDSS, trust-related constructs, and quantitative measurement. AI-related terms included clinical decision support system, CDSS, AI-assisted diagnosis, artificial intelligence, machine learning, deep learning, computer-aided diagnosis, and related terms. Trust-related terms included trust, distrust, mistrust, trustworthiness, reliance, trust calibration, overtrust, undertrust, confidence in AI, and acceptance. Measurement-related terms included scale, questionnaire, instrument, measurement, survey, psychometric, reliability, validity, Likert, and factor analysis. Search syntax was adapted to the requirements of each database. The complete database-specific search strategies are provided in Supplementary Information. Reference lists of included studies and relevant existing reviews were also examined to identify additional potentially eligible studies.

2.3 Eligibility Criteria

Studies were eligible if they examined an AI-CDSS, including applications such as imaging or pathology diagnosis, clinical prediction, or AI-supported diagnostic or treatment recommendations. Studies were required to quantitatively assess trust, reliance, trustworthiness, or confidence in the AI system, its outputs, or AI-assisted decision-making using a questionnaire, scale, or structured rating instrument.

The trust measure also had to be sufficiently identifiable from the article, such as through the scale name, measurement items, number of items, dimensional structure, or scoring method. Studies relying solely on a single-item trust measure were excluded. Eligible publications included original quantitative or mixed-methods empirical studies and methodological studies focused on instrument development or validation. Studies were limited to English-language publications from January 1, 2010 through June 24, 2026.

Studies were excluded if they focused on AI applications without an identifiable diagnostic or clinical decision-support function; used only qualitative methods or behavioral indicators without an eligible quantitative trust measure; did not provide sufficient information to identify the trust instrument; or were conference abstracts without full text, editorials, commentaries, letters, case reports, narrative reviews, systematic reviews, or scoping reviews. When multiple publications used the same dataset, the publication providing the most complete information relevant to the review was retained.

2.4 Study Selection and Screening

All retrieved records were imported into Covidence for deduplication and screening. Two reviewers (Q.L. and A.I.) independently screened titles and abstracts according to the prespecified eligibility criteria. Records considered potentially eligible, or for which eligibility could not be determined from the title and abstract alone, proceeded to full-text review. Disagreements were resolved through discussion and consensus; when necessary, a third reviewer (S.Q.) was consulted.

The same two reviewers independently assessed potentially eligible articles at the full-text stage using the prespecified eligibility criteria. Reasons for exclusion were documented for articles excluded after full-text review, and disagreements were resolved using the same consensus process. Studies identified through reference checking of included studies and relevant existing reviews were assessed using the same eligibility criteria. The study selection process is presented in the PRISMA flow diagram in Fig. 1.

2.5 Data Extraction

Data extraction was conducted using a standardized data-charting form developed specifically for this review. Basic study characteristics included publication year, geographic location, study design, population and sample size, participant characteristics, AI application, and clinical context. For trust conceptualization and measurement, extracted information included the trust definition, trust target or object, theoretical framework, trust dimensions, instrument source or development, trust measurement, measurement context or type, and reported reliability or validity. For factors associated with trust, extracted factors were organized into five non-mutually exclusive domains: AI/system-level factors, task or clinical-context factors, human–AI interaction or relational factors, organizational or implementation factors, and social factors. Individual-level characteristics were examined separately as an additional category of trust-related factors. We also extracted the subgroups or comparison groups examined and any reported subgroup differences or moderation findings. For the cross-population synthesis, studies were grouped by population type to examine trust patterns primarily among HCPs, patients or patient-facing populations, and members of the general population, while considering differences in the variables, clinical contexts, and study designs emphasized across these groups. Mixed or other population studies were retained in the overall synthesis and considered where they provided relevant cross-population evidence. For AI-related behavioral outcomes, we extracted independent AI-related behavioral outcomes, behavior type, and reported trust–behavior relationships or findings.

Results

3.1 Study Selection and Characteristics of Included Studies

A total of 2,769 records were retrieved from five databases, including PubMed (n = 542), Scopus (n = 603), Web of Science (n = 1,061), Embase (n = 518), and PsycINFO (n = 45). After removing 1,159 duplicate records, 1,610 records remained for title and abstract screening. Of these, 1,531 records were excluded, leaving 79 articles for full-text eligibility assessment.

Following full-text screening, 56 articles were excluded for reasons highlighted in Figure 1: use of a single-item trust measure (n = 26), no eligible quantitative trust measure (n = 18), irrelevant context (n = 6), unclear or unidentifiable trust instrument (n = 3), and ineligible publication type (n = 3). Twenty-three studies met the eligibility criteria following full-text review. An additional 10 eligible studies were identified through reference checking of included studies and existing literature reviews. A total of 33 studies were included in the study (**Figure 1**).

The basic characteristics of the included studies were presented in **Table 1**. These studies were published between 2019 and 2026, with 28 (84.8%) published from 2022 onward and 14 (42.4%) published in 2025 or 2026. Geographically, thirteen studies (39.4%) were conducted in Asia, eight (24.2%) in Europe, six (18.2%) in North America, and one (3.0%) in Africa; four studies (12.1%) recruited participants across multiple countries or regions, and one (3.0%) did not clearly report the recruitment location. Sample sizes ranged from 20 to 13,806 participants. Among single-country studies, China was the most frequently represented (n=6, 18.2%), followed by the United States (n=5, 15.2%), India, Denmark, and Germany (n=2 each, 6.1%).

Healthcare professionals constituted the largest population group (17/33, 51.5%), followed by patients (7/33, 21.2%), mixed or other populations (5/33, 15.2%), and members of the general population (4/33, 12.1%). Twenty studies (60.6%) used primarily cross-sectional or survey-based observational designs, 10 (30.3%) used experimental or controlled user-study designs, and three (9.1%) used prospective or time-lagged designs.

The AI applications and clinical contexts also varied considerably. Most studies (25/33, 75.8%) examined AI-CDSS within a defined clinical specialty or specific context, whereas a smaller proportion (8/33, 24.2%) examined AI-CDSS more broadly, without restriction to a single specialty or narrowly defined clinical context. Among clinical specialties or specific contexts, the most frequently represented were ophthalmology/optometry (3/33, 9.1%) and psychiatry or behavioral health (n=3, 9.1%), followed by critical care or intensive care, radiology or medical imaging, and cancer-related care in oncology or urology (n=2 each, 6.1%). Other contexts were each represented by one study and included digital self-diagnosis and triage, otolaryngology and rhinology, antimicrobial stewardship, telehealth and urgent care, pathology and cervical cytology, clinical neurophysiology, hand surgery, emergency medicine, dentistry, blood transfusion, internal medicine, and gastroenterology. One additional study examined AI-CDSS across multiple specialty areas, including radiology, robotic surgery, and dermatology.

3.2 Conceptualization and Measurement of Trust in AI-CDSS

Trust was conceptualized and operationalized heterogeneously across the 33 included studies (**Table 2**). Fourteen studies (42.4%) provided an explicit conceptual or operational definition of trust, while 22 (66.7%) used an identifiable theoretical or conceptual framework. A dimensional or structural characterization of trust was identified in 23 studies (69.7%), and 24 (72.7%) reported some form of reliability or validity evidence relevant to the trust measure or its development.

3.2.1 Conceptualization and Targets of Trust

Among studies that explicitly defined trust, recurring elements included expectations of AI reliability or competence, vulnerability under conditions of uncertainty, and willingness to rely on AI-generated recommendations or decisions. Theoretical approaches were similarly diverse, spanning technology-adoption models and human–automation or human–computer trust frameworks. Trust was also inconsistently distinguished from related constructs such as confidence, reliability, belief, reliance, acceptance, and distrust. For example, Perfalk, et al. [15] and Perfalk, et al. [16] treated trust and distrust as distinct constructs, whereas other studies incorporated reliability, competence, or reliance directly into trust measurement.

Importantly, the target or object of trust differed substantially across studies. Some assessed trust broadly in AI-CDSS or AI-assisted diagnosis and treatment [17-21], whereas others focused on a specific AI-CDSS [22-29]. Trust was also directed toward particular outputs, including diagnostic results, treatment recommendations, clinical advice, and image interpretation [30-34](Arango et al., 2025; Asokan et al., 2023; Maleki et al., 2026; Radeva et al., 2026; Yang et al., 2019). Other studies located trust within a broader human–AI or sociotechnical configuration, including physician-supervised versus autonomous AI [35], AI versus human decision-makers [36], delegation of diagnostic responsibility [37], and trust in developers, data governance, or institutional handling of health information [13,38,39].

3.2.2 Assessment Context

The assessment context in which trust was measured also varied substantially. Sixteen studies (48.5%) primarily assessed anticipatory, hypothetical, or scenario-based trust without direct interaction with the study AI system, while eight (24.2%) assessed general or experience-informed trust without a standardized study interaction. Seven studies (21.2%) measured post-interaction, task-specific trust after participants viewed, simulated, or directly interacted with a particular AI system [23,25-29,40]. One study (3.0%) examined post-failure, experience-informed trust, referring to trust assessed after participants had prior experience with a medical AI service failure or error [41], and one study (3.0%) assessed trust both before and after AI-assisted interaction [37]. Overall, most quantitative evidence reflected anticipated or generalized trust rather than trust formed through a standardized direct interaction with AI.

3.2.3 Trust Measures and Psychometric Properties

Trust measurement was varied, and no single instrument predominated. Studies used established or adapted human–automation and technology-trust instruments, newly developed multi-item measures, and task- or context-specific questionnaires. Although a discernible dimensional or structural characterization of trust was identified in 23 studies (69.7%), most studies still treated trust as a single overall or latent construct rather than modeling multiple distinct trust dimensions. Psychometric reporting was also inconsistent. Twenty-four studies (72.7%) provided some form of reliability or validity evidence, most commonly internal consistency, factor-analytic evidence, or convergent and discriminant validity. However, the scope of evaluation varied

substantially, and in some studies psychometric evidence applied to the broader questionnaire rather than specifically to the trust measure. These differences in measurement approach and psychometric evaluation contributed to limited comparability of trust estimates across studies.

3.3 Factors Associated with Trust in AI-CDSS

Factors associated with trust were examined in 27 of the 33 studies (81.8%) and were organized into five non-mutually exclusive domains: AI/system-level (17/33, 51.5%), task or clinical-context (12/33, 36.4%), human–AI interaction or relational (19/33, 57.6%), organizational or implementation (7/33, 21.2%), and social factors (6/33, 18.2%). Individual-level characteristics were examined separately as an additional category of trust-related factors. Several studies also examined whether associations between these factors and trust varied across participant characteristics. Identified factors and findings are summarized in **Table 3**.

AI/system-level factors. Perceived AI performance, reliability, competence, usefulness, security, and risk were among the most commonly examined system-related factors. Greater performance expectancy was positively associated with trust in Cheng, et al. [18] while overall expectancy toward the Blood Utilization Calculator was positively associated with trust in Choudhury, et al. [22]. Perceived security was strongly positively associated with trust in Calisto, et al. [17], whereas perceived risk was negatively associated with trust in Choudhury, et al. [22] and Calisto, et al. [17]; in contrast, perceived risk was not significantly associated with trust in Ghozali, et al. [24]. System configuration also mattered in experimental settings. In Ghaffar, et al. [23], optometrists reported significantly greater trust in a multimodal than a unimodal AI system, while Yokoi et al [36] found lower trust in an AI medical decision-maker than in a human physician under otherwise comparable conditions. Other studies descriptively identified accuracy, reliability, transparency, reproducibility, validation, and technical limitations as considerations relevant to trustworthiness [31,37,38].

Task and clinical-context factors. Trust varied across clinical tasks, disease contexts, specialties, and different levels of AI responsibility. In Woodcock, et al. [40], trust-related responses differed by disease context, with the effects of explanation type varying between migraine and temporal arteritis scenarios. Yakar, et al. [42] found significant differences across specialties, with trust highest for dermatology, followed by radiology and robotic surgery. In Rodler, et al. [35], trust became more cautious as clinical situations were perceived as more severe or complex, while Zhang, et al. [37] found greater acceptance of AI for preliminary review and highlighting suspicious regions than for independently making final diagnoses. Dunker, et al. [38] similarly reported greater caution when AI-supported decisions could reduce subsequent specialist involvement. Several studies also assessed trust separately across clinical functions without formally testing whether the differences were statistically significant. For example, Radeva, et al. [33] reported low trust in both AI-based glaucoma diagnosis and treatment recommendations (7.5% and 5.7%, respectively), while Arango et al [30], Maleki, et al. [32], and Yang, et al. [34] assessed trust separately across diagnosis, treatment, complication management, or other clinical functions. Yokoi et al. [36] found that whether the AI or physician recommended the participant's preferred versus non-preferred treatment did not significantly affect trust. Across studies, trust appeared to vary with the nature and stakes of the clinical task and with the degree of decision authority assigned to AI.

Human–AI interaction and relational factors. Explainability, information relevance, interactivity, human oversight, and alignment between AI support and users' clinical reasoning were commonly examined trust-related features. Physician oversight showed a particularly strong relationship with trust in Rodler, et al. [35]: trust in an AI-generated diagnosis was substantially higher when the output was verified by a physician than when AI operated autonomously. In Katzburg, et al. [27], higher perceived information relevance and interactivity were associated with greater trust, although the relationship between interactivity and trust varied by clinical experience. Liu et al. reported a positive association between perceived AI explainability and trust [20]. In Zhang, et al. [29], an emotion-adaptive, personalized CDSS received higher trustworthiness and reliability ratings than a basic static system. Explainability did not consistently increase trust, however. In Goel, et al. [25], trust differed descriptively across strong-, weak-, and no-explanation conditions, but the overall difference was not statistically significant; Woodcock et al. [40] and Gomez, et al. [26] likewise showed that trust depended on the type and context of explanations rather than their mere presence. Other studies descriptively emphasized concordance with clinicians' own reasoning or trusted expert judgment, preservation of human involvement, and appropriate reliance as important relational conditions for trust [23,31,37].

Organizational and implementation factors. Organizational and implementation factors were less frequently examined and included facilitating conditions, information provision, external validation, workflow integration, accountability, training, and clinical implementation safeguards. In Ghozali, et al. [24], facilitating conditions, including access to technical support, training, infrastructure, and organizational resources, were positively associated with trust in technology. Information provision was experimentally examined in Perfalk, et al. [16] and Perfalk, et al. [15]. In both randomized survey experiments, participants received brief, neutral information explaining how ML-based CDSS work and support clinical decision-making, and those receiving this information reported greater trust than participants in the control conditions. Other studies identified external validation and expert evaluation [31], clear responsibility and accountability, clinician AI literacy, workflow integration, and safeguards against overreliance or deskilling [38], and preservation of clinician oversight within existing workflows [35,37] as implementation conditions relevant to trust.

Social factors. Trust was also associated with interpersonal and institutional influences surrounding AI use. Social influence reflected perceived expectations or endorsement from colleagues, supervisors or leaders, patients, and the broader professional environment. Social influence was positively associated with trust in Tran, et al. [21], Cheng, et al. [18], [19], and [24]. In Hameed, et al. [19], perceived expectations from leaders and patients, observation that other clinicians were using AI, and the perception that AI use was becoming a professional norm showed the strongest direct association with initial trust. In Ghozali, et al. [24], colleague expectations, managerial support, professional recommendations, and perceived professional norms were positively associated with trust in technology. Broader social and institutional sources of trust were also reported. In Asokan, et al. [31], agreement with trusted specialists and expert endorsement were identified as relevant to perceived trustworthiness, while Dunker, et al. [38] reported greater trust in public healthcare and research institutions than in private or commercial actors, as well as greater trust in AI systems associated with domestic institutions and data sources. Collectively,

these studies indicate that trust in healthcare AI was situated not only in perceptions of the technology itself but also within professional norms, expert endorsement, and institutional relationships.

Individual-level factors. Individual-level factors examined in relation to trust included AI knowledge and technology familiarity, general attitudes toward AI, and psychological or attitudinal characteristics. More favorable general attitudes toward AI were positively associated with trust in AI-generated findings among gastroenterology clinicians [43]. Greater AI knowledge or prior awareness was associated with more favorable trust in several studies [32,34,44]. Technology-related characteristics showed inconsistent associations with trust. Rodler et al. [35] found a weak positive association between technology affinity and general trust in AI, whereas Arango et al. [30] reported that greater technology familiarity was associated with lower trust in AI-generated treatment recommendations. Psychological factors were less frequently examined. For example, Huo et al. [41] found that greater self-responsibility attribution was associated with higher levels of human–computer trust. Other studies examined general trust disposition and related individual characteristics, but findings were less consistent across studies.

Moderating effects. Four studies (12.1%) examined whether associations between trust-related factors or interventions and trust varied across participant characteristics. In Katzburg, et al. [27], clinical experience moderated the relationship between interface interactivity and trust, with the positive association becoming weaker among more experienced physicians. In Calisto, et al. [17], the association between perceived clinical workflow impact and trust was significant among older and expert clinicians but not among younger or novice clinicians. Stratified analyses in the two randomized ML-CDSS information studies [15,16] also suggested some heterogeneity in intervention effects. Among patients [16], increases in trust were more evident among women and those with affective or anxiety disorders, whereas among psychiatric staff [15], increases were more evident among older participants and staff in professions requiring bachelor-level education or less.

3.4 Trust Pattern across HCP, Patients and General Population

Among HCP studies (17/33), trust was generally not uniformly high, but instead appeared moderate, cautious, and conditional. HCP trust often depended on system performance, task characteristics, explanations, and human oversight. For example, Radeva et al. [33] reported that only 7.5% of ophthalmologists reported full trust in AI diagnosis, while 5.7% expressed full trust in AI-generated treatment recommendations. In Goel, et al. [25], direct trust did not appear particularly high or stable across explanation conditions. In Ghaffar, et al. [23], however, optometrists reported significantly greater trust in a multimodal AI system than in a unimodal system. These findings suggest that HCP trust may be especially sensitive to system capability and the clinical configuration in which AI is used.

Among patient or patient-facing studies (7/33), stated trust could be relatively favorable, but trust in autonomous AI decision-making appeared more cautious. In Yang, et al. [34], 90.0% of cancer patients reported at least some trust in an independent AI diagnosis and 85.1% reported some trust in AI-generated treatment advice; however, only 11.2% preferred AI when its diagnosis conflicted with that of a physician, and only 8.7% chose AI when treatment recommendations disagreed [34]. In Rodler, et al. [35], trust in physician-verified AI was substantially higher than trust in autonomous AI. These findings suggest that patients may express trust in AI capability while still preferring physician involvement when AI is used for consequential clinical decisions.

General-population studies (4/33) appeared to show more moderate or conditional trust, and most of this evidence reflected anticipatory or hypothetical trust. In Yokoi et al [36], trust in an AI medical decision-maker was lower than trust in a human physician. Yakar et al. [42] reported moderate mean trust levels across three clinical AI applications, including dermatology, radiology, and robotic surgery. In Woodcock, et al. [40], trust varied according to disease context and explanation condition rather than remaining stable across scenarios. Because these studies were largely based on hypothetical scenarios or broad attitudes toward AI, the trust measured in this population may be better understood as expected or anticipatory trust rather than trust formed through direct clinical use.

Overall, HCP trust appeared more contingent on technical performance and clinical integration, patient trust appeared particularly sensitive to physician involvement and AI autonomy, and general-population trust was more often anticipatory and shaped by broader attitudes and hypothetical scenarios. These patterns primarily reflect comparisons across separate studies rather than direct comparisons between population groups. Direct comparisons between HCPs, patients, and other participant groups were rare. In Dunker, et al. [38], scenario-based trust did not differ significantly overall among physicians, neurophysiology technicians, and patients, although some pairwise patterns suggested greater trust among technicians than patients.

The populations also differed in the variables and study contexts emphasized. HCP studies more often examined system performance and reliability, workflow impact, explainability, interface and interaction design, human oversight, clinical experience, and professional role, often in relation to clinical tasks or AI-supported decision-making. Patient-focused studies more often examined AI autonomy, physician involvement, and preferences when AI and physician recommendations conflicted. General-population studies more often examined broader attitudes toward AI, perceived capability, social influence, explanation effects, and willingness to use AI, commonly through hypothetical or scenario-based assessments rather than actual clinical use.

These differences in research emphasis and study context limit direct comparison of trust levels across populations. Observed differences may therefore reflect not only population characteristics but also differences in the AI applications, clinical scenarios, trust measures, and study designs used across studies.

3.5 Trust and AI application-Related Behavioral Outcomes

Independent AI-related behavioral outcomes were assessed in 27 of the 33 studies (81.8%) (**Table 4**). These outcomes were defined as AI-related intentions, preferences, acceptance, choices, or actions measured separately from the study's trust construct. Behavior-like items embedded within a trust scale, such

as willingness to rely on or follow AI advice, were treated as part of trust measurement rather than as independent behavioral outcomes. Most outcomes were self-reported intentions, willingness, preferences, acceptance, or recommendation intentions rather than directly observed AI use. These included intentions to adopt or use AI [18-22,24,39], acceptance of AI-assisted or autonomous decision-making [41,43], and preferences for AI versus human decision-making [32,34,35]. Several studies also assessed more behaviorally proximal outcomes, including agreement with AI recommendations, diagnostic decisions or performance during AI-supported tasks, and confirmatory-testing decisions [23,26,27,37]. For example, Curiello et al. assessed clinicians' intentions to use AI tools in clinical practice but did not formally test the association between trust and behavioral intention [45].

Among the 27 studies assessing independent AI-related behavioral outcomes, 11 (40.7%) formally tested the relationship between trust and behavioral outcomes. Direct positive associations between trust and behavioral intention or use were reported in Ghozali, et al. [24], Liu, et al. [20], and Hameed, et al. [19], and direct trust–behavior pathways were also identified within the mediation models of Cheng, et al. [18] and Choudhury, et al. [22]. In contrast, Tran, et al. [21] found no significant association between initial trust and behavioral intention. Trust also functioned as a mediator in pathways leading to adoption, acceptance, or use [13,18,22,41,43]. Moderation effects were identified in Huo et al [41] and Ye, et al. [39]. Awad, et al. [46] reported a positive association between trust and AI-CDSS reliance. Overall, trust was associated with AI-related behavioral outcomes through direct, mediated, moderated, or nonsignificant pathways.

Self-reported trust did not always align with behavioral preferences or decisions, particularly when AI recommendations conflicted with those of healthcare professionals. In Yang, et al. [34], 90.0% of cancer patients reported at least some trust in an independent AI diagnosis, yet only 11.2% preferred AI when its diagnosis conflicted with that of a physician. Similarly, 85.1% reported some trust in AI-generated treatment advice, but only 8.7% chose AI when treatment recommendations disagreed. Comparable patterns were observed in Maleki, et al. [32] and Rodler, et al. [35], where relatively favorable trust in AI coexisted with a strong preference for physician involvement or oversight. These findings suggest that expressed trust in AI capability does not necessarily translate into a preference to defer to AI when it conflicts with human clinical judgment. Gomez et al. found that participants showed greater agreement with diagnostic advice from all three AI-based CDSS conditions than with the Centor Score, despite reporting lower self-reported trust in the Performance Characteristics and Explain by Feature Contribution conditions [26].

Discussion

This scoping review synthesized quantitative evidence on trust in AI-CDSS across HCPs, patients, and the general population. Trust was defined and measured inconsistently and was usually assessed without direct use of the AI system in question. It was conditional rather than uniformly high or low, depending on system performance, clinical task, and human involvement, whereas organizational and social influences received comparatively little attention. Trust patterns appeared to differ across populations, although these groups were rarely compared directly, and reported trust did not always translate into willingness to rely on AI, particularly when AI and clinician recommendations conflicted.

The first knowledge gap identified by the scoping review is inconsistency in how trust was defined and measured emphasizes the conceptual fragmentation described in prior reviews of trust in healthcare AI [8,10]. Our findings add a practical consequence: because the object of trust also varied, studies may share the label "trust" while measuring different constructs directed at different targets, so that trust in "AI" reported in a population survey and trust in a specific AI-CDSS system's recommendation are not directly comparable. Trust measured without direct use may reflect general attitudes, expectations, and context rather than experience with a specific AI system, corresponding to dispositional and situational influences rather than the learned trust that develops through interaction [47]. It therefore cannot show whether trust tracks a system's actual performance. The scarcity of post-failure evidence is a particular gap, because confidence in algorithms may decline more sharply than confidence in human advisors after the same error [48].

The second knowledge gap is a limited investigation regarding the factors associated with trust beyond a technology perspective. Trust was shaped by a broad range of factors, but research attention was unevenly distributed. Most studies focused on the AI-CDSS system and the human–AI interaction, such as performance, explainability, and human oversight, whereas organizational and social contexts in which the system is used were examined far less often. The limited attention to organizational and social conditions matters, because social influence, where measured, was consistently associated with trust, suggesting that clinicians' trust may depend partly on professional norms and endorsement, not only on the AI-CDSS system itself. Even explainability, a commonly examined feature, did not reliably increase trust, consistent with arguments that explaining a model's logic may not calibrate trust when individual predictions vary in reliability [49].

The third knowledge gap is a lack of direct comparation and dialogues about trust toward a same AI-CDSS across different key stakeholders. Although differences across populations partly reflect how each group was studied, they may also reflect the different positions these groups occupy in the healthcare practice. Clinicians evaluate AI-CDSS as a tool they use, whereas patients experience AI-CDSS as something used in their care, which may explain why clinicians' trust centered on performance and patients' trust on physician involvement. This is consistent with evidence that AI is trusted less than physicians partly because it is perceived as less caring and less similar in values [36]. These patterns suggest that populations may differ not only in how much they trust AI-CDSS, but also in what their trust is anchored to.

The fourth knowledge gap is a mixed result regarding the relationship between trust and AI-related behaviors. The mismatch between reported trust and AI-related behavior ran in both directions: favorable trust coexisted with a preference for physician judgment [34], while lower reported trust coexisted with agreement with AI advice [26]. Because most behavioral outcomes were self-reported intentions, the link between trust and actual use remains largely untested. Self-reported trust may therefore be an imperfect proxy for reliance, consistent with foundational work treating trust as an attitude that guides reliance rather than reliance itself, with appropriate reliance as the central goal [50]. For AI-CDSS, the aim should be calibrated trust, matched to the AI-CDSS system's purpose and performance in context, rather than simply more trust [51].

The final knowledge gap is the methodology limitations. Most included studies were cross-sectional, which limits causal inference. Some empirical studies were not guided by a solid conceptual framework. Not many trust measurements used in the studies were strictly validated with a complete psychometric report. In addition, the evidence was concentrated among HCPs and in studies from Asia, Europe, or North America, with few general-population studies and only one from Africa. Although it may reflect the disparities of AI research infrastructure in the globe, findings may therefore not generalize to the public or to health systems where institutional trust and AI governance differ.

Future studies should address these knowledge gaps. Researchers should define trust explicitly, specify what is being trusted, and examine organizational and social influences more systematically. Longitudinal designs that follow users through repeated real-world use, including after errors, are needed to show how trust develops, erodes, and recovers. Studies should include more patients, members of the public, and underrepresented regions, and compare populations directly by having different groups evaluate the same AI system with the same measure. Finally, studies should pair trust measures with observed behaviors, such as acceptance of correct versus incorrect AI advice, which would allow trust calibration to be assessed directly.

This review has several limitations. Restricting inclusion to identifiable multi-item trust measures enabled analysis of how trust was conceptualized and measured but excluded single-item trust measures and studies that assessed only behavioral indicators or qualitative methods. The English-language restriction may also have contributed to geographic imbalance.

Despite these limitations, our findings suggest that successful implementation of AI-CDSS should focus on appropriate use rather than simply increasing trust. Healthcare organizations, healthcare system regulation, and AI governance should ensure that AI-CDSS is introduced with clear clinical oversight, transparent roles, and communication that helps clinicians and patients understand when AI system should inform decisions and when human judgment should remain central. Because patient trust and preferences were sensitive to physician involvement, implementation strategies that position AI as a support for clinical decision-making may be more acceptable than approaches perceived as replacing clinicians. Trust should also be evaluated across population groups, since systematic differences in trust or reliance could translate into unequal uptake and potentially widen existing health inequities. Overall, trust should be treated as an implementation outcome that needs to be calibrated, supported, and evaluated alongside safety, effectiveness, and equity. Future research should prioritize clearer definitions, more comparable trust measures, and longitudinal or real-world studies examining how trust develops and influences clinical decision-making over time. Such evidence will be essential for supporting the safe, effective, and equitable implementation of AI-CDSS in healthcare.

Declarations

Competing Interests

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

Author Contributions

S.Q., Q.L., and A.I. conceived the study. Q.L. and A.I. jointly contributed to the study methodology, literature screening, data extraction, and data synthesis. Q.L. drafted the Methods and Results sections, while A.I. drafted the Discussion section. Y.T. drafted and revised the Introduction section. S.Q. supervised the study, contributed to the methodology, resolved screening disagreements when necessary, and critically reviewed and revised the manuscript. All authors critically reviewed the manuscript for important intellectual content, provided feedback, and approved the final version for publication.

Data Availability

All data extracted and analyzed in this scoping review were obtained from previously published studies. The data supporting the findings are provided in the main article and its Supplementary Information.

Acknowledgements / Funding

This research received no specific funding from any funding agency in the public, commercial, or not-for-profit sectors. The authors thank Rajon Banik for his assistance with reference organization.

References

1. Al Kuwaiti A et al (2023) A Review of the Role of Artificial Intelligence in Healthcare. J Pers Med 13. https://doi.org/10.3390/jpm13060951
2. Laxar D et al (2023) The influence of explainable vs non-explainable clinical decision support systems on rapid triage decisions: a mixed methods study. BMC Med 21:359. https://doi.org/10.1186/s12916-023-03068-2
3. Rezaeian O, Bayrak AE, Asan O (2024) An Architecture to Support Graduated Levels of Trust for Cancer Diagnosis with AI. Comm Com Inf Sc 2119:344–351. https://doi.org/10.1007/978-3-031-61966-3_37
4. Stacy J et al (2022) Qualitative Evaluation of an Artificial Intelligence-Based Clinical Decision Support System to Guide Rhythm Management of Atrial Fibrillation: Survey Study. Jmir Form Res 6. https://doi.org/10.2196/36443
5. Kaplan AD, Kessler TT, Brill JC, Hancock PA (2023) Trust in Artificial Intelligence: Meta-Analytic Findings. Hum Factors 65:337–359. https://doi.org/10.1177/00187208211013

6. Kim DKD, Kim S (2022) What if you have a humanoid AI robot doctor? An investigation of public trust in South Korea. J Commun Healthc 15:276–285. https://doi.org/10.1080/17538068.2021.1994825
7. York TJ, Raj S, Ashdown T, Jones G (2023) Clinician and computer: a study on doctors' perceptions of artificial intelligence in skeletal radiography. Bmc Med Educ 23. https://doi.org/10.1186/s12909-022-03976-6
8. Piehl HS, Janssen R, Penders B, Fijten RR (2026) R. Between map and maze: reframing trust in healthcare AI. AI Soc 41:6925–6935. https://doi.org/10.1007/s00146-026-02997-9
9. Steerling E, Siira E, Nilsen P, Svedberg P, Nygren J (2023) Implementing AI in healthcare-the relevance of trust: a scoping review. Front Health Serv 3. https://doi.org/10.3389/frhs.2023.1211150
10. Mertz M et al (2025) Exploring trust factors in AI-healthcare integration: a rapid review. Front Artif Intell 8:1658510. https://doi.org/10.3389/frai.2025.1658510
11. Astobiza M, Alonso A, M., Ortega Lozano R (2025) Trust and AI in healthcare: a systematic review. Monash Bioeth Rev. https://doi.org/10.1007/s40592-025-00272-z
12. Tun HM, Rahman HA, Naing L, Malik OA (2025) Trust in Artificial Intelligence-Based Clinical Decision Support Systems Among Health Care Workers: Systematic Review. J Med Internet Res 27:e69678. https://doi.org/10.2196/69678
13. Ratta R, Sodhi J, Saxana U (2025) The Relevance of Trust in the Implementation of AI-Driven Clinical Decision Support Systems by Healthcare Professionals: An Extended UTAUT Model. Electron J Knowl Manage 23:47–66. https://doi.org/10.34190/ejkm.23.1.3499
14. Aquilino L et al (2025) Decoding Trust in Artificial Intelligence: A Systematic Review of Quantitative Measures and Related Variables. *Informatics-Basel* 12 https://doi.org/10.3390/informatics12030070
15. Perfalk E, Bernstorff M, Danielsen AA, Ostergaard SD (2025) Receiving Information on Machine Learning-Based Clinical Decision Support Systems in Psychiatric Services Increases Staff Trust in These Systems: A Randomized Survey Experiment. Acta Psychiatr Scand 152:39–48. https://doi.org/10.1111/acps.13791
16. Perfalk E, Bernstorff M, Danielsen AA, Ostergaard SD (2024) Patient trust in the use of machine learning-based clinical decision support systems in psychiatric services: A randomized survey experiment. Eur Psychiatry 67:e72. https://doi.org/10.1192/j.eurpsy.2024.1790
17. Calisto FM, Nunes N, Nascimento JC (2022) Modeling adoption of intelligent agents in medical imaging. Int J Hum Comput Stud 168. https://doi.org/10.1016/j.ijhcs.2022.102922
18. Cheng M, Li X, Xu J (2022) Promoting Healthcare Workers' Adoption Intention of Artificial-Intelligence-Assisted Diagnosis and Treatment: The Chain Mediation of Social Influence and Human-Computer Trust. Int J Environ Res Public Health 19. https://doi.org/10.3390/ijerph192013311
19. Hameed BMZ et al (2023) Breaking Barriers: Unveiling Factors Influencing the Adoption of Artificial Intelligence by Healthcare Providers. Big Data Cogn Comput 7. https://doi.org/10.3390/bdcc7020105
20. Liu CF, Chen ZC, Kuo SC, Lin TC (2022) Does AI explainability affect physicians' intention to use AI? Int J Med Inf 168:104884. https://doi.org/10.1016/j.ijmedinf.2022.104884
21. Tran AQ et al (2021) Determinants of Intention to Use Artificial Intelligence-Based Diagnosis Support System Among Prospective Physicians. Front Public Health 9:755644. https://doi.org/10.3389/fpubh.2021.755644
22. Choudhury A, Asan O, Medow JE (2022) Effect of risk, expectancy, and trust on clinicians' intent to use an artificial intelligence system -- Blood Utilization Calculator. Appl Ergon 101:103708. https://doi.org/10.1016/j.apergo.2022.103708
23. Ghaffar F, Islam YZ, Furtado N, Burns C (2025) Trust and accuracy in AI: Optometrists favor multimodal AI systems over unimodal for glaucoma diagnosis in collaborative environment. Comput Biol Med 198:111132. https://doi.org/10.1016/j.compbiomed.2025.111132
24. Ghozali MT, Satibi, Fortwengel G (2025) Adoption Intentions Toward AI-based Clinical Decision Support Tools: A Tam Study on Hospital Pharmacists. J Inform Communication Technol 24:44–69. https://doi.org/10.32890/jict2025.24.3.3
25. Goel K, Sindhgatta R, Kalra S, Goel R, Mutreja P (2022) The effect of machine learning explanations on user trust for automated diagnosis of COVID-19. Comput Biol Med 146:105587. https://doi.org/10.1016/j.compbiomed.2022.105587
26. Gomez C, Smith BL, Zayas A, Unberath M, Canares T (2024) Explainable AI decision support improves accuracy during telehealth strep throat screening. Commun Med (Lond) 4:149. https://doi.org/10.1038/s43856-024-00568-x
27. Katzburg O, Roimi M, Frenkel A, Ilan R, Bitan Y (2024) The Impact of Information Relevancy and Interactivity on Intensivists' Trust in a Machine Learning-Based Bacteremia Prediction System: Simulation Study. JMIR Hum Factors 11:e56924. https://doi.org/10.2196/56924
28. Russ P et al (2025) Feasibility of a multimodal AI-based clinical assessment platform in emergency care: an exploratory pilot study. Front Digit Health 7:1657583. https://doi.org/10.3389/fdgth.2025.1657583
29. Zhang T, Bae SW, Chung T, Dey AK (2026) Emotion-Adaptive Large Language Model-Driven Clinical Decision Support: User Evaluation of the Empathic Clinical Decision Support System Framework for Trust and Explainability. JMIR Hum Factors 13:e89005. https://doi.org/10.2196/89005
30. Arango SD et al (2025) Patient Perceptions of Artificial Intelligence in Hand Surgery: A Survey of 511 Patients Presenting to a Hand Surgery Clinic. *J Hand Surg Am* 50, 1410 e1411-1410 e1418 https://doi.org/10.1016/j.jhsa.2025.01.004
31. Asokan A et al (2023) Physician views of artificial intelligence in otolaryngology and rhinology: A mixed methods study. Laryngoscope Investig Otolaryngol 8:1468–1475. https://doi.org/10.1002/lio2.1177
32. Maleki MS, Sahebi L, Shahvari Z, Samadi S (2026) Attitudes of the Iranian Public Toward the Clinical Use of Artificial Intelligence in Medicine: A Cross-Sectional Survey. Patient Prefer Adherence 20:563494. https://doi.org/10.2147/ppa.S563494

33. Radeva MN, Hristova E, Georgiev RT, Zlatarova ZI (2026) Awareness, trust, and expectations of AI for glaucoma care among Bulgarian ophthalmologists: Role of demographic factors. PLOS Digit Health 5:e0001199. https://doi.org/10.1371/journal.pdig.0001199
34. Yang K, Zeng Z, Peng H, Jiang Y (2019) Attitudes Of Chinese Cancer Patients Toward The Clinical Use Of Artificial Intelligence. Patient Prefer Adherence 13:1867–1875. https://doi.org/10.2147/PPA.S225952
35. Rodler S et al (2024) Patients' Trust in Artificial Intelligence-based Decision-making for Localized Prostate Cancer: Results from a Prospective Trial. Eur Urol Focus 10:654–661. https://doi.org/10.1016/j.euf.2023.10.020
36. Yokoi R, Eguchi Y, Fujita T, Nakayachi K (2021) Artificial Intelligence Is Trusted Less than a Doctor in Medical Treatment Decisions: Influence of Perceived Care and Value Similarity. Int J Human–Computer Interact 37:981–990. https://doi.org/10.1080/10447318.2020.1861763
37. Zhang P et al (2026) The effect of AI assistance timing on performance and user perceptions in pathological slide diagnosis. Appl Ergon 131:104663. https://doi.org/10.1016/j.apergo.2025.104663
38. Dunker O, Bergsjo LO, Severinsen GH, Glette H, Nilsen KB (2026) Is accuracy enough? trust and barriers to AI-based clinical decision support in clinical neurophysiology. Clin Neurophysiol Pract 11:262–270. https://doi.org/10.1016/j.cnp.2026.04.003
39. Ye T et al (2019) Psychosocial Factors Affecting Artificial Intelligence Adoption in Health Care in China: Cross-Sectional Study. J Med Internet Res 21:e14316. https://doi.org/10.2196/14316
40. Woodcock C, Mittelstadt B, Busbridge D, Blank G (2021) The Impact of Explanations on Layperson Trust in Artificial Intelligence-Driven Symptom Checker Apps: Experimental Study. J Med Internet Res 23:e29386. https://doi.org/10.2196/29386
41. Huo W, Zheng G, Yan J, Sun L, Han L (2022) Interacting with medical artificial intelligence: Integrating self-responsibility attribution, human–computer trust, and personality. Comput Hum Behav 132:107253. https://doi.org/10.1016/j.chb.2022.107253
42. Yakar D, Ongena YP, Kwee TC, Haan M (2022) Do People Favor Artificial Intelligence Over Physicians? A Survey Among the General Population and Their View on Artificial Intelligence in Medicine. Value Health 25:374–381. https://doi.org/10.1016/j.jval.2021.09.004
43. Schulz PJ et al (2023) Modeling the influence of attitudes, trust, and beliefs on endoscopists' acceptance of artificial intelligence applications in medical practice. Front Public Health 11:1301563. https://doi.org/10.3389/fpubh.2023.1301563
44. Busch F et al (2025) Multinational Attitudes Toward AI in Health Care and Diagnostics Among Hospital Patients. *JAMA Netw Open* 8, e2514452 https://doi.org/10.1001/jamanetworkopen.2025.14452
45. Curiello S, Chapman A, Wyatt JC (2026) AI and clinicians growing together: A cross-sectional survey of clinicians' attitudes toward AI-CDSS with comparison to 2020 data. Clin Med (Lond) 26:100589. https://doi.org/10.1016/j.clinme.2026.100589
46. Awad NHA et al (2025) When Machines Decide: Exploring How Trust in AI Shapes the Relationship Between Clinical Decision Support Systems and Nurses' Decision Regret: A Cross-Sectional Study. Nurs Crit Care 30:e70157. https://doi.org/10.1111/nicc.70157
47. Hoff KA, Bashir M (2015) Trust in Automation:Integrating Empirical Evidence on Factors That Influence Trust. Hum Factors 57:407–434. https://doi.org/10.1177/0018720814547570
48. Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: people erroneously avoid algorithms after seeing them err. J Exp Psychol Gen 144:114
49. Davis SE, Salwei ME (2026) Explainability in context: calibrating appropriate trust and reliance in artificial intelligence. J Am Med Inform Assoc 33:1554–1563. https://doi.org/10.1093/jamia/ocag082
50. Lee JD, See KA (2004) Trust in automation: designing for appropriate reliance. Hum Factors 46:50–80. https://doi.org/10.1518/hfes.46.1.50_30392
51. Benda NC, Novak LL, Reale C, Ancker JS (2022) Trust in AI: why we should be designing for APPROPRIATE reliance. J Am Med Inform Assoc 29:207–212. https://doi.org/10.1093/jamia/ocab238

Tables

Table 1. Characteristics of Included Studies

Study	Year	Location	Study Design	Population (N)	Participant Characteristics	AI Application
Asokan et al.	2023	United States	Mixed-methods study; survey and semistructured interviews	HCPs (survey n=76; interviews n=19)	Otolaryngologists/rhinologists; 49% academic and 51% private practice; varied experience	Clinical AI for diagnosis, imaging, treatment recommendations, risk profiling
Awad et al.	2025	Egypt	Cross-sectional survey; regression/moderation	HCPs (N=250)	ICU nurses; 76% female; 48% aged 30–39	AI-CDSS providing predictive information for high-stakes ICU decisions
Calisto et al.	2022	International	Cross-sectional survey; CFA/SEM	HCPs (N=322)	Medical-imaging professionals; 59.9% male; 65.2% had >10 years' experience	AI/intelligent agents supporting medical-imaging diagnosis and workflow
Curiello et al.	2026	United Kingdom and Italy	Cross-sectional survey; 2020-sample comparison	HCPs (N=215)	Multidisciplinary clinicians; 57.7% female; 51.1% aged ≥50	General AI-CDSS for diagnosis, prognosis, medication, risk, treatment
Ghaffar et al.	2025	Canada	Within-subject experiment; repeated measures	HCPs (N=20)	Optometrists; mean experience 6.6 years; 10 women, 8 men, 2 trans/non-binary	Unimodal and multimodal AI diagnostic-support systems
Ghozali et al.	2025	Indonesia	Cross-sectional survey; PLS-SEM	HCPs (N=235)	Hospital pharmacists; 56.6% female; 82.6% had no prior CDSS use	ML-based antibiotic-resistance prediction CDSS
Gomez et al.	2024	United States	Randomized experiment; three-arm repeated measures	HCPs (N=121)	Physicians, NPs, and PAs; 74.4% female; varied experience	Explainable AI-CDSS using throat images to predict streptococcal pharyngitis
Katzburg et al.	2024	Israel	Randomized experiment; simulation study	HCPs (N=47)	Critical-care physicians; 33 male and 14 female	ML-CDSS predicting bacteremia with varying information relevance/interactivity
Perfalk et al.	2025	Denmark	Randomized experiment; 3-arm survey	HCPs (N=780)	Multidisciplinary psychiatric staff; median age 47; 81% female	ML-CDSS for psychiatric diagnosis/treatment support
Radeva et al.	2026	Bulgaria	Cross-sectional survey; subgroup analyses	HCPs (N=156)	Ophthalmologists/residents; 73.1% female; median age 35	AI for glaucoma screening, diagnosis, monitoring, and treatment support
Zhang et al.	2026	China	Within-subject experiment; AI-assistance timing comparison	HCPs (N=108)	Cytopathologists; 80.6% female; mean experience 7.8 years	AI-assisted cervical-cytology slide-reading system
Cheng et al.	2022	China	Cross-sectional survey; mediation/path analysis	HCPs (N=343)	Dental HCPs; 71.1% female; 77.0% aged ≤40	AI-assisted diagnosis/treatment systems
Choudhury et al.	2022	United States	Cross-sectional survey; CFA/SEM	HCPs (N=119)	Mostly physician residents; 53.8% female; most aged 25–35	EHR-integrated AI Blood Utilization Calculator
Liu et al.	2022	China, Taiwan	Cross-sectional survey; PLS-SEM	HCPs (N=295)	Physicians; 74.2% male; 37.6% residents and 62.4% attendings	General AI/ML decision support
Goel et al.	2022	Not clearly specified	Within-subject experiment; explanation-condition comparison	HCPs (N=30)	Clinicians with CT-reading experience; varied experience	Deep-learning COVID-19 CT diagnostic system with visual explanations
Schulz et al.	2023	Asia-Pacific region	Cross-sectional survey; SEM	HCPs (N=164)	Gastroenterology clinicians; mean age ~44; >50% had >10 years' specialty experience	AI for colorectal-polyp detection/classification/intervention
Ratta et al.	2025	India	Cross-sectional survey; PLS-SEM; mediation	HCPs (N=320)	Physicians, consultants, and technicians; 71.9% male; 62.5% aged ≤40	AI-CDSS for diagnosis/clinical decisions
Arango et al.	2025	United States	Cross-sectional survey; multivariable regression	Patients (N=511)	55.4% female; age 18–≥75; 83.0% recruited from urban clinic	Patient-facing GenAI chatbots
Busch et al.	2025	43 countries	Cross-sectional multicenter survey; mixed-effects models	Patients (N=13,806)	Median age 48; 50.5% male; 64.8% from Global North	General healthcare/diagnostic AI

Study	Year	Location	Study Design	Population (N)	Participant Characteristics	AI Application
Perfalk et al.	2024	Denmark	Randomized experiment; 3-arm online survey	Patients (N=992)	Psychiatric patients; median age 35; 70% female	ML-CDSS for psychiatric diagnosis/treatment
Rodler et al.	2024	Germany	Prospective cross-sectional survey; 3 clinical-pathway stages	Patients (N=466)	Localized prostate-cancer patients; median age 66–70 across groups	Hypothetical multimodal AI: diagnosis, risk, treatment
Russ et al.	2025	Germany	Feasibility pilot; single-arm post-interaction evaluation	Patients (N=20)	Low-acuity ED patients; 70% female; mean age 45.1	Multimodal AI platform for history-taking, triage, diagnosis, and report generation
Yang et al.	2019	China	Cross-sectional survey; multivariable regression	Patients (N=402)	Cancer patients; mean age 47.9; 49.5% male	General medical AI for independent or physician-assisted care
Huo et al.	2022	China	Time-lagged survey; mediation/moderation	Patients (N=249)	Prior medical-AI users; 60% female; 67.9% aged <30	Medical AI: independent/assistive diagnosis/treatment
Woodcock et al.	2021	United Kingdom	Randomized experiment; 2×4 between-subjects design	General population (N=750)	Age 18–87 (M=35.8); 68.3% employed; 96.5% had prior headache experience	AI symptom-checker chatbot providing diagnosis, explanation, and triage guidance
Ye et al.	2019	China	Cross-sectional survey; CFA/SEM	General population (N=474)	64.3% female; ~91.7% aged 18–40; 71.9% had a bachelor's degree	Ophthalmic AI for eye screening, early detection, and diagnostic support
Yakar et al.	2022	The Netherlands	Cross-sectional survey; nationally representative sample	General population (N=1,909)	Mean age 55.1; 52.4% female; nationally representative household sample	AI in radiology/robotic surgery/dermatology
Yokoi et al.	2021	Japan	Randomized experiment; 2×2×2 factorial scenario design	General population (N=272)	College graduates aged 23–39; 55.5% female	Hypothetical AI decision-maker: diagnosis/medication
Maleki et al.	2026	Iran	Cross-sectional survey; multivariable regression	Mixed—patients and companions (N=388)	Inpatients, outpatients, and companions; mean age 40.95; 53.9% male	Hypothetical medical AI: autonomous/physician-assisted
Tran et al.	2021	Vietnam	Cross-sectional survey; PLS-SEM	Medical students / prospective HCPs (N=211)	Mean age 20.6; 73.5% female; 89.1% urban residence	Hypothetical AI diagnosis-support system
Zhang et al.	2026	United States	Within-subject mixed-methods experiment	Mixed (N=33)	Mostly graduate students; 60.6% male; included clinicians/medical experts and medical product managers	LLM-based empathic CDSS integrating prediction, XAI, and adaptive dialogue
Dunker et al.	2026	Norway	Cross-sectional survey; between-group comparisons	Mixed—patients and HCPs (N=193)	129 patients and 64 HCPs; mean age 54 and 49 years, respectively	AI-CDSS for neurophysiology diagnosis
Hameed et al.	2023	India	Cross-sectional survey; PLS-SEM	Mixed—HCPs and medical trainees (N=392)	Consultants, medical students, residents, and technicians/nurses; ~72% male; most had <5 years' experience	Healthcare AI for diagnosis/treatment

Abbreviations: *AI, artificial intelligence; AI-CDSS, AI-based clinical decision support system; CDSS, clinical decision support system; CT, computed tomography; ED, emergency department; EHR, electronic health record; GenAI, generative AI; HCP, healthcare professional; ICU, intensive care unit; LLM, large language model; ML, machine learning; NP, nurse practitioner; PA, physician assistant; XAI, explainable AI.*

Table 2. Trust Conceptualization and Measurement

Study	Trust Definition	Trust Target	Theoretical Framework	Trust Dimensions	Instrument S
Asokan et al.	No explicit definition; operationalized as willingness to rely on AI for specific clinical tasks and conditions perceived to make AI trustworthy	AI-generated radiographic interpretations, malignancy identification, laboratory interpretation, and decision support	No explicit trust theory; survey domains informed by AI-in-medicine literature	Reliability/competence; accuracy; concordance with physician judgment; validation; transparency; autonomy	Study-develo item survey; literature info and tailored otolaryngolo interviews ac qualitative exploration
Awad et al.	No explicit definition; operationalized as faith in AI-CDSS recommendations, procedures, and conclusions	AI-CDSS recommendations, procedures, and conclusions	Conflict Theory of Decision Making; Cognitive Appraisal Theory; trust modeled as moderator	Dependability of AI-CDSS recommendations, procedures, and conclusions	Trust in AI Sc from Ratta et (2025); expe reviewed and tested
Calisto et al.	Defined as belief that a system/party will perform according to user expectations, particularly under uncertainty and reduced control	Intelligent agents/AI in medical-imaging workflow and patient-information handling	Extended UTAUT including security, risk, trust, workflow impact, and AI guidance	Unidimensional perceived trust; expected performance, security, and privacy	28-item questionnair adapted from UTAUT and technology-a measures
Curiello et al.	No explicit definition; operationalized as confidence in AI tools, algorithms, predictive/diagnostic capability, accuracy, and adaptability	AI clinical tools and algorithms	Technology-acceptance constructs, particularly UTAUT; no dedicated trust theory	System/algorithm trust; diagnostic/predictive competence; accuracy; adaptability	Survey partia replicated/ex Petkus et al. AI-specific it added; Italian version forwa back translat
Ghaffar et al.	Defined as willingness to be vulnerable based on expectations of reliability, accuracy, and integrity; appropriate trust framed as calibrated trust	Input data modality; information quality; technical accuracy; output explainability; system quality	Performance–Process–Purpose (PPP) framework	Input Data Modality; Information Quality; Technical Accuracy; Output Explainability; System Quality	Study-develo questionnair informed by Checklist for TOAST, Körb MDMT, and r measures
Ghozali et al.	Defined as confidence in AI reliability, accuracy, and dependable clinical recommendations	ML-based antibiotic-resistance predictor and recommendations	Extended TAM including TiT, SI, FC, and Perceived Risk	Reliability/dependability; scientific validity; fairness; accuracy	Multi-item sc adapted from TAM/extend literature; co tailored and pretested
Gomez et al.	No explicit definition; operationalized as confidence in and willingness to rely on CDSS diagnostic advice	AI-CDSS diagnostic advice; comparison with Modified Centor Score	Human-centered XAI and HCI principles; no dedicated trust theory	Overall perceived trust in CDSS advice	3 trust items adapted from and HCT sur
Katzburg et al.	Defined as an attitude that an agent will help achieve one's goals under uncertainty and vulnerability	ML bacteremia-prediction system, including functioning, methods, predictions, competence, reliability, and understandability	Lee and See; Madsen and Gregor HCT model; cognition-based trust emphasized	Understandability; Technical Competence; Reliability	14 items fron Madsen and Gregor's HCT questionnair
Perfalk et al.	No explicit definition; operationalized as perceived safety, appropriate use, and belief that ML-CDSS advantages outweigh disadvantages	ML-CDSS and psychiatric services' ability to use it safely/appropriately	No dedicated trust theory; informed by trust, transparency, explainability, and implementation literature	Safety/appropriateness; overall favorable trust; trust/distrust treated separately	Study-specifi experiment questionnair adapted from patient study trust/distrust domains eva by PCA
Radeva et al.	No explicit definition; operationalized as confidence in AI diagnostic accuracy and trust in AI-generated treatment recommendations	AI diagnostic capability and glaucoma-treatment recommendations	TAM-informed; no dedicated trust-specific theory	Diagnostic accuracy/competence; treatment-recommendation trust	Study-develo item survey; tested and ex reviewed
Zhang et al.	No explicit definition; operationalized through confidence in performance, safety of reliance, comparative competence, reliability, and explanation-related trust	AI diagnostic-assistance tool and its diagnostic suggestions	Human–AI collaboration and XAI literature; measure informed by Hoffman et al.	Performance confidence; decision authority/reliance; comparative competence; reliability; explainability-related trust	13-item post questionnair on Hoffman (2018); 4 iter administered interaction
Cheng et al.	Defined as confidence in and willingness to act on recommendations, actions, and decisions of an AI decision aid	AI-assisted diagnosis/treatment systems and their recommendations/functions	Integrated UTAUT and HCT theory	Functional effectiveness; dependability; understanding of user needs; expected functionality; information trustworthiness	Established s translated/ad HCT from Gu al. (2018); ex reviewed and pretested

Study	Trust Definition	Trust Target	Theoretical Framework	Trust Dimensions	Instrument S
Choudhury et al.	Defined as the extent to which clinicians rely on the AI-enabled BUC to achieve transfusion and patient-safety goals	Blood Utilization Calculator and its information/recommendations	Expectancy-value theory; modified extended UTAUT-2	Overall system trust; trust in system-provided information	Items adapte validated UT and prior trus scales
Liu et al.	Defined as willingness to be vulnerable to technology based on expectations of predictability, reliability, and utility	AI/ML algorithm and algorithm-generated recommendations/results	AI–human interaction framework integrating XAI, TT, Perceived Value, and BI	Reliability and credibility of algorithm-generated recommendations/results	Items adapte literature; translated/ba translated, ex reviewed, an tested
Goel et al.	No explicit definition; operationalized as clinicians' ability to rely on and trust the automated diagnostic system	Automated COVID-19 diagnostic system and predictions	Körber's Trust in Automation framework	Direct system trust/reliance	8-item quest adapted fron Körber; doma Reliability, Understandir Trust
Schulz et al.	Trust conceptualized as a mechanism supporting AI adoption under clinical uncertainty; operationalized as trust in AI-generated findings/results	AI technology and findings/results from tested/licensed AI applications	Expectancy-value framework; TRA; TPB; TAM	Credibility/reliability of AI-generated findings/results	Measures ba theoretical frameworks previously va questionnair pretested wit gastroentero
Ratta et al.	Defined as belief that an entity is honest, dependable, effective, or worthy of reliance; operationalized as confidence in AI-CDSS recommendations, procedures, and conclusions	AI-CDSS recommendations, system performance, patient privacy protection, and developer competence	Extended UTAUT with perceived risk, self-efficacy, and trust; trust modeled as mediator	Recommendation accuracy; privacy/security; expected performance; developer competence	Items adapte validated technology-adoption/tru literature; tru items adapte McKnight an Chervany; pr
Arango et al.	No explicit definition; operationalized as confidence in AI chatbots' ability to provide accurate information, diagnosis, treatment recommendations, and complication management	Patient-facing AI chatbots and their clinical information/decision-support functions	No formal trust or technology-acceptance framework	Informational accuracy; diagnostic competence; treatment appropriateness; complication-management capability	Study-develo item survey; literature info and reviewed hand surgeo
Busch et al.	No explicit definition; operationalized as confidence in AI improving healthcare, providing reliable/accurate information, predicting treatment response, and making consequential decisions	Healthcare/diagnostic AI and AI-generated health, diagnostic, and treatment-response information	No trust-specific framework	General confidence/benefit; information reliability; diagnostic accuracy; treatment-response accuracy; severity-dependent acceptance; high-stakes decision trust	Study-develo item instrum informed by systematic r patient interv expert panel, cognitive inte and pilot vali
Perfalk et al.	No explicit definition; operationalized as perceived safety/appropriateness of ML-supported psychiatric decisions and belief that benefits outweigh disadvantages	ML-CDSS and psychiatric services' ability to use it safely/appropriately	No dedicated trust framework; informed by trust, explainability, human oversight, and implementation literature	Safety of ML-supported decisions; institutional/implementation trust; perceived benefit versus disadvantage	Study-specifi experiment questionnair trust/distrus domains spe priori and ex with PCA
Rodler et al.	No explicit definition; operationalized as perceived safety/confidence in AI-supported or autonomous diagnosis across varying physician oversight and clinical severity	AI-supported prostate-cancer diagnostic decision-making, with and without physician verification	Human–machine interaction and shared decision-making; no formal trust-specific theory	Diagnostic competence; safety; physician supervision vs autonomy; disease severity; personalization; explainability; evidence currency	Study-develo questionnair LMU Human-Machine Inte group; adapt prior AI-perc instrument
Russ et al.	No explicit definition; operationalized through reliability/competence, understandability/predictability, developer intentions, general trust, familiarity, and propensity to trust automation	Multimodal AI clinical assessment platform	Körber's Trust in Automation framework; interpreted through Performance–Process–Purpose	Reliability/Competence; Understandability/Predictability; Intention of Developers; General Trust; Familiarity; Propensity to Trust	Validated 19 Trust in Auto questionnair
Yang et al.	No explicit definition; operationalized as belief in diagnostic/therapeutic advice from an independently acting AI doctor	Autonomous AI doctor and its diagnostic/therapeutic recommendations	No trust-specific/technology-acceptance framework	Autonomous diagnostic trust; autonomous therapeutic-advice trust	Study-develo questionnair tested with 3 cancer patie
Huo et al.	HCT conceptualized as a trust attitude arising in human–AI	Medical AI for diagnosis/treatment following	Attribution Theory; Human–Computer	Four unnamed aspects; items reflected reliance/dependability,	12-item HCT from Gulati e

Study	Trust Definition	Trust Target	Theoretical Framework	Trust Dimensions	Instrument S
	interaction and belief that technology can support goal achievement under uncertainty/vulnerability	prior service failure	Interaction/Trust literature; Big Five personality framework	perceived risk, caution, and negative consequences	(2018), adap medical AI
Woodcock et al.	No explicit definition; operationalized through faith, reliance, satisfaction, and comprehension toward the AI symptom checker and its explanations	Symptom-checker explanation, system, and advice	Contrastive explanation; Information Gap Theory; explanatory-coherence and social/cognitive perspectives	Faith; Comprehension; Depth; reliance/satisfaction items loaded into derived factors	Adapted fror Hoffman's Explanation Satisfaction and prior HCI/automat trust measur tailored to symptom-ch use
Ye et al.	Defined as belief in an entity's honesty, reliability, effectiveness, and dependability; operationalized as confidence in ophthalmic AI reliability, accuracy, effectiveness, security, and privacy protection	Ophthalmic AI devices, diagnostic reports, and data-protection stakeholders	TAM and TPB, supplemented by HBM, DFT, and SQB; trust modeled as moderator	Reliability/accuracy; comparative confidence versus ophthalmologists; security/privacy	3 trust items adapted fron technology-acceptance/ literature; co adapted, translated/ba translated, a expert reviev
Yakar et al.	No explicit definition; operationalized as acceptance of AI implementation, willingness to allow AI to perform physician tasks, comfort with reduced human involvement, implementation readiness, procedural clarity, and comparative confidence	AI implementation in radiology, robotic surgery, and dermatology	No dedicated trust theory; adapted from validated AI-implementation questionnaire	Specialty-specific trust; task substitution; human oversight; procedural clarity; communication; implementation readiness; comparative expertise	Three specia specific scal adapted fron validated rad AI trust scale modified for and dermato
Yokoi et al.	Defined as a psychological state in which an agent is expected to help achieve one's goals under uncertainty and vulnerability	AI or physician responsible for diagnosis and medication recommendation	Lee and See's trust-in-automation conceptualization; risk-perception, media-equation, and uniqueness-neglect literature	General agent trust; reliance; willingness to entrust	Trust items a from Yokoi/Nakay research; rel constructs a from establis scales
Maleki et al.	No explicit definition; operationalized as confidence in AI acting independently for diagnosis/treatment recommendations	Autonomous medical AI providing diagnosis/treatment recommendations	No dedicated trust framework; broader technology-acceptance literature used as context	Autonomous diagnostic trust; autonomous treatment-recommendation trust	Adapted fror et al.; translated/ba translated in Persian, expe reviewed, an tested
Tran et al.	Defined as belief in a technology's capability before direct use or prior experience	AI-based Diagnosis Support System and its diagnostic accuracy/reliability	Extended UTAUT plus trust-formation model	Diagnostic accuracy; reliability	26-item questionnair across nine constructs, a from UTAUT, technology-adoption, an literature
Zhang et al.	No explicit definition; operationalized as perceived trustworthiness in meeting user needs/expectations and overall system reliability	Interactive CDSS, including personalized predictions, explanations, and responses	Human-centered XAI; affective computing; emotion-adaptive human–AI interaction; prior automation-trust literature	Trustworthiness; reliability	Structured p interaction questionnair trust/reliabili items adapte literature
Dunker et al.	No explicit definition; operationalized through confidence in AI-supported decisions, institutional data handling, and conditions affecting trust	AI-CDSS decisions; technician decisions; institutions/commercial actors handling data; implementation process	No dedicated trust theory; literature on AI attitudes, trust, implementation, and neurophysiology	Clinical decision trust; institutional/data-governance trust; provenance/data-origin trust; implementation trust	Study-develo questionnair literature info iteratively ref and pilot test
Hameed et al.	Defined as belief in technology capability independent of motives/intentions; operationalized as belief that AI can provide accurate, reliable, safe, and convenient diagnosis	Healthcare AI and its diagnostic capability	Integrated UTAUT/UTAUT2 and initial-trust literature	Accuracy; reliability; safety; convenience	Items adapte validated technology/A adoption stu

Abbreviations: *AI, artificial intelligence; AI-CDSS, AI-based clinical decision support system; AVE, average variance extracted; CFA, confirmatory factor analysis; CFI, comparative fit index; CR, composite reliability; EFA, exploratory factor analysis; HBM, Health Belief Model; HCI, human–computer interaction; HCT, human–computer trust; KMO, Kaiser–Meyer–Olkin; ML, machine learning; PCA, principal component analysis; RMSEA, root mean square error of approximation; SEM, structural equation modeling; SRMR, standardized root mean square residual; TAM, Technology Acceptance Model; TiA, Trust in Automation; TLI, Tucker–Lewis index; TPB, Theory of Planned Behavior; UTAUT, Unified Theory of Acceptance and Use of Technology; XAI, explainable AI.*

Table 3. Factors Associated With Trust

Study/Population (N)	AI/System Factors	Task/Clinical Factors	Human–AI Interaction	Org./Implementation Factors	Individual Factors
Asokan et al. HCPs (survey n=76; interviews n=19)	Accuracy, algorithm quality, transparency, reproducibility, and validation identified as trust-related characteristics; primarily descriptive/qualitative	Greater trust for simpler, lower-risk, objective, and augmentative tasks; lower trust for complex/high-stakes or interpretive tasks	Concordance with clinician judgment/expert opinion and preservation of physician autonomy supported trust	Peer-reviewed validation, expert evaluation, testing, liability, workflow burden, and implementation efficiency identified as implementation considerations	NE; practice settir and clinical experi reported descripti by subgroup.
Awad et al. HCPs (N=250)	NE	NE	Self-reported AI-CDSS reliance positively correlated with trust (r=.51, p<.01); directionality cannot be established	NE	NE
Calisto et al. HCPs (N=322)	Perceived security positively associated with trust (β=.715, p<.001); perceived risk weakly negatively associated (β=−.124, p=.095)	Positive perceived clinical-workflow impact weakly associated with trust (β=.179, p=.055)	NE	Facilitating conditions included but not modeled as a trust predictor	NE; participant characteristics we assessed primaril moderators of fac trust relationships
Curiello et al. HCPs (N=215)	Performance problems, accuracy, bias, privacy, risk, and benefits measured; not tested as a trust predictor	Task-specific trust differences not tested	Communication barriers and reduced clinician–patient interaction measured; not tested as a trust predictor	Facilitating conditions, training, regulation, liability, infrastructure, and support measured; not tested as a trust predictor	NE
Ghaffar et al. HCPs (N=20)	Multimodal AI produced higher overall trust than unimodal AI (M=3.98 vs 3.00, p<.001, d=1.53) and higher ratings across all five trust targets	Clinical task constant as glaucoma diagnosis	Alignment of multimodal information with clinicians' mental models/workflow proposed as mechanism; not separately tested	NE	NE; professional expertise and gen reported under subgroup analyse
Ghozali et al. HCPs (N=235)	Perceived Risk not significantly associated with trust (β=.069, p=.124)	Clinical context constant	Ease of Use measured but not modeled as direct trust predictor	FC strongly positively associated with trust (β=.507, p<.001)	NE
Gomez et al. HCPs (N=121)	Decision-support source differed between Modified Centor Score and AI-CDSS	Strep-throat task constant; case difficulty not independently modeled	Explanation condition varied. Relative to Centor Score, trust was lower for Performance Characteristics (estimate=−.41, p<.001) and Feature Contribution (−.24, p=.019), but not Examples (−.10, p=.325); AI explanation conditions did not significantly differ from one another	Familiarity/workflow integration discussed; not tested	NE; provider type clinical experience reported under gr comparisons.
Katzburg et al. HCPs (N=47)	Clinically relevant information produced higher trust than algorithm-centered/nonrelevant information (β=2.684, p=.05); strongest effect for Technical Competence (β=4.5, p<.001)	Bacteremia-prediction task constant	Perceived information relevancy (β=.824, p<.001) and perceived interactivity (β=.613, p<.001) positively associated with trust; objective UI interactivity alone was nonsignificant	NE	NE; clinical experi was examined primarily as a moderator of the interactivity–trust relationship.

Study/Population (N)	AI/System Factors	Task/Clinical Factors	Human–AI Interaction	Org./Implementation Factors	Individual Factors
Perfalk et al. HCPs (N=780)	Intervention provided general information about ML-CDSS; individual system features were not isolated	Psychiatric decision-support context constant	Information emphasized ML as clinician support rather than replacement; components not independently isolated	Reliable performance and clinician final authority embedded in intervention	General trust disposition and baseline ML-CDSS knowledge were examined in strati analyses but not established as dir predictors of trust
Radeva et al. HCPs (N=156)	Accuracy, reliability, transparency, and misinterpretation discussed; not tested as a trust predictor	Diagnosis and treatment-recommendation trust measured separately; task difference not formally tested	Preference for AI as support rather than replacement described but not linked statistically to trust	Training, integration, financial, legal, and ethical barriers identified qualitatively	NE; demographic professional characteristics reported under subgroup analyse
Zhang et al. HCPs (N=108)	Reliability/performance characteristics informed trust judgments but were not formally modeled as predictors	Greater trust for preliminary/assistive roles than autonomous final diagnosis	Direct AI interaction recalibrated reliance: safety of bypassing preliminary human review decreased (M=3.83→3.52, p=.002) and autonomous diagnosis decreased (3.33→2.81, p<.001); explainability/transparency viewed as important	Human oversight, workflow integration, and clinician control identified as implementation conditions	NE
Cheng et al. HCPs (N=343)	PE positively associated with HCT (β=.278, p<.001; r=.559, p<.01)	Dental diagnosis/treatment context constant	EE positively associated with HCT (β=.482, p<.001; r=.558, p<.01)	NE	NE
Choudhury et al. HCPs (N=119)	Perceived Risk negatively associated with trust (β=−.27, p=.005); higher-order Expectancy positively associated with trust (β=.55, p<.001)	Blood-transfusion context constant	EE contributed to higher-order Expectancy but was not estimated separately against trust	EHR integration and clinician decision authority characterized implementation context but were not modeled as trust predictors	NE
Liu et al. HCPs (N=295)	Objective AI performance not independently modeled; reliability/credibility were part of the trust construct	Specialty differences treated as RQ4 subgroup analysis	AI Explainability positively associated with TT (β≈.52); XAI explained 27.1% of trust variance	Not formally tested	NE; specialty and professional posit reported under subgroup analyse
Goel et al. HCPs (N=30)	Diagnostic model held constant; model accuracy not manipulated as a trust factor	COVID-19 CT classification task constant	Explanation quality varied: no explanation, strong explanation, weak explanation. Trust pattern was nonsignificant (Friedman p=.240); weak explanations significantly reduced Understanding (p=.004)	NE	NE
Schulz et al. HCPs (N=164)	Positive expectations regarding AI usefulness/performance positively associated with trust (b=.483, p<.01)	Gastroenterology scenarios varied in complexity, but scenario complexity was not modeled against general trust	No interaction feature independently modeled	Testing/licensing embedded in trust items but not separately tested	More favorable ge attitudes toward A were positively associated with tr AI-generated findi
Ratta et al. HCPs (N=320)	PE, Perceived Risk, and related perceptions modeled against Adoption Intention rather than Trust	NE	EE modeled against Adoption Intention rather than Trust	No implementation factor directly tested against Trust	NE
Arango et al.	Reliability, hallucination/misinformation	Trust measured across information,	NE	Transparency and patient education	Greater technolog familiarity was

Study/Population (N)	AI/System Factors	Task/Clinical Factors	Human–AI Interaction	Org./Implementation Factors	Individual Factors
Patients (N=511)	risk, and transparency discussed; not tested as a trust predictor	diagnosis, treatment, and complication-management tasks; no formal within-person task comparison		recommended but not tested	associated with lo trust in AI-generat treatment recommendations
Busch et al. Patients (N=13,806)	Accuracy/reliability embedded within trust scenarios rather than modeled as independent predictors	Disease severity incorporated into Trust in AI scale, not modeled as independent factor	Explainability and physician collaboration assessed separately but not modeled against trust	No implementation factor tested	Greater AI knowle was associated w more favorable tru responses; poorer reported health w associated with lo trust, while greate technology literac generally associat with more favorab trust.
Perfalk et al. Patients (N=992)	Intervention described how ML learns, generates recommendations, and is used after reliable performance is established; components not isolated	Psychiatric context constant	Human oversight embedded in intervention; not separately estimated	Validation, clinician oversight, and integration into clinical assessment embedded in intervention package	General trust disposition and baseline ML-CDSS knowledge were examined in strati analyses but not established as dir predictors of trust
Rodler et al. Patients (N=466)	AI capability ratings included accuracy, current evidence, personalization, and explainability; not independently modeled as a trust predictor	Trust decreased with increasing perceived complexity/severity; higher illness perception associated with lower autonomous-AI trust	Physician verification markedly increased trust in AI diagnosis (M=4.31 vs 1.75, p<.001, d=2.818); AI checking did not increase trust in physician diagnosis	Human oversight and integration into clinician-led workflow were central implementation conditions	Greater technolog affinity was weakl positively associa with general trust age, education, an overall illness perception were n significantly associated with general trust.
Russ et al. Patients (N=20)	No platform feature independently tested; multimodal sensing, conversational AI, triage, diagnosis, and reporting were bundled	Low-acuity ED context constant	Direct avatar interaction, understandability, information quality, comfort, and safety measured but not statistically related to TiA trust	Clinician review, local processing, safety termination, and physician availability embedded as safeguards; not tested	NE
Yang et al. Patients (N=402)	Accuracy, objectivity, standardization, currency, limited individualization, complexity limitations, and lack of humane care elicited descriptively	Trust assessed separately for diagnosis and treatment; no formal task comparison	Physician involvement strongly preferred but not modeled against trust	National qualification/licensing assumed in scenario; not tested	Prior awareness o knowledge of AI ir medicine was associated with g trust in autonomo diagnosis.

Study/Population (N)	AI/System Factors	Task/Clinical Factors	Human–AI Interaction	Org./Implementation Factors	Individual Factors
Huo et al. Patients (N=249)	AI service failure was common context; error severity, accuracy, explainability, and related system characteristics were not independently modeled	Independent vs assistive AI roles distinguished for downstream acceptance, not HCT comparison	Self-responsibility attribution shaped interpretation of AI failure but is an individual-level cognitive factor rather than an interaction characteristic	Governance/safeguards discussed; not tested	Greater self-responsibility attribution was associated with h HCT; openness ar neuroticism were positively correlat with trust.
Woodcock et al. General population (N=750)	Explanation type experimentally varied: input-influence, social-proof, counterfactual, or none	Disease context varied: migraine vs temporal arteritis, reflecting differences in familiarity/information gaps	Fit between explanation and users' mental models/information needs proposed as an explanatory mechanism; not separately tested	NE	NE
Ye et al. General population (N=474)	No system-level factor formally tested against trust; usefulness, ease of use, and risk were modeled against other adoption constructs	NE	NE	NE	NE
Yakar et al. General population (N=1,909)	Perceived AI efficiency positively correlated with trust across radiology (r=.531), surgery (r=.523), and dermatology (r=.564), all p<.001; positive AI attitudes also correlated with trust	Trust differed significantly by specialty: dermatology M=2.90, radiology M=2.82, robotic surgery M=2.75	Preference for personal interaction negatively correlated with trust: radiology r=−.286, surgery r=−.278, dermatology r=−.229, all p<.001	Implementation readiness and procedural clarity were components of the trust scales, not independent predictors	NE; age, gender, education, and rel characteristics we examined primaril demographic subgroup/correlat differences and ar reported in the subgroup column
Yokoi et al. General population (N=272)	Decision-maker type experimentally varied; AI was trusted less than a physician ($F(1,262)=8.26$, $p=.004$, partial $\eta^2=.031$)	Preferred vs non-preferred treatment recommendation had no significant effect on trust ($p=.298$)	Perceived ability strongly positively associated with trust ($F=168.16$, $p<.001$, partial $\eta^2=.391$); uniqueness neglect negatively associated ($F=5.33$, $p=.022$); manipulated understanding of preferences was nonsignificant	NE	Greater perceived ability of the medi decision-maker w associated with h trust, whereas gre perceived uniquer neglect was associated with lo trust.
Maleki et al. Mixed—patients and companions (N=388)	Accuracy, objectivity, standardization, empathy limitations, and complexity limitations elicited descriptively; not modeled against trust	Trust assessed separately for autonomous diagnosis and treatment; task difference not formally tested	Physician oversight and human–AI collaboration strongly preferred but not modeled against trust	Licensing, oversight, accountability, and regulation framed the scenario but were not separately tested	Greater prior awareness or knowledge of mec AI was associated greater trust in autonomous AI diagnosis; educat was also associat with trust in some analyses.
Tran et al. Medical students / prospective HCPs (N=211)	PE was not significantly associated with Initial Trust ($p>.05$)	Task Complexity influenced PE but had no direct trust path	EE positively associated with Initial Trust ($\beta=.201$, $p<.05$)	NE	NE

Study/Population (N)	AI/System Factors	Task/Clinical Factors	Human–AI Interaction	Org./Implementation Factors	Individual Factors
Zhang et al. Mixed (N=33)	Empathic-CDSS bundled emotion sensing, personalization, XAI, causal/counterfactual reasoning, and LLM explanations; individual components not isolated	Behavioral-health prediction task constant	Emotion-adaptive/personalized interaction increased trustworthiness (M=7.85 vs 5.91, p<.001) and reliability (M=7.79 vs 6.12, p<.001)	Privacy, cost, latency, and implementation constraints discussed; not tested	NE
Dunker et al. Mixed—patients and HCPs (N=193)	Performance, representative training data, transparency, explainability, and clinician agreement descriptively trust-enhancing; poor prior AI performance reduced trust	Trust decreased when AI altered access to specialist care despite physician-equivalent stated accuracy	Human oversight, clinician agreement, and preservation of physician involvement were central trust themes	Accountability, workflow integration, AI literacy/training, governance, and prevention of deskilling identified as important	NE; professional r was treated as a population/compa characteristic.
Hameed et al. Mixed—HCPs and medical trainees (N=392)	PE positively associated with Initial Trust (β=.229, p<.001); technological characteristics affected EE rather than trust directly	Task Complexity affected PE but had no direct trust path	EE positively associated with Initial Trust (β=.225, p=.001)	NE	NE

Abbreviations: *AI, artificial intelligence; AI-CDSS, AI-based clinical decision support system; AOR, adjusted odds ratio; BI, behavioral intention; CDSS, clinical decision support system; CI, confidence interval; EE, effort expectancy; FC, facilitating conditions; HCP, healthcare professional; HCT, human–computer trust; ML, machine learning; NE, not examined; NP, nurse practitioner; OR, odds ratio; PA, physician assistant; PE, performance expectancy; SI, social influence; TiT, trust in technology; TT, technology trust; XAI, explainable AI.*

Table 4. AI-Related Behavioral Outcomes and Trust–Behavior Associations

Study	AI-Related Behavioral Outcome	Behavior Type	Trust–Behavior Findings
Asokan et al.	Willingness to use AI clinical decision aids and accept AI-assisted diagnosis/treatment under various conditions	Self-reported adoption/acceptance	No formal trust–behavior association tested. 82% would use an AI clinical decision aid; qualitative findings indicated that perceived trustworthiness and concordance with clinical judgment influenced willingness to adopt AI
Awad et al.	AI-CDSS reliance/use-related engagement measured separately from trust; decision regret also assessed	Self-reported reliance / use-related engagement	Trust positively correlated with AI-CDSS reliance (r=.51, p<.01). Trust also moderated the Reliance–Decision Regret relationship (β=−.21, p=.012), although regret was not classified as a behavioral outcome
Calisto et al.	Behavioral intention to use AI and willingness to follow AI-generated recommendations (“AI Guidance”)	Self-reported adoption/intended reliance	No direct Trust→Behavioral Intention or Trust→AI Guidance path tested. Trust was embedded in an indirect adoption pathway through Effort Expectancy and Performance Expectancy, but no formal trust→intention indirect effect was estimated
Curiello et al.	Intention to use AI tools in clinical practice, patient management, diagnosis/treatment, and care planning	Self-reported adoption/recommendation-following	No formal trust–intention relationship tested; trust was neither included in the cluster analysis nor modeled as a predictor of AI use intention
Ghaffar et al.	Case-level diagnostic decisions, agreement with AI recommendations, and diagnostic performance	Observed decision/performance	No direct trust→behavior model. Multimodal AI produced both higher trust and generally better performance when AI was correct; when AI was wrong, accuracy fell to 15.1% versus 54.1% with incorrect unimodal AI, suggesting possible overreliance on the more trusted multimodal system
Ghozali et al.	Intention to adopt, regularly use, and recommend the ML-driven predictor	Self-reported adoption/future-use intention	Trust in Technology positively associated with Behavioral Intention (β=.179, p=.017); trust also positively associated with Perceived Usefulness (β=.245, p=.005)
Gomez et al.	Agreement with diagnostic advice; final diagnosis/performance; antibiotic/supportive-care decisions; confirmatory testing	Observed advice-taking/clinical decisions	No individual-level trust→behavior model. Participants agreed more often with AI than with the Centor Score despite lower self-reported trust in two AI conditions. Lower trust in those AI conditions also coincided with more confirmatory testing, indicating partial trust–behavior misalignment
Katzburg et al.	Agreement with ML bacteremia predictions across simulated cases	Observed advice agreement/reliance	No direct trust→agreement model. Agreement did not differ by UI condition; participants consistently agreeing with AI reported greater perceived information relevance (p=.05), a factor also positively associated with trust
Perfalk et al.	No independent behavioral outcome	NE	Not examined; study focused on intervention-induced changes in trust and distrust
Radeva et al.	No clear independent AI-use intention or observed-use outcome measured separately from trust	NE	Not examined; future-role and utility items reflected attitudes/preferences rather than independent behavioral outcomes
Zhang et al.	Diagnostic decisions/performance across AI-assistance modes; human–AI mismatch; collaboration-mode preferences; future-use intention	Observed experimental decision-making / performance; self-reported preference and use intention	No direct trust→behavior model. AI assistance improved performance and was strongly preferred despite only moderate trust. Concurrent assistance was preferred, while participants remained reluctant to grant AI autonomous authority—indicating acceptance of assistance alongside cautious autonomous-AI trust
Cheng et al.	Intention to learn/use AI-assisted diagnosis/treatment, use it in future, and recommend it to others	Self-reported adoption/recommendation intention	HCT positively associated with adoption intention (r=.604, p<.01; β=.155–.181, p<.01) and mediated Performance Expectancy→Adoption Intention (indirect=.043, 95% CI [.010, .088]) and Effort Expectancy→Adoption Intention (.088, 95% CI [.027, .157])
Choudhury et al.	Preferred future frequency of BUC use	Self-reported intended use frequency	Trust positively associated with use intention (β=.55 and .51, both p<.001) and mediated Expectancy→Use Intention (β=.31, p<.001) and Perceived Risk→Use Intention (β=−.15, p=.009)
Liu et al.	Intention to use available AI, expand AI use, and use AI in more clinical work	Self-reported adoption/use intention	Technology Trust positively associated with Behavioral Intention (β=.134, p<.05) and partially mediated the XAI→Behavioral Intention relationship (indirect effect=.070; VAF=.574)
Goel et al.	No independent behavioral outcome	NE	Not examined; “I can rely on the system” was part of the trust measure rather than an independent reliance outcome
Schulz et al.	Acceptance of AI for polyp detection, classification, and AI-supported polypectomy	Scenario-based AI acceptance	Trust positively associated with acceptance (b=.193, p<.01; r=.48, p<.01) and mediated Expectations→Acceptance (indirect=.093, 95% CI [.032, .169]) and the sequential Expectations→Attitude→Trust→Acceptance pathway (.030, 95% CI [.008, .065])
Ratta et al.	adoption intention and self-reported actual use of AI-CDSS	Self-reported adoption intention and current/actual use	Trust mediated Adoption Intention→Actual Usage: direct effect nonsignificant (β=.082, p=.129), indirect effect=.181 (p=.002), total effect=.263 (p<.001)

Study	AI-Related Behavioral Outcome	Behavior Type	Trust–Behavior Findings
Arango et al.	Prior condition-related AI use; general chatbot use; future chatbot-use intention; perceived future importance of AI	Self-reported prior/current/future use	No formal trust→behavior association tested. Only 3.9% had used AI for their condition; 14.1% were likely/very likely to use chatbots for future upper-extremity questions
Busch et al.	Preferences for explainable vs more accurate AI; diagnostic error trade-offs; human–AI decision authority; AI-enabled facility preference	Hypothetical diagnostic / decision-authority preference and facility-choice intention	No individual-level trust→preference model. Despite measurable trust, 70.2% preferred explainable AI and 72.9% preferred AI–physician collaboration with physician final authority; only 4.4% preferred AI-only diagnosis
Perfalk et al.	No independent behavioral outcome	NE	Not examined; study focused on intervention-induced changes in trust/distrust rather than downstream behavior
Rodler et al.	Preference among AI-only, AI-controlled physician, AI-assisted physician, or physician-only care; preference during AI–urologist disagreement; acceptance of greater AI use	Hypothetical clinical choice/AI acceptance	No individual-level trust→choice model. Preferences strongly favored AI-assisted physicians (67.2%) or physicians alone (29.8%); only 0.6% preferred AI alone. During disagreement, 70.8% chose the urologist versus 2.2% AI
Russ et al.	Willingness to recommend the AI platform; future-use item embedded in SUS excluded as independent outcome	Self-reported recommendation intention	No formal trust→recommendation model. 80% selected the most positive recommendation response and trust was generally favorable, but the two were not statistically linked
Yang et al.	Preference for AI versus physician during diagnostic/treatment disagreement; follow-up provider; physician presence; future human–AI role	Hypothetical clinical choice/human–AI role acceptance	No individual-level trust→preference model. Although 90.0% reported some trust in autonomous AI diagnosis, only 11.2% preferred AI during diagnostic disagreement; 85.1% trusted AI treatment advice, but only 8.7% preferred AI during treatment disagreement
Huo et al.	Acceptance of independent and assistive medical AI for diagnosis and treatment	Self-reported AI acceptance / future willingness to use	HCT positively associated with independent-AI acceptance ($r=.418$, $p<.01$) and assistive-AI acceptance ($r=.171$, $p<.01$), and mediated Self-responsibility Attribution→Acceptance for both. Personality traits further moderated HCT–acceptance relationships
Woodcock et al.	Inclination to use the symptom checker in different situations; reliance-like items embedded within the trust measure were excluded as independent outcomes	Self-reported use intention / propensity	No formal trust–behavior association tested. Most participants (91.5%) would consider using the symptom checker in ≥1 situation; use propensity did not differ significantly across experimental conditions
Ye et al.	Intention to use ophthalmic AI and recommend AI-based eye screening to others	Self-reported adoption intention	Trust positively associated with use intention ($\beta=.302$, $p=.003$) and moderated the Perceived Usefulness–Intention relationship (PU×Trust $\beta=-.095$, $p=.049$); usefulness was more strongly associated with intention at lower trust
Yakar et al.	No independent behavioral outcome measured separately from the trust constructs	NE	Not examined; physician-versus-AI preferences and willingness to allow AI to perform clinical tasks were embedded within the trust measures
Yokoi et al.	No independent behavioral outcome	NE	Not examined; reliance and willingness to entrust work were components of the trust measure
Maleki et al.	Preference for AI versus physician during disagreement; physician supervision; preferred communication and future AI role	Hypothetical clinical choice; collaborative-care acceptance	No individual-level trust→preference model. Despite 86.6% reporting some trust in AI diagnosis and 87.1% in AI treatment, only 33.8% and 33.2%, respectively, preferred AI over physicians during disagreement; strong preference remained for physician oversight
Tran et al.	Intention to use an AI-based Diagnosis Support System in future clinical practice	Self-reported future adoption / use intention	Initial Trust was not significantly associated with Behavioral Intention; Social Influence was the only significant direct predictor ($\beta=.527$, $p<.05$)
Zhang et al.	Willingness to recommend the CDSS and continue future use	Self-reported recommendation/future-use intention	No direct trust→behavior model. Empathic-CDSS produced both higher trust and more favorable recommendation ($p=.006$) and continued-use intentions, but individual-level association was not estimated
Dunker et al.	Willingness to share health data for AI development and preference for retaining human final decision authority	Self-reported data-sharing willingness / governance preference	No formal trust→behavior model. Data-sharing willingness was generally high under transparency/security/opt-out conditions, while respondents strongly preferred healthcare professionals to retain final decision authority
Hameed et al.	Intention to adopt/use healthcare AI when available and whenever possible	Self-reported adoption/use intention	Initial Trust strongly positively associated with Behavioral Intention ($\beta=.468$, $p<.001$); Performance Expectancy also significant ($\beta=.255$, $p<.001$), whereas Effort Expectancy, Social Influence, and substitution concerns were not significant direct predictors

Abbreviations: *AI, artificial intelligence; AI-CDSS, AI-based clinical decision support system; BI, behavioral intention; BUC, Blood Utilization Calculator; CDSS, clinical decision support system; HCT, human–computer trust; ML, machine learning; NE, not examined; SUS, System Usability Scale.*

Figures

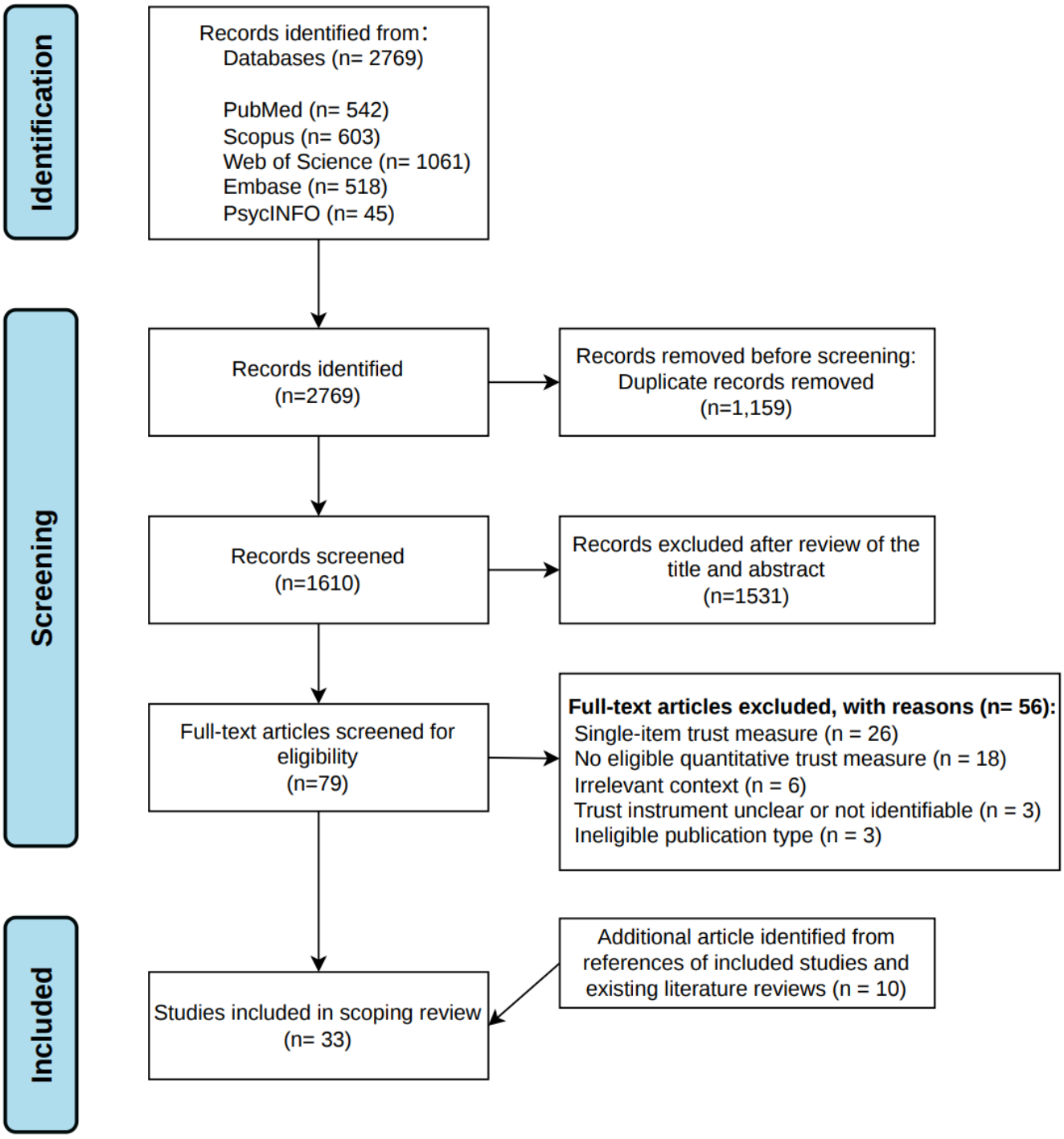


Figure 1

PRISMA flow diagram for the scoping review process (publications from 2010 to 2026).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Searchsyntax.docx