

Supplementary Materials

Table S1. Definition of Extracted Information.....	2
Table S2. Summary of Potential Reasons for RCTs with Non-significant Outcomes .	5
Table S3. Reporting Guideline Use Among Included RCTs Overall and by Publication Period.....	7
Table S4. Distribution of ROBUST-RCT Risk-of-Bias Judgments for Each Assessed Item Across 113 Randomized Controlled Trials	8
Table S5. Database-specific search strategies	8
Figure S1. Distribution of AI Medical Studies by Medical Specialty	11
Figure S2. Traffic-Light Plot of ROBUST-RCT Risk-of-Bias Judgments Across 113 Randomized Controlled Trials	12
Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist	17

This supplementary material has been provided by the authors to give readers additional information about their work.

Table S1. Definition of Extracted Information

Name	Definition	Possible Values / Standardization Notes
First Author	The last name and initials of the first author of the publication.	Free text (e.g., Smith J)
Year	The year the article was published.	Numeric (e.g., 2023)
Title	The full title of the publication.	Free text
Country	The country of origin for the research, based on the institutional affiliations of all authors.	Determined by the intersection of countries listed across all author affiliations. Note: Chinese Mainland, Hong Kong, Macao, and Taiwan are to be collectively categorized as “China”.
Stage	The evaluation phase of the AI system presented in the study.	1. Preclinical Development and Offline Validation: Model development, in silico experiments, retrospective validation, diagnostic accuracy studies, and other offline evaluations in which AI outputs do not guide real-time patient management. 2. Early Live Clinical Evaluation: Prospective evaluations in live clinical settings, where the AI system is deployed within routine workflows and can influence clinical decisions, but studies are primarily exploratory and nonrandomized (e.g., pilot cohorts, prospective before–after studies, or other nonrandomized comparative designs). 3. Comparative Prospective Evaluation: Late-stage assessment in routine clinical practice, defined exclusively as randomized controlled trials that prospectively compare AI-assisted care with a predefined control condition (e.g., usual care, clinician alone, or an alternative workflow).
Specialty	The medical specialty or disease domain of the study, standardized using the International Classification of Diseases 11th Revision (ICD-11).	Use first-level ICD-11 classifications only. Examples: Neoplasms; Diseases of the Musculoskeletal System or Connective Tissue; Diseases of the Nervous System; Diseases of the Visual System.

Centers	The number and geographic scope of the clinical sites involved in the study.	1. Single-center: The study sample is sourced from one single healthcare institution.
		2. Multi-center: The study sample is sourced from multiple healthcare institutions within a single country.
		3. Multinational: The study sample is sourced from multiple healthcare institutions located in more than one country.
Sample Size	The total number of units included in the study's primary analysis.	Prioritize the number of patients. If patient count is unavailable, use other units (e.g., images, lesions, samples) and specify the unit type in a separate notes column.
AI Role	The primary function or application of the AI model within the clinical context described.	1. Care Management This category includes studies where AI is used to optimize clinical workflows, manage treatment plans, or improve the efficiency of healthcare delivery. The primary goal is to enhance the management of care processes and resource allocation, rather than directly diagnosing a disease. Examples include using AI to optimize insulin dosage or to automate radiation therapy planning.
		2. Clinical Decision-Making This category covers AI applications designed to directly assist clinicians with diagnosis, prognosis, or treatment selection. These tools aim to enhance clinical judgment and improve patient outcomes by providing risk assessments or predictions that inform the decision-making process. For instance, AI that estimates mortality to guide serious illness conversations or assists in breast cancer screening is included here.
		3. Patient Behavior and Symptoms This classification pertains to patient-facing AI tools that support health self-management, facilitate behavioral change, or help monitor symptoms. The emphasis is on the direct interaction between the patient and the AI system to improve treatment adherence, lifestyle, and overall quality of life. Common examples include AI-powered apps for chronic disease management and digital interventions to improve medication adherence.
		4. Diagnostic Yield/Performance This category focuses on studies that evaluate the technical performance and accuracy of an AI model as an independent or assistant diagnostic tool. The primary objective is to measure the model's ability to detect or classify disease, typically quantified by metrics such as sensitivity, specificity, and detection rates. Examples include using AI to improve polyp detection rates in colonoscopy or to identify abnormalities on radiological images.

Reporting Standard	For RCTs only. The specific reporting guideline the study claims to have followed.	Free text. CONSORT 2010, CONSORT-AI, Not Specified.
Primary Result	For RCTs only. A concise summary of the study's main finding as it relates to the primary outcome.	1. Significant Improvement 2. No Significant Effect 3. Demonstrated Noninferiority 4. Significant Deterioration
Comparison	For RCTs only. The control or standard practice against which the AI intervention was evaluated.	1. Assisted AI vs. Unassisted AI: Compares a clinician aided by an AI tool to a fully autonomous AI system. 2. AI vs. Routine Care: Compares an AI-driven intervention to the established standard of care without AI. 3. AI vs. Clinician: Compares an independent AI system to a human clinician performing the same task without AI.

Table S2. Summary of Potential Reasons for RCTs with Non-significant Outcomes

Study (First Author, Year)	Clinical Domain	1. HCI & Workflow Integration Issues	2. AI Model Performance Limitations	3. Study Design & Endpoint Selection Issues	4. Insufficient Statistical Power	5. Strong External Confounding Factors	6. Intervention Implementation Variability
Blomberg, 2021	Out-of-Hospital Cardiac Arrest	✓	✓				
Hwang, 2023	Emergency Chest Radiography	✓	✓	✓	✓	✓	
Maheshwari, 2020	Intraoperative Hypotension	✓	✓	✓			
Schenk, 2021	Postoperative Hypotension			✓	✓	✓	
Kumar, 2020	Clinical Order Entry	✓		✓			
Agarwal, 2019	Diabetes Self-Management App	✓		✓	✓		✓
Wang, 2019	Atrial Fibrillation Anticoagulation	✓	✓			✓	
Xu, 2021	Colonoscopy Polyp Detection		✓	✓			✓
Pavel, 2020	Neonatal Seizure Recognition		✓	✓			
Lui, 2023	Colonoscopy Polyp Miss Rate		✓	✓	✓		
Seol, 2021	Childhood Asthma Management	✓		✓	✓		
Luna, 2021	Physical Therapy (Squat Evaluation)		✓	✓	✓		

Lin, 2019	Childhood Cataract Diagnosis	✓	✓		✓	
Vilkoite, 2023	Colonoscopy Neoplasm Diagnosis			✓	✓	
Karsenti, 2023	Colonoscopy Polyp Detection	✓		✓		✓
Marcuzzi, 2023	Neck/Low Back Pain App	✓		✓		✓
Yamaguchi, 2024	Colonoscopy Training			✓	✓	
Wei, 2023	Colonoscopy (Community Setting)	✓	✓	✓	✓	
Mangas-Sanjuan, 2023	Colonoscopy Advanced Neoplasia		✓	✓		
Yuan, 2023	Esophageal Cancer Diagnosis			✓	✓	

Table S3. Reporting Guideline Use Among Included RCTs Overall and by Publication Period

Reporting Standard	Overall (n=113)	Pre-2021	2021–2023
Not Specified	81 (71.7%)	39 (88.6%)	42 (60.9%)
CONSORT (2010)	23 (20.4%)	5 (11.4%)	18 (26.1%)
CONSORT-AI	6 (5.3%)	NA	6 (8.7%)
CONSORT-AI + SPIRIT-AI	2 (1.8%)	NA	2 (2.9%)
CONSORT-eHealth	1 (0.9%)	NA	1 (1.4%)

Note: Publication period was stratified at 2021 because CONSORT-AI was published in October 2020; trials published in 2021 or later were therefore considered to have had the opportunity to reference AI-specific reporting guidance. NA, not applicable.

Table S4. Distribution of ROBUST-RCT Risk-of-Bias Judgments for Each Assessed Item Across 113 Randomized Controlled Trials

ROBUST-RCT Item	Definitely low	Probably low	Probably high	Definitely high
Core items				
Item 1: Random sequence generation	34 (30.1%)	64 (56.6%)	15 (13.3%)	0 (0%)
Item 2: Allocation concealment	4 (3.5%)	24 (21.2%)	85 (75.2%)	0 (0%)
Item 3: Blinding of participants	26 (23%)	5 (4.4%)	82 (72.6%)	0 (0%)
Item 4: Blinding of healthcare providers	4 (3.5%)	11 (9.7%)	97 (85.8%)	1 (0.9%)
Item 5: Blinding of outcome assessors	50 (44.2%)	28 (24.8%)	33 (29.2%)	2 (1.8%)
Item 6: Outcome data not included in analysis	56 (49.6%)	40 (35.4%)	11 (9.7%)	6 (5.3%)
Optional items				
Item 7: Balance of baseline prognostic factors	58 (51.3%)	48 (42.5%)	7 (6.2%)	0 (0%)
Item 8: Balance of co-interventions in blinded trials	4 (3.5%)	101 (89.4%)	8 (7.1%)	0 (0%)
Item 9: Different outcome assessment between groups	58 (51.3%)	52 (46%)	2 (1.8%)	1 (0.9%)
Item 10: Different follow-up between groups	30 (26.5%)	74 (65.5%)	8 (7.1%)	1 (0.9%)
Item 11: Validity of outcome measurement methods	49 (43.4%)	63 (55.8%)	1 (0.9%)	0 (0%)
Item 12: As-treated analysis concerns	0 (0%)	113 (100%)	0 (0%)	0 (0%)
Item 13: Selective outcome reporting	0 (0%)	96 (85%)	17 (15%)	0 (0%)
Item 14: Early trial termination for benefit	3 (2.7%)	109 (96.5%)	1 (0.9%)	0 (0%)

Note: Percentages were calculated using 113 included randomized controlled trials as the denominator for each item. ROBUST-RCT includes 6 core items (Items 1–6) and 8 optional items (Items 7–14).

Abbreviation: ROBUST-RCT, Risk Of Bias instrument for Use in SysTematic reviews—for Randomized Controlled Trials.

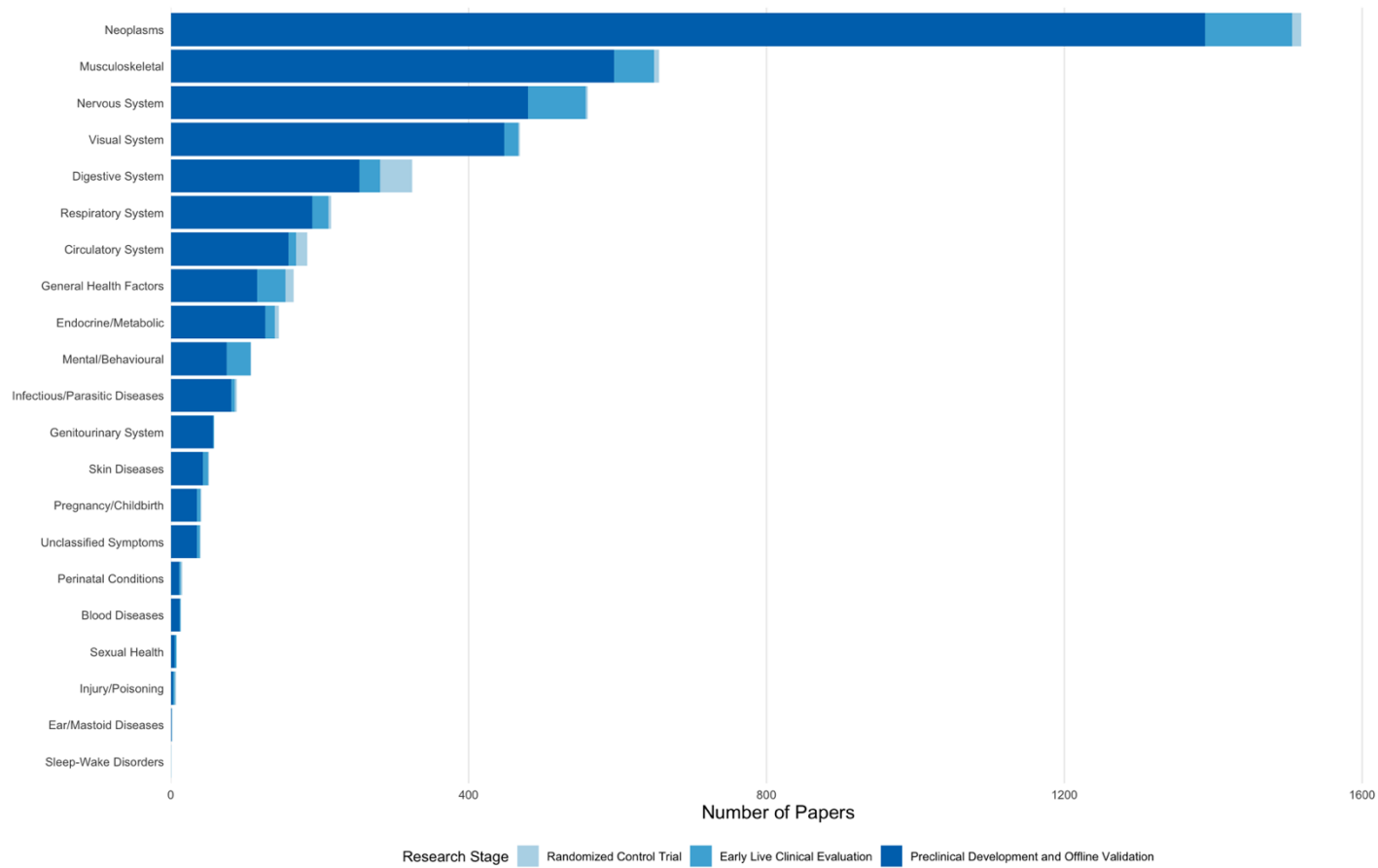
Table S5. Database-specific search strategies

Database	Search strategy
PubMed	(“Artificial Intelligence”[Mesh:noexp] OR “artificial intelligence”[tiab] OR “computational intelligence”[tiab] OR “machine intelligence”[tiab] OR “computer vision”[tiab] OR “computer reasoning”[tiab] OR “automated reasoning”[tiab] OR “computer aid*”[tiab] OR “computer assist*”[tiab] OR “machine learning”[tiab] OR “deep learning”[tiab] OR “hierarchical learning”[tiab] OR “support vector machine*”[tiab] OR “support vector network*”[tiab] OR “Neural Networks, Computer”[Mesh] OR “neural network*”[tiab] OR perceptron*[tiab] OR “Natural Language Processing”[Mesh] OR “natural language processing*”[tiab] OR “computation analysis”[tiab] OR “real time prediction”[tiab]) AND (“Physicians”[Mesh] OR “Nurses”[Mesh] OR “Internship and Residency”[Mesh] OR “Physical Therapists”[Mesh] OR “Anesthetists”[Mesh] OR “Audiologists”[Mesh] OR “Psychotherapists”[Mesh] OR “Pharmacists”[Mesh] OR “board certified”[tiab] OR expert*[tiab] OR surgeon*[tiab] OR clinician*[tiab] OR physician*[tiab] OR doctor*[tiab] OR nurse*[tiab] OR fellow*[tiab] OR residen*[tiab] OR attending*[tiab] OR specialist*[tiab] OR practitioner*[tiab] OR hospitalist*[tiab] OR anesthesiologist*[tiab] OR anaesthetist*[tiab] OR audiologist*[tiab] OR cardiologist*[tiab] OR dermatologist*[tiab] OR endocrinologist*[tiab] OR gastroenterologist*[tiab] OR geriatrician*[tiab] OR gynecologist*[tiab] OR haematologist*[tiab] OR histopathologist*[tiab] OR immunologist*[tiab] OR internist*[tiab] OR intensivist*[tiab] OR nephrologist*[tiab] OR neurologist*[tiab] OR oncologist*[tiab] OR ophthalmologist*[tiab] OR orthopedic*[tiab] OR radiologist*[tiab] OR psychiatrist*[tiab] OR psychologist*[tiab] OR therapist*[tiab]) AND (“systematic review”[pt] OR “systematic review”[tiab] OR “meta-analysis”[pt] OR “meta-analysis”[tiab])
Embase	(‘artificial intelligence’/exp OR ‘artificial intelligence’:ti,ab OR ‘computational intelligence’:ti,ab OR ‘machine intelligence’:ti,ab OR ‘computer vision’:ti,ab OR ‘computer reasoning’:ti,ab OR ‘automated reasoning’:ti,ab OR ‘computer aid*’:ti,ab OR ‘computer assist*’:ti,ab OR ‘machine learning’:ti,ab OR ‘deep learning’:ti,ab OR ‘hierarchical learning’:ti,ab OR ‘support vector machine*’:ti,ab OR ‘neural network’/exp OR ‘neural network*’:ti,ab OR ‘perceptron*’:ti,ab OR ‘natural language processing’/exp OR ‘natural language processing*’:ti,ab OR ‘computation analysis’:ti,ab OR ‘real time prediction’:ti,ab) AND (‘physician’/exp OR ‘nurse’/exp OR ‘internship’/exp OR ‘residency’/exp OR ‘physical therapist’/exp OR ‘anesthetist’/exp OR ‘audiologist’/exp OR ‘psychotherapist’/exp OR ‘pharmacist’/exp OR ‘board certified’:ti,ab OR expert*:ti,ab OR surgeon*:ti,ab OR clinician*:ti,ab OR physician*:ti,ab OR doctor*:ti,ab OR nurse*:ti,ab OR fellow*:ti,ab OR residen*:ti,ab OR attending*:ti,ab OR specialist*:ti,ab OR practitioner*:ti,ab OR hospitalist*:ti,ab OR anesthesiologist*:ti,ab OR anaesthetist*:ti,ab OR audiologist*:ti,ab OR cardiologist*:ti,ab OR dermatologist*:ti,ab OR endocrinologist*:ti,ab OR gastroenterologist*:ti,ab OR geriatrician*:ti,ab OR gynecologist*:ti,ab OR

	haematologist*:ti,ab OR immunologist*:ti,ab OR internist*:ti,ab OR intensivist*:ti,ab OR nephrologist*:ti,ab OR neurologist*:ti,ab OR oncologist*:ti,ab OR ophthalmologist*:ti,ab OR orthopedist*:ti,ab OR radiologist*:ti,ab OR psychiatrist*:ti,ab OR psychologist*:ti,ab OR therapist*:ti,ab) AND ('systematic review'/exp OR 'meta analysis'/exp OR 'systematic review':ti,ab OR 'meta- analysis':ti,ab)
Cochrane Library	("Artificial Intelligence" OR "computational intelligence" OR "machine intelligence" OR "computer vision" OR "computer reasoning" OR "automated reasoning" OR "computer aid*" OR "computer assist*" OR "machine learning" OR "deep learning" OR "hierarchical learning" OR "support vector machine*" OR "support vector network*" OR "neural network*" OR "perceptron*" OR "Natural Language Processing" OR "natural language processing*" OR "computation analysis" OR "real time prediction") AND ("Physicians" OR "Nurses" OR "Internship and Residency" OR "Physical Therapists" OR "Anesthetists" OR "Audiologists" OR "Psychotherapists" OR "Pharmacists" OR "board certified" OR expert* OR surgeon* OR clinician* OR physician* OR doctor* OR nurse* OR fellow* OR residen* OR attending* OR specialist* OR practitioner* OR hospitalist* OR anesthesiologist* OR anaesthetist* OR audiologist* OR cardiologist* OR dermatologist* OR endocrinologist* OR gastroenterologist* OR geriatrician* OR gynecologist* OR haematologist* OR histopathologist* OR immunologist* OR internist* OR intensivist* OR nephrologist* OR neurologist* OR oncologist* OR ophthalmologist* OR orthopedic* OR radiologist* OR psychiatrist* OR psychologist* OR therapist*) AND ("systematic review" OR "meta-analysis")

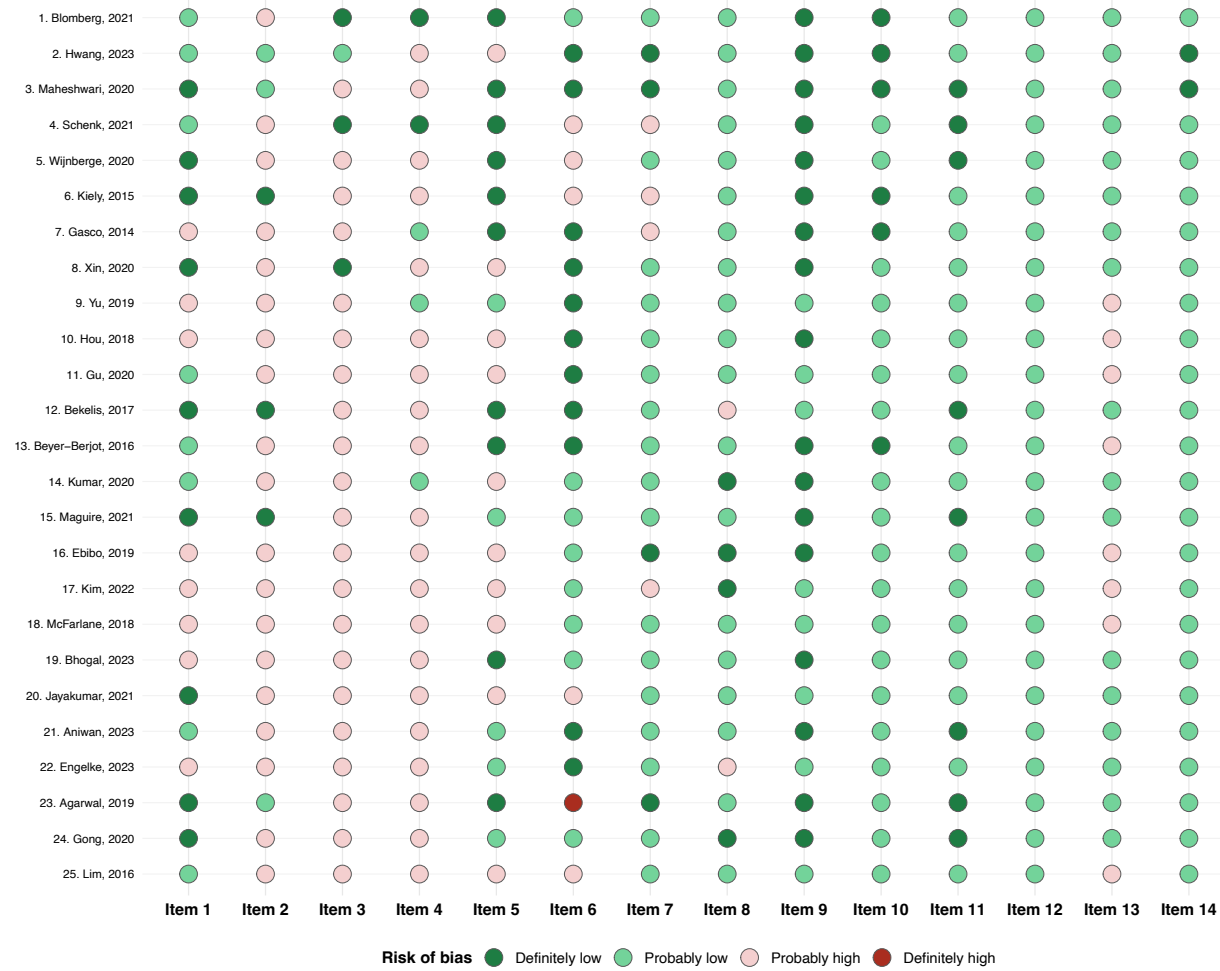
Note: Searches were conducted from September 20, 2023 to September 20, 2024.

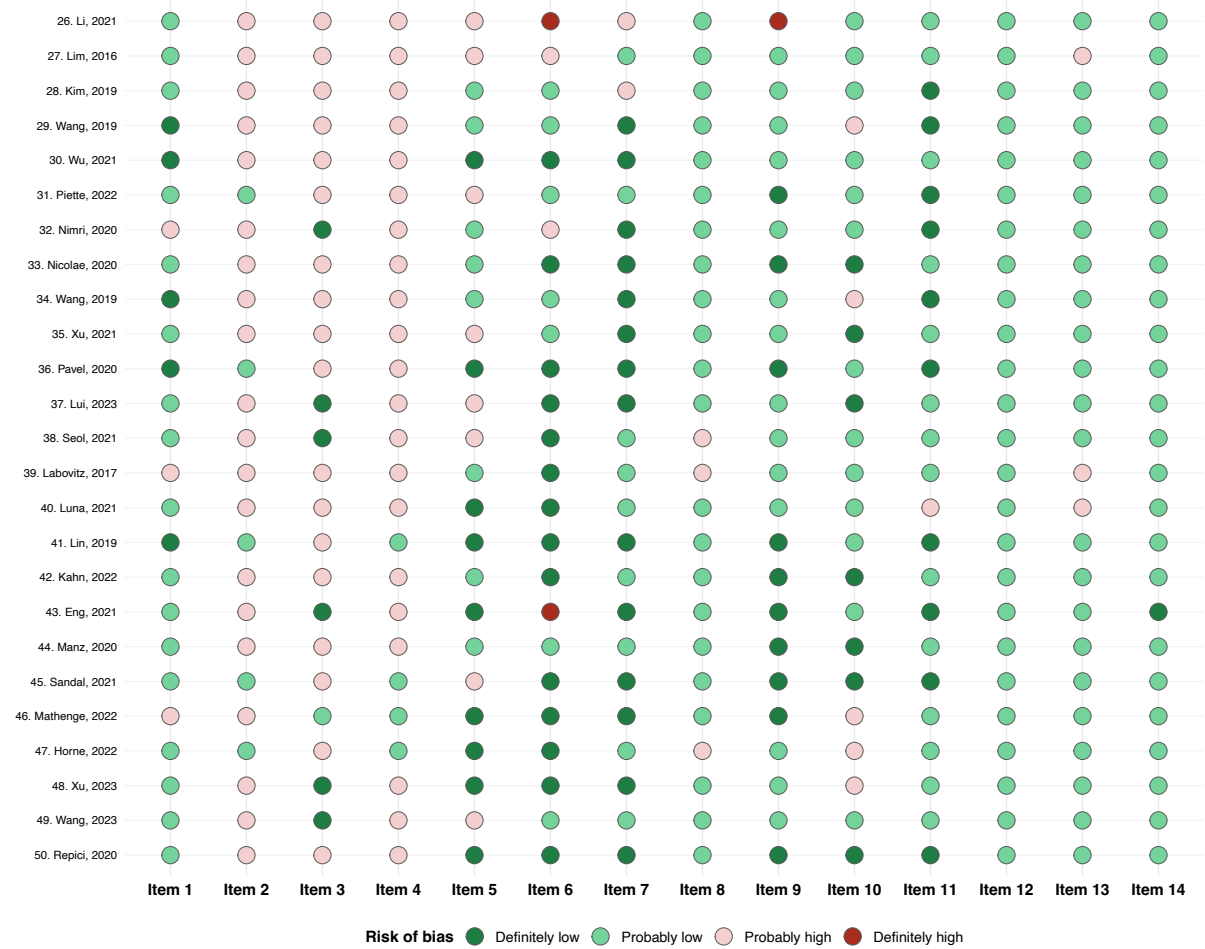
Figure S1. Distribution of AI Medical Studies by Medical Specialty

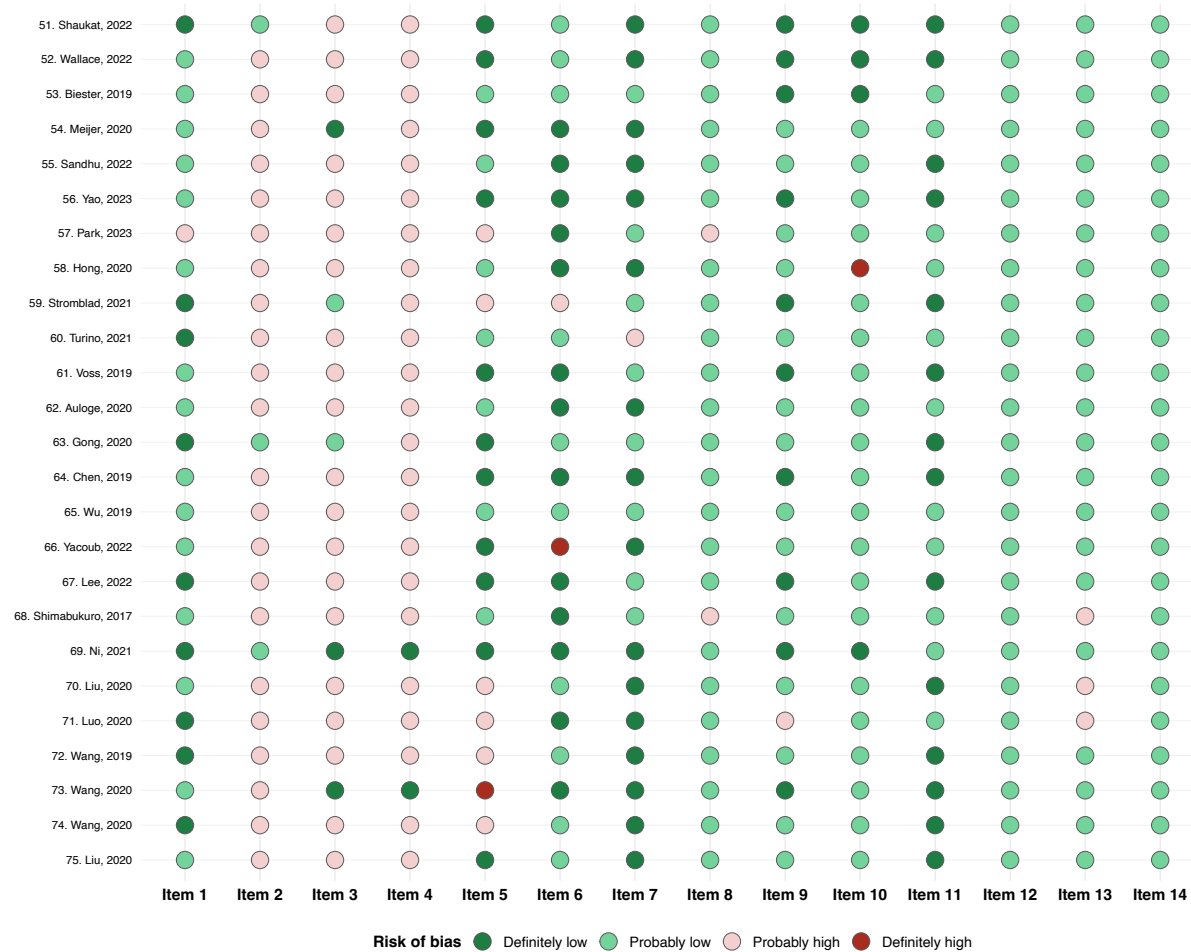


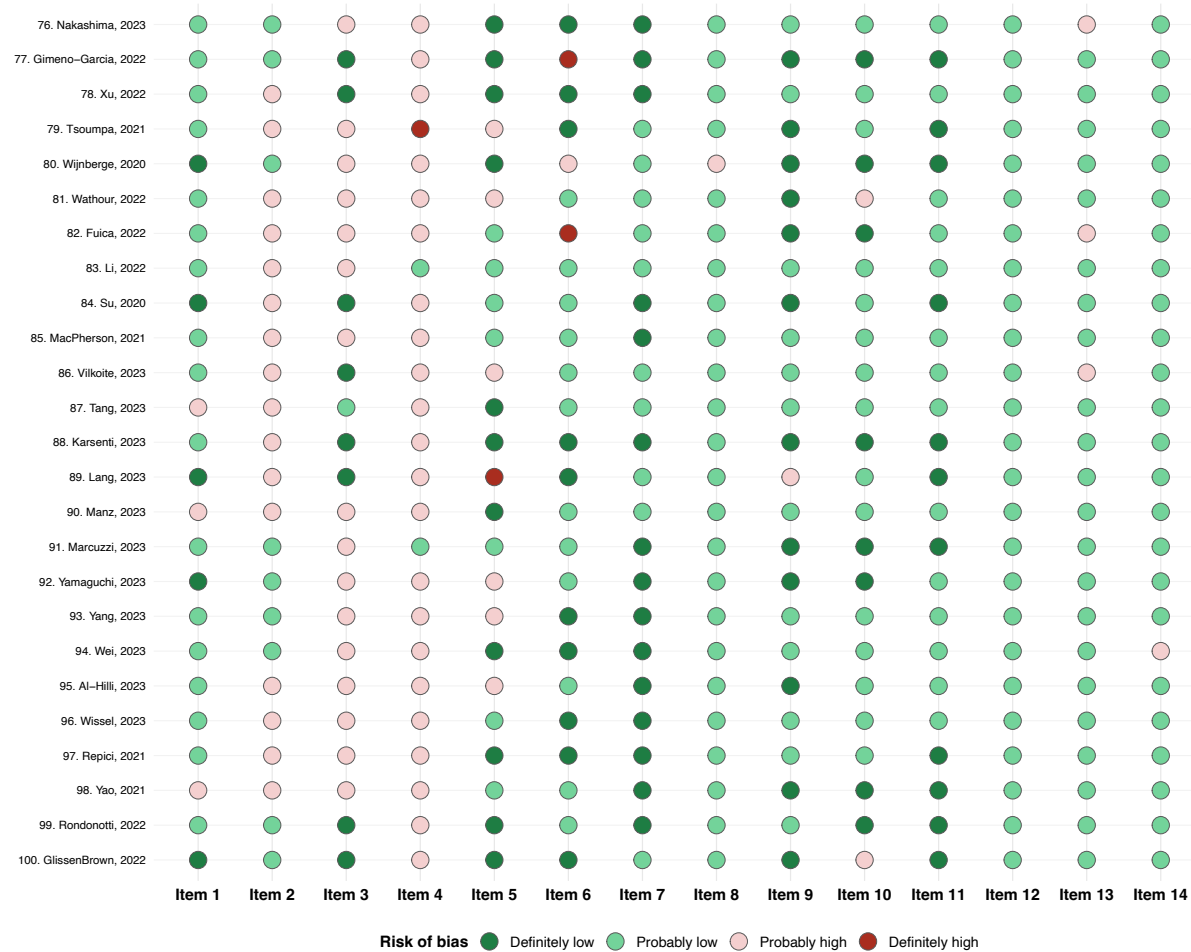
The figure shows the number of included studies across ICD-11 medical specialties, stratified by research stage. Research stages include preclinical development and offline validation, early live clinical evaluation, and randomized controlled trial.

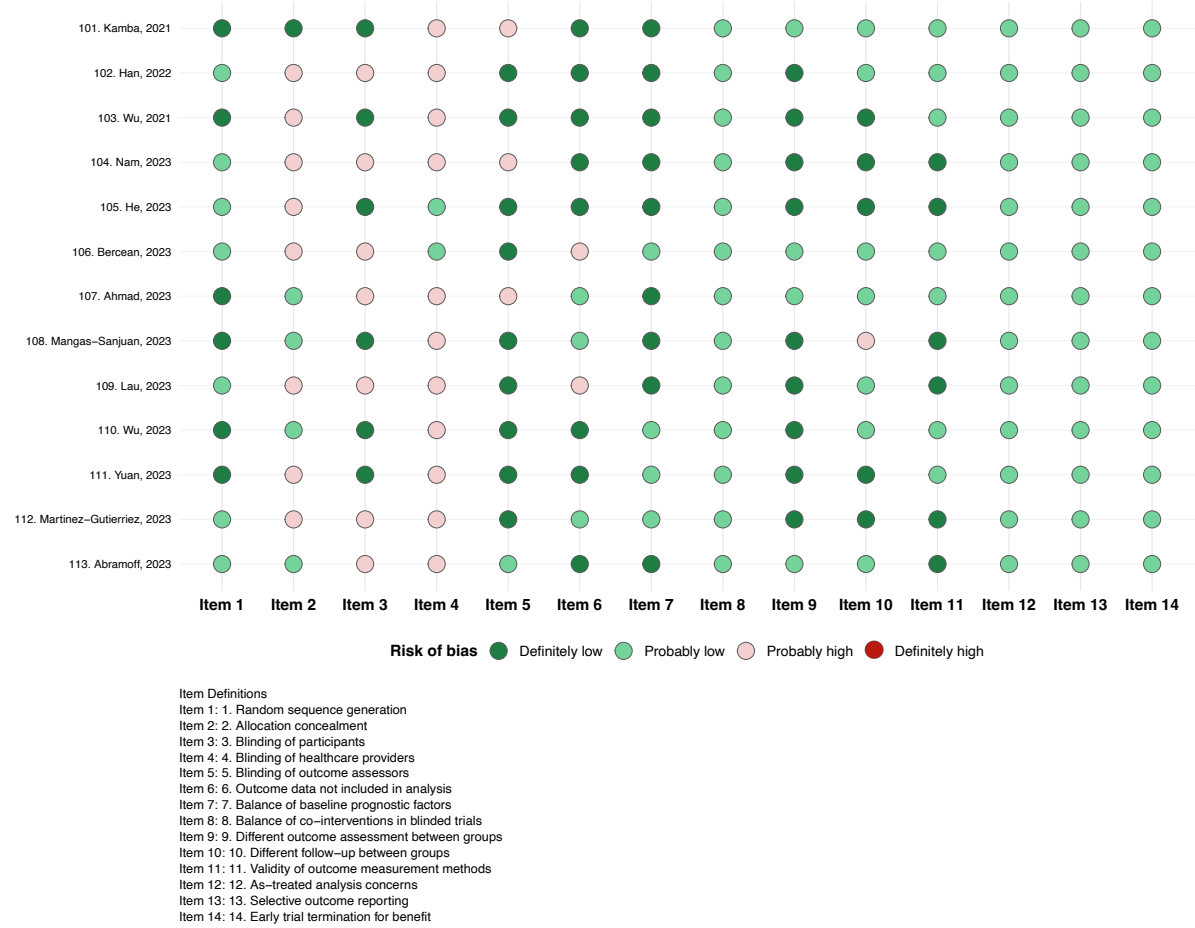
Figure S2. Traffic-Light Plot of ROBUST-RCT Risk-of-Bias Judgments Across 113 Randomized Controlled Trials











Each row represents an individual trial, labeled by study number, first author surname, and publication year, and each column represents a ROBUST-RCT item. Colors indicate risk-of-bias judgments (definitely low, probably low, probably high, definitely high). Items 1-6 are core items and Items 7-14 are optional items.

Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist

Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
TITLE			
Title	1	Identify the report as a scoping review.	1
ABSTRACT			
Structured summary	2	Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach.	3
Objectives	4	Provide an explicit statement of the questions and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives.	4
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number.	13
Eligibility criteria	6	Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale.	13-14; Table S1
Information sources*	7	Describe all information sources in the search (e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed.	14; Figure 1
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	Table S5
Selection of sources of evidence†	9	State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review.	14
Data charting process‡	10	Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators.	14-15
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made.	Table S1
Critical appraisal of individual sources of evidence§	12	If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate).	15
Synthesis of results	13	Describe the methods of handling and summarizing the data that were charted.	16



St. Michael's
Inspired Care.
Inspiring Science.

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
RESULTS			
Selection of sources of evidence	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram.	4
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	4
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12).	8
Results of individual sources of evidence	17	For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives.	4-8
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives.	7
DISCUSSION			
Summary of evidence	19	Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups.	8-12
Limitations	20	Discuss the limitations of the scoping review process.	13
Conclusions	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps.	12
FUNDING			
Funding	22	Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review.	17

JB1 = Joanna Briggs Institute; PRISMA-ScR = Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews.

* Where *sources of evidence* (see second footnote) are compiled from, such as bibliographic databases, social media platforms, and Web sites.

† A more inclusive/heterogeneous term used to account for the different types of evidence or data sources (e.g., quantitative and/or qualitative research, expert opinion, and policy documents) that may be eligible in a scoping review as opposed to only studies. This is not to be confused with *information sources* (see first footnote).

‡ The frameworks by Arksey and O'Malley (6) and Levac and colleagues (7) and the JBI guidance (4, 5) refer to the process of data extraction in a scoping review as data charting.

§ The process of systematically examining research evidence to assess its validity, results, and relevance before using it to inform a decision. This term is used for items 12 and 19 instead of "risk of bias" (which is more applicable to systematic reviews of interventions) to include and acknowledge the various sources of evidence that may be used in a scoping review (e.g., quantitative and/or qualitative research, expert opinion, and policy document).

From: Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. ;169:467–473. doi: 10.7326/M18-0850



St. Michael's
Inspired Care.
Inspiring Science.