Review

Stakeholder Acceptance, Trust, and Implementation of AI in Oncology: Mixed Methods Systematic Review

Sebastian Fitzek[1], PhD; Birgit Pohn[2,3], BSc, DI, MSc; Alisa Mohebbi[2], MD; Ali Abbasian Ardakani[2], PhD; Kyung-Eun (Anna) Choi[1], PhD; Sepideh Hatamikia[2,4], PhD

[1]Research Center for Medical Image Analysis and Artificial Intelligence, Department of Medicine, Faculty of Medicine and Dentistry, Danube Private University, Krems an der Donau, Lower Austria, Austria

[2]Department of Medicine, Faculty of Medicine and Dentistry, Danube Private University, Krems an der Donau, Lower Austria, Austria

[3]Faculty Computer Science and Applied Mathematics, University of Applied Sciences Technikum Wien, Vienna, Vienna, Austria

[4]Austrian Center for Medical Innovation and Technology (ACMIT), Wiener Neustadt, Lower Austria, Austria

Corresponding Author:
Sebastian Fitzek, PhD
Research Center for Medical Image Analysis and Artificial Intelligence, Department of Medicine
Faculty of Medicine and Dentistry
Danube Private University
Steiner Landstraße 124
Krems an der Donau, Lower Austria, 3500
Austria
Phone: 43 676 842 419 305
Email: sebastian.fitzek@dp-uni.ac.at

Abstract

Background: AI is entering oncology workflows faster than evidence on stakeholder acceptance, trust, and sustained use. Existing syntheses rarely distinguish anticipated attitudes from actual use across oncology stakeholders.

Objective: The aim of the study is to synthesize primary empirical evidence on stakeholder acceptance, trust, and implementation of AI in oncology and examine variation by stakeholder group, clinical task, AI exposure, and implementation maturity.

Methods: We conducted a Joanna Briggs Institute–guided mixed methods systematic review with convergent integrated synthesis; the protocol was registered before formal eligibility screening. Eight databases were searched without date or language limits (August 6-9, 2026), yielding 46,206 records and 31,802 unique records. Title and abstract screening used 2 balanced primary allocations, joint bidirectional readjudication, and blinded verification of a stratified 10% (n=3182) sample. Full-text assessment, extraction, and critical appraisal were dual independent. Acceptance/Trust modes and F1-F8 coverage were independently coded across 309 studies and reconciled. Eight integrated findings were anchored to the nonadoption, abandonment, scale-up, spread, and sustainability framework; an emergent economic finding was reported separately. Adjusted stage models controlled for study design, publication format, and publication period; robustness analyses examined screening provenance and appraisal sensitivity.

Results: We included 309 primary empirical studies; 187 (60.5%) were published in 2025 or 2026. Exposure ranged from hypothetical opinion to real-world implementation (n=82 studies), but only 6 described routine use. Acceptance was empirically assessed in 267 studies and Trust in 127. Relative to anticipatory stage 1, Acceptance assessment was less frequent in evaluated-interaction stage 2 (adjusted odds ratio [aOR] 0.07, 95% CI 0.02-0.25; P<.001) and stages 3+4 (aOR 0.12, 95% CI 0.03-0.45; P=.001). Transparency/explainability coverage (F3) was lower in stages 3+4 (aOR 0.20, 95% CI 0.08-0.47; q<0.001). Workflow/interoperability coverage (F5) was higher and governance coverage (F7) lower in both later contrasts after false discovery rate correction. Trust declined descriptively across stages, but adjusted contrasts were nonsignificant. Eight integrated findings spanned clinical value, performance, transparency, oversight, workflow, readiness, governance, and implementation, plus an emergent stage 1 economic finding. The principal Acceptance, F3, F5, and F7 findings persisted in the 229-study original screening subset.

Conclusions: In oncology, the AI evidence profile appears to differ from anticipated use to implementation. Acceptance was empirically assessed most often before deployment; later-stage studies more often examined workflow and less often examined transparency and governance. The challenge is whether favorable attitudes and early deployment translate into calibrated, governed, and sustained routine use. An interpretive stage-specific model identifies assessment priorities across exposure and implementation

XSL•FO
RenderX

contexts, but is not a validated causal or predictive model. Later-stage evaluations should retain measurement of Acceptance, trust calibration, governance, durability, and workflow outcomes.

Trial Registration: OSF Registries osf.io/nmep8; https://osf.io/nmep8

(J Med Internet Res 2026;28:e101445) doi: 10.2196/101445

KEYWORDS

AI; oncology; acceptance; trust; implementation; mixed methods systematic review; clinical decision support; explainable AI; human factors

Introduction

AI is moving from model development into clinically consequential oncology workflows, including screening, imaging, pathology, radiotherapy planning, treatment decision support, supportive care, and patient communication [1]. Technical performance is necessary but not sufficient for clinical impact; evaluations repeatedly describe a gap between accurate models and systems that are adopted, used safely, and sustained in practice [2]. Implementation research therefore treats adoption as a sociotechnical process in which the condition, the technology, its value proposition, adopters, organizations, and the wider system interact and change over time, as formalized in the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework [3]. Oncology intensifies these interactions because decisions are high-stakes, preference-sensitive, multidisciplinary, and embedded in long care relationships, and oncology-specific frameworks have accordingly emphasized ethical deployment, accountability, and lifecycle governance [4,5].

Acceptance has most often been studied through technology acceptance models, in which perceived usefulness and ease of use, or performance expectancy, effort expectancy, social influence, and facilitating conditions, predict the intention to use a technology [6,7]. These models were designed to explain intention, and intention is an imperfect predictor of sustained behavior [8]. Stated willingness to use a hypothetical AI system therefore says little about whether a deployed system will be used, overridden, or abandoned, which is why implementation outcomes such as adoption, fidelity, and sustainability need to be measured in their own right [9].

Trust is a distinct construct. Human factors research defines appropriate trust as trust calibrated to the actual capabilities of an automated system and shows that both distrust and overtrust impair performance [10]. Complacency and automation bias are well documented, including in clinical decision support [11,12], and clinicians' trust in AI depends on reliability, explanation, and accountability as well as on accuracy [13]. Greater subjective trust can therefore be unsafe when it is not matched by system reliability. Acceptance and Trust should consequently be reported separately, and Trust should, where possible, be observed as reliance behavior and not inferred solely from attitudes.

Several syntheses define the adjacent evidence base. Reviews across health care have described patient and public attitudes toward clinical AI [14], health professionals' acceptance of AI in hospitals [15], and barriers and facilitators to patient acceptance [16]. Hogg et al [17] synthesized qualitative stakeholder perspectives on clinical AI implementation, Kuo et al [18] coproduced a qualitative evidence synthesis on diagnostic AI [18], and Tun et al [19] reviewed health care workers' trust in AI-based decision support. Oncology-specific reviews have been narrower in scope, addressing patient attitudes toward AI in cancer care [20], implementation challenges in breast cancer screening programs [21], AI in palliative and supportive cancer care [22], and empathic AI for patient-centered cancer care [23], while Wenderott et al [24] examined routine AI implementation in medical imaging. Together, these reviews establish the importance of trust, oversight, workflow fit, and governance. Among the syntheses compared here, none integrates quantitative and qualitative primary evidence across oncology stakeholder groups and clinical tasks while also separating anticipated attitudes from direct interaction and early deployment from sustained routine use. Most also predate the rapid growth of this literature since 2024, including studies of generative AI.

The present review addresses that gap in 5 ways. First, empirical denominators are restricted to primary studies; reviews and frameworks are used only for comparison. Second, the synthesis is multistakeholder and spans oncology tasks, from screening and imaging through pathology, radiotherapy planning, treatment decision support, supportive care, and patient communication. Third, exposure type and implementation maturity were coded as separate, prespecified axes and grouped post hoc into 4 analytic stages: anticipated use, evaluated interaction, local implementation, and routine use. Fourth, construct measurement and methodological quality are examined alongside the findings, which are anchored to NASSS. Fifth, and central to the oncology contribution, the review asks whether the questions studied change as a tool moves from a hypothetical cancer scenario to interaction and local deployment. That shift may be clinically consequential in oncology because recommendations can affect recall, contouring, diagnostic sign-out, multidisciplinary recommendations, and preference-sensitive treatment choices, often inside long care relationships and with limited opportunity for reversal. The resulting stage pattern is then translated into an interpretive clinical model for what a service might assess at each stage. This model is not a validated implementation score and does not imply that stages form a required sequence.

The registered objectives were to identify the reported determinants of acceptance, trust, use, and implementation in primary empirical studies; to examine how findings differ across stakeholder groups, oncology tasks, exposure types, and implementation stages; and to map the constructs and

XSL•FO
RenderX

instruments used, appraise study quality, and determine where evidence is consistent, uncertain, or absent.

Methods

Review Design and Reporting

We conducted a Joanna Briggs Institute (JBI)–guided mixed methods systematic review with convergent integrated synthesis, transforming relevant quantitative results into textual findings for integration with qualitative evidence [25]. Reporting followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 and PRISMA-S (PRISMA Extension for Literature Searches) [26,27]. Completed PRISMA 2020 and PRISMA-S reporting checklists are provided in Multimedia Appendices 1 and 2, respectively. Meta-analysis was not planned because stakeholders, AI tasks, exposures, constructs, outcomes, and implementation maturity were too heterogeneous for a defensible pooled effect.

Protocol, Registration, and Eligibility Criteria

The protocol specified eligibility, screening, appraisal, and synthesis and was registered on Open Science Framework on August 10, 2026 [28]. Database searches and cross-database deduplication preceded registration, but formal eligibility screening had not begun. Because the protocol had specified registration before deduplication, this timing is reported as a deviation. Protocol version 1.2 and methodological documentation are archived with the registration.

Studies were eligible when they met all of the following criteria: (1) concerned an oncology context, including prevention, screening, diagnosis, prognosis, treatment, radiotherapy, monitoring, supportive care, communication, or oncology workflow; (2) included human stakeholders such as patients, survivors, caregivers, members of the public directly concerned, clinicians, oncology staff, implementers, managers, or other relevant participants; (3) reported primary empirical research; (4) addressed acceptance, trust, attitudes, preferences, willingness to use, readiness, usability, experience, workflow, adoption, abandonment, governance, or implementation barriers or facilitators; and (5) explicitly concerned AI, machine learning, deep learning, natural language processing, large language models, computer vision, or another inference system presented by the authors as AI.

We excluded purely technical performance studies without stakeholder outcomes, reviews/frameworks, protocols, editorials/commentaries, and studies without a separable oncology context. Secondary literature was contextual only and never entered empirical denominators. Conference material required sufficient empirical information for extraction and appraisal; unvalidated preprints were excluded. No date or language filter was applied, and no dedicated gray-literature repository was searched.

Information Sources and Search Strategy

We searched PubMed, Embase, CINAHL Complete, APA PsycINFO, Scopus, Web of Science Core Collection, IEEE Xplore, and the ACM Guide to Computing Literature on August 6-9, 2026. The searches yielded 46,206 records and combined oncology, AI/clinical decision support, human stakeholders, acceptance/trust, and broader implementation, workflow, ethics, and governance concepts. PubMed underwent PRESS (Peer Review of Electronic Search Strategies) peer review [29]. Prespecified known-item checks recovered all retrievable sentinels in PubMed (12/12), Scopus (13/13), and Web of Science (8/8), and all indexed sentinels tested in Embase. Full strategies, interfaces, dates, source counts, and validation records are in Multimedia Appendix 3.

Deduplication and Study Selection

Records were conservatively deduplicated using corroborated DOI/title evidence, with conflicts kept separate. Of 2918 candidate pairs reviewed, 1414 were merged and 1504 retained separately; 14,404 duplicates were removed, leaving 31,802 records. Under protocol version 1.2, both reviewers (SF and BP) independently screened a shared 200-record calibration stratum and resolved it by consensus; the remaining 31,602 records were split into 2 balanced, nonoverlapping allocations of 15,801 for one primary human decision each. The first pass retained 388 records. An additional joint human quality-control and bidirectional readjudication layer recovered 76 initially excluded records and re-examined the resulting 464-record pool; 70 remained reinstated, and 99 initially retained records were excluded, leaving 359. A reproducible stratified 10% (n=3182) sample, disjoint from calibration, was then blind-verified by the other reviewer. Against original first-pass decisions, retain/exclude agreement was 99.15% (Cohen κ=0.72). Among 3120 sampled first-pass exclusions outside the jointly reviewed pool, the verifier retained 3 (0.10%, 95% CI 0.03%-0.28%). All verification disagreements were reconciled.

All retained reports underwent dual independent full-text assessment and reconciliation. Of 321 exact primary reports assessed at this stage, 13 had been dual-finalized earlier, and 308 formed the main batch, with 98.1% (302/308) agreement before reconciliation (κ=0.82). In total, 10 reports opened during the later screening remediation also underwent dual assessment with complete agreement. Multiple reports from one study were linked to avoid double counting; one primary exclusion reason was retained per excluded report (Multimedia Appendix 4).

Exposure Type and Implementation Maturity

Exposure and implementation maturity were coded separately. Hypothetical opinion denoted attitudes without interaction with a functioning system; information/vignette, responses to a description, scenario, demonstration, or presented output without hands-on use; prototype/simulation, interaction with a nonproduction system or simulated workflow; controlled evaluation, structured evaluation of a functioning system under study conditions; and real-world implementation, use within an actual clinical or organizational workflow. Maturity was development, validation, pilot, implementation, or routine use; routine use required established operation beyond a bounded pilot/evaluation. The 4-stage grouping was post hoc: routine use defined stage 4; among remaining studies, hypothetical opinion/information-vignette formed stage 1, prototype/simulation/controlled evaluation stage 2, and real-world implementation stage 3. Underlying exposure and

XSL•FO
RenderX

maturity variables remained separate, and stages were analytic strata rather than a required sequence.

Data Extraction and Critical Appraisal

Two reviewers piloted the extraction form on 10 diverse studies, reconciled the codebook, and independently extracted the remaining 295 primary-workflow studies; the 4 remediation studies later underwent the same dual procedure, yielding 309 studies. Extracted domains covered study context, participants and stakeholders, cancer task, AI/workflow role, exposure and maturity, sample and response rate, Acceptance, Trust, implementation domains, instruments, quantitative/qualitative findings, direction, funding/conflicts, and source traceability (Multimedia Appendix 5).

All revised-review extraction, eligibility, appraisal, thematic/construct coding, synthesis, and sensitivity judgments were human. Codex assisted only with Python script development; scripts executed locally in PyCharm supported deterministic data handling, sampling/allocation, blinded materials, and quality assurance, and, after reviewer coding, analytic tables and statistics. No revised-review extraction field or scientific judgment was assigned programmatically.

Methodological quality was independently appraised and reconciled. Mixed Methods Appraisal Tool (MMAT) 2018 was applied criterion by criterion to 307 studies [30]; one instrument-validation study used an adapted Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) Risk of Bias approach [31], and one other study used a prespecified fit-for-purpose appraisal. Prereconciliation agreement was 2558 of 2739 (93.4%) applicable ratings; all 181 disagreements were resolved. Ratings are source-traceable in Multimedia Appendix 5. No global MMAT score was calculated. Because no appraisal threshold was prespecified, 3 post hoc robustness subsets used applicable criteria only: no criterion rated No, at least 80% Yes, and at least 60% Yes. These were sensitivity operationalizations, not validated quality grades or exclusion rules.

Stakeholder and Construct Coding

Stakeholder coding was multilabel and referred to participants actually studied. SF coded the 305 primary-workflow studies; the 4 remediation studies were dual-coded, and BP blindly verified a stratified 50-study sample (82% exact stakeholder-set agreement). Both reviewers independently coded Acceptance and Trust across all 309 studies. Initial agreement was 61.5% (κ=0.406) and 73.8% (κ=0.589), respectively; all disagreements were then independently readjudicated blind, with residual consensus after agreement of 83.2% (κ=0.781) for Acceptance and 84% (κ=0.775) for Trust in those disagreement subsets. Acceptance was empirical when measured or explicitly elicited from participants; Trust required measured/elicited trust, reliance, or confidence, including observed reliance. Inference or author discussion alone was coded separately. Throughout, capitalized Acceptance and Trust denote these coded constructs; lowercase acceptance and trust denote the general concepts. Wilson CIs described stage proportions. Exploratory adjusted logistic models compared stage 2 and pooled stages 3+4 with stage 1, adjusting for study design, publication format, and publication period; Acceptance and Trust were modeled separately.

Convergent Integrated Synthesis

For convergent integration, relevant quantitative results were transformed into contextualized textual evidence and integrated with qualitative findings; no pooled effect was generated. NASSS was the prespecified implementation framework [3]. A purposive set of 47 studies yielded 57 evidence pairs across 8 candidate findings; BP independently verified all findings and pairs, and 30 disputed decisions were reconciled to 32 final key pairs. Both reviewers then independently coded all 309 studies against the final findings F1 to F8, which span clinical value, performance, transparency, oversight, workflow, readiness, governance, and implementation (Table 1), with prereconciliation κ 0.593-0.702; all 428 disagreements were resolved. The resulting matrix represents coverage of the final reconciled finding wording, not prevalence. A separate blind completeness audit adjudicated 14 residual signals: 9 mapped to F1-F8 and 5 stage 1 studies supported the emergent F9 economic finding (κ=1.000 for the primary decision). Full audit material is in Multimedia Appendix 5.

XSL•FO
RenderX

Table 1. Human-verified integrated findings, nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) anchoring, and key evidence.

Finding	NASSS domains	Convergence	Integrated interpretation	Key evidence[a]
Clinical value and task fit	Condition; value proposition	Mixed or context-dependent	Clinical usefulness and task fit favored acceptance, conditional on performance, workflow, oversight, and governance.	• [32-35] • 4 journal articles
Performance, reliability, validation, and calibrated reliance	Technology; value proposition	Convergent	Credible performance and validation supported confidence; failures undermined reliance, while confidence cues could still produce overreliance.	• [36-40] • 4 journal articles
Transparency, explainability, and calibrated trust	Technology; adopter system	Mixed or context-dependent	Explanations and uncertainty information could help calibration, but quality and representation could also create burden or confusion.	• [38,41-43] • 3 journal articles, 1 conference abstract
Human oversight, autonomy, and relationships	Adopter system; wider system	Convergent	Stakeholders favored decision support with retained human authority and interaction.	• [33,44-46] • 4 journal articles
Usability, workflow, and interoperability	Technology; organization	Convergent	Workflow fit, interoperability, and burden distinguished favorable attitudes from implementable use.	• [47-50] • 4 journal articles
Training, AI literacy, and organizational readiness	Adopter system; organization	Convergent	Training, support, infrastructure, and organizational capacity shaped readiness.	• [51-54] • 3 journal articles, 1 conference abstract
Governance, privacy, fairness, and accountability	Wider system	Mixed or context-dependent	Governance conditions shaped acceptability and deployment, with context-dependent preferred arrangements.	• [55-58] • 4 journal articles
Implementation, adaptation, and sustained adoption	Organization; embedding and adaptation over time	Mixed or context-dependent	Local adaptation, stakeholder involvement, support, and monitoring aided embedding; sustained routine use remained sparse.	• [33,50,59,60] • 3 journal articles, 1 conference abstract
Cost, affordability, and economic trade-offs	Not assigned; emergent completeness finding	Emergent; not convergence-classified	Cost and willingness-to-pay materially shaped stated preferences for AI configurations, but evidence was confined to anticipatory stage 1 studies and does not establish realized cost-effectiveness.	• [55,61-64] • 5 journal articles; study IDs and criterion-level appraisal in Multimedia Appendix 5

[a]For F1-F8, key evidence comprises 32 human-confirmed theme-study pairs (4 per finding) from 28 unique studies; each finding has at least 3 full journal reports. Criterion-level appraisal is in Multimedia Appendix 5; no composite Mixed Methods Appraisal Tool score was used. For F2, a fifth controlled study was added as a boundary refinement after sensitivity testing. F9 was established separately through completeness adjudication and is supported by 5 stage 1 journal studies; its study IDs and appraisal are in Multimedia Appendix 5.

Coverage Sensitivity Analysis

Coverage sensitivity testing sampled 10 studies per F1-F8 finding from studies outside the purposive set, using fixed finding-specific seeds (80 decisions; overlap allowed). SF classified each as supports, qualifies, contradicts, or does not address; any evidence proposed to modify a finding required BP verification. Additional post hoc analyses excluded conference abstracts, restricted the analysis to the 229 studies already retained at the original title/abstract pass, or excluded stage 3 quantitative-descriptive studies. Adjusted F1-F8 coverage models coded supports/qualifies/contradicts as addressed and used Benjamini-Hochberg correction across 16 stage contrasts (14 estimable contrasts in the first-pass restriction). Acceptance and Trust were analyzed separately.

XSL•FO
RenderX

These were robustness analyses, not prevalence estimates or changes to eligibility/coding.

Protocol Deviations and Special Evidence Handling

Protocol deviations were registration after search/deduplication; unperformed reference-list screening, forward citation searching, and final database update; and initial exclusion of a documented 611-record missing/insufficient-abstract subgroup under residual code E8 (other clear ineligibility, note required), contrary to the prespecified uncertainty rule. During title/abstract screening, an additional joint human quality-control and bidirectional readjudication layer not specified in protocol version 1.2 was introduced; original individual decisions were preserved and used for all agreement statistics. One report excluded under code E4 (no relevant stakeholder outcome) was subsequently reopened after an independent pre–full-text title/abstract reassessment and underwent dual independent full-text assessment. The 4-stage grouping and adjusted stage models were defined post hoc from the prespecified exposure and maturity axes. Of the 611-record subgroup, 609 remained excluded after joint readjudication, and 2 were reinstated. Separately, a targeted nonrandom quality control (QC) set of 17 preidentified E8 records was jointly retained for full text (Include: n=13; Maybe: n=4); its enriched yield was not extrapolated to all E8 exclusions. Appraisal sensitivity thresholds were also post hoc. The review therefore represents the August 6-9, 2026, database searches, with known-item validation as the principal search-completeness check.

Two full-text records required documented special handling: one was excluded using official retraction/report-status evidence and one using a validated parallel open-access source-language report. Both remain traceable in the exclusion audit.

Post Hoc Screening Remediation

To remediate the E8 deviation, all 609 records remaining in the missing/insufficient-abstract subgroup were reopened without a new search. Two reviewers independently and blindly reassessed them against locked eligibility criteria. In total, 3 records already had definitive downstream resolutions and were excluded from the net-new denominator, leaving 606; retain/exclude agreement was 98.3% (596/606; κ=0.66). A total of 20 records retained by at least 1 reviewer advanced; 5 required no new eligibility decision because of linkage/prior resolution, leaving 15 retrieval candidates. In total, 10 were opened for dual full-text assessment; 8 yielded unanimous definitive decisions, while 7 reports overall remained unretrieved/uninspectable and unresolved. Four studies were added, increasing the corpus from 305 to 309, and underwent the same dual extraction/appraisal workflow. All reported denominators, tables, and figures were recomputed from N=309. This remediation did not substitute for the unperformed update or citation searches.

Ethical Considerations

This review synthesized published or publicly available research reports and did not involve collection of new participant-level data; ethics approval was therefore not required.

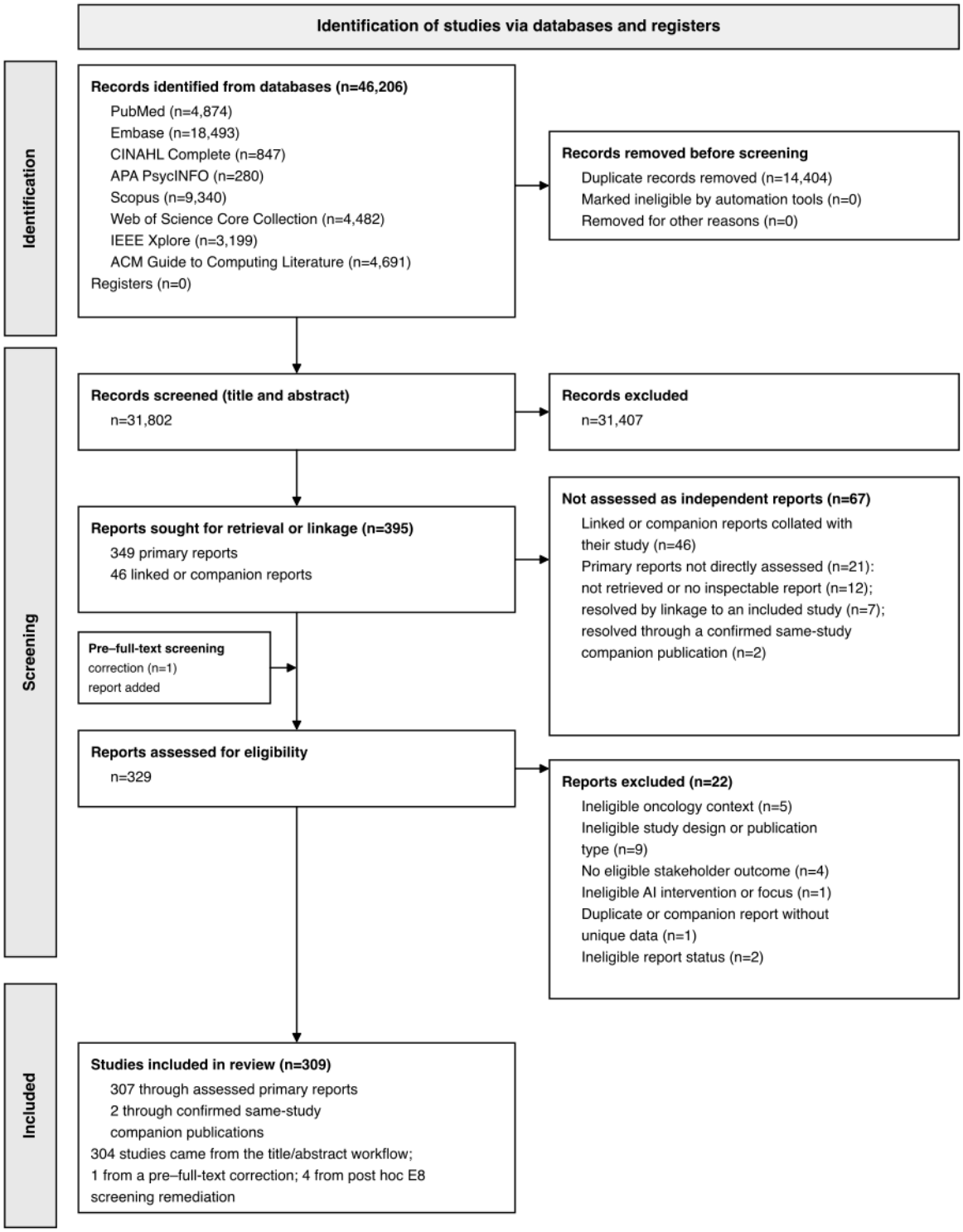
Results

Study Selection

The searches yielded 46,206 records; deduplication left 31,802 for title/abstract screening, of which 31,407 were excluded (Figure 1). The 395 records carried forward comprised 349 primary reports and 46 linked/companion reports. One further report excluded at title/abstract was added to full-text assessment through a documented pre–full-text screening correction. Of the 350 primary reports, 329 were assessed in full text, and 21 were not directly assessed (12 not retrieved or without an inspectable report, 7 resolved by linkage, and 2 through a confirmed companion publication). In total, 22 reports were excluded at full text (Multimedia Appendix 4). The final corpus comprised 309 unique studies (Multimedia Appendix 6): 307 from assessed primary reports and 2 through confirmed companions; 304 came from the title/abstract workflow, 1 from the pre–full-text correction, and 4 from E8 remediation.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram of study selection, incorporating the post hoc E8 screening remediation. Linked or companion reports were collated at study level and were not counted as separate studies. One report excluded at title/abstract was added to full-text assessment through a documented pre–full-text screening correction. One record already assessed at full text through the targeted E8 quality-control set re-entered through the remediation worklist and was counted once. Of the 21 primary reports not directly assessed, 7 were resolved by linkage to an already included study, 2 through a confirmed same-study companion publication, and 12 were not retrieved or had no inspectable report. The remediation rescreened records already retrieved by the August 2026 searches and did not involve a new search.



Study Characteristics

The 309 studies comprised 196 quantitative, 59 mixed methods, 52 qualitative, 1 instrument-validation, and 1 other empirical study (Table 2). There were 227 journal articles, 71 conference abstracts, 8 full conference/proceedings papers, and 3 other reports. Exposure included 86 hypothetical-opinion, 47 information/vignette, 26 prototype/simulation, 68 controlled evaluation, and 82 real-world implementation studies. Maturity remained concentrated in development (n=141), validation

XSL•FO
RenderX

(n=56), pilot (n=53), and implementation (n=53), with only 6 routine-use studies. Most evidence was recent (187/309, 60.5%, in 2025-2026). Of 270 single-country studies, 225 (83.3%) were from high-income countries, and none solely from a low-income country. In total, 74 (23.9%) studies concerned generative/conversational AI (Multimedia Appendix 5).

Table 2. Summary characteristics of the included primary empirical evidence base (N=309).

Domain and category	Values, n (%)[a]
Study design	
Quantitative	196 (63.4)
Mixed methods	59 (19.1)
Qualitative	52 (16.8)
Instrument validation	1 (0.3)
Other empirical	1 (0.3)
Publication type	
Journal article	227 (73.5)
Conference abstract	71 (23)
Full conference or proceedings paper	8 (2.6)
Book chapter	1 (0.3)
Thesis or dissertation	1 (0.3)
Research letter or other	1 (0.3)
AI exposure	
Hypothetical opinion	86 (27.8)
Information or vignette	47 (15.2)
Prototype or simulation	26 (8.4)
Controlled evaluation	68 (22)
Real-world implementation	82 (26.5)
Implementation maturity	
Development	141 (45.6)
Validation	56 (18.1)
Pilot	53 (17.2)
Implementation	53 (17.2)
Routine use	6 (1.9)
Publication year	
2025 or 2026	187 (60.5)
2024	50 (16.2)
2023 or earlier	72 (23.3)
Setting income group	
Single-country, high income	225 (72.8)
Single-country, upper-middle income	29 (9.4)
Single-country, lower-middle income	16 (5.2)
Multicountry or international	38 (12.3)
Not reported	1 (0.3)

[a]Percentages are of the 309 included studies and may not sum to 100 because of rounding.

XSL•FO
RenderX

Stakeholder Representation and Construct Assessment

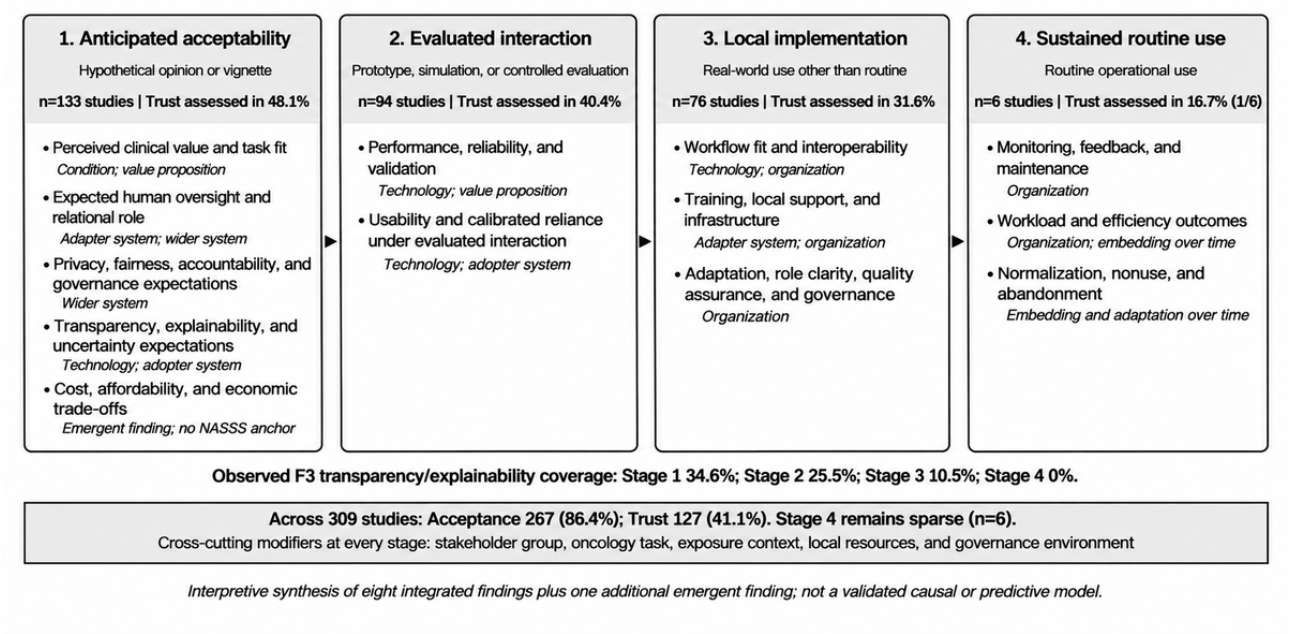
Stakeholder groups were multilabel and represented participants actually studied. Clinicians or licensed health professionals were represented in 190 studies, patients or survivors of cancer in 100, allied or technical health professionals in 34, general-public or lay participants in 31, screening-eligible or at-risk populations in 17, researchers, developers, or experts in 14, caregivers or family members in 11, administrators or decision-makers in 4, and students or trainees and organizations or services or centers in 3 studies each. Acceptance was empirically assessed in 267 of 309 (86.4%) studies: 178 studies measured it quantitatively, 65 elicited it qualitatively, and 24 used both approaches; 29 treated Acceptance only as a proxy, inference, or discussion point, and 13 did not assess it. Trust was empirically assessed in 127 (41.1%) studies: 57 quantitatively, 65 qualitatively, and 5 by both approaches; 150 treated Trust only as a proxy, inference, or discussion point, and 32 did not assess it.

Evidence Across Stages of Exposure and Implementation

Exposure type and implementation maturity were related but not interchangeable. Hypothetical-opinion and vignette studies were almost all at the development stage (129/133), controlled evaluations were mostly validation (47/68) or pilot (20/68) studies, and real-world implementation studies were mostly at the implementation (52/82) or pilot (21/82) stage; all 6 routine-use studies involved real-world implementation (Multimedia Appendix 5). Table 3 groups the studies into the 4 stages used in Figure 2.

Evidence composition differed by stage. General-public participants were mainly anticipatory (23/31), whereas allied/technical professionals were mainly studied in real-world implementation (21/34). Acceptance was empirically assessed in 97.7% (130/133) of stage 1, 73.4% (69/94) of stage 2, 82.9% (63/76) of stage 3, and 83.3% (5/6) of stage 4 studies. Versus stage 1, adjusted Acceptance assessment was less frequent in stage 2 (adjusted odds ratio [aOR] 0.07, 95% CI 0.02-0.25; $P<.001$) and stages 3+4 (aOR 0.12, 95% CI 0.03-0.45; $P=.001$). Trust was assessed in 48.1% (64/133), 40.4% (38/94), 31.6% (24/76), and 16.7% (1/6), respectively, but adjusted contrasts were nonsignificant (stage 2: aOR 1.06, 95% CI 0.56-2.00; $P=.86$; stages 3+4: aOR 0.57, 95% CI 0.29-1.14; $P=.11$). Direction-of-findings distributions in Table 3 are descriptive only.

Figure 2. Stage-contingent interpretation of conditions shaping oncology AI acceptance and implementation. Stages were grouped post hoc from the prespecified exposure and maturity axes: stage 1, hypothetical opinion or vignette (n=133); stage 2, prototype, simulation, or controlled evaluation (n=94); stage 3, real-world implementation other than routine use (n=76); and stage 4, routine use (n=6). Percentages in the stage headers give the share of studies at each stage in which Trust was empirically assessed. The descriptive Trust gradient shown in the stage headers was not significant after adjustment (Table 3). The annotation beneath the stage panels reports study-level coverage of F3, transparency, explainability, and calibrated trust, across the 4 stages; these percentages describe evidence coverage, not importance or effect size. NASSS domains are shown beneath each integrated condition; the emergent economic finding was identified in the subsequent completeness adjudication, was not assigned a NASSS anchor, and was observed only in stage 1. The figure is an interpretive synthesis, not a validated causal or predictive model. NASSS: nonadoption, abandonment, scale-up, spread, and sustainability.



XSL•FO
RenderX

Table 3. Evidence by stage of exposure and implementation (N=309).

Stage	Studies, n	Acceptance empirically assessed[a]		Trust empirically assessed[a]		Positive or enabling, n (%)[b]	Mixed or context-dependent, n (%)[b]	Main participant groups, n[c]
		Values, n (%)	Wilson 95% CI[d]	Values, n (%)	Wilson 95% CI[d]			
1. Anticipated acceptability (hypothetical opinion; information or vignette)	133	130 (97.7)	93.6-99.2	64 (48.1)	39.8-56.5	13 (9.8)	116 (87.2)	Clinicians 70; patients 40; general public 23
2. Evaluated interaction (prototype or simulation; controlled evaluation)	94	69 (73.4)	63.7-81.3	38 (40.4)	31.1-50.5	47 (50)	45 (47.9)	Clinicians 66; patients 32; researchers 7
3. Real-world use/local implementation (other than routine)	76	63 (82.9)	72.9-89.7	24 (31.6)	22.2-42.7	21 (27.6)	51 (67.1)	Clinicians 49; patients 26; allied 19
4. Sustained routine use	6	5 (83.3)	43.6-97.0	1 (16.7)	3.0-56.4	3 (50)	3 (50)	Clinicians 5; patients 2; allied 2
Total	309	267 (86.4)	82.1-89.8	127 (41.1)	35.8-46.7	84 (27.2)	215 (69.6)	Multilabel

[a]Adjusted binomial logistic regression used stage 1 as the reference, pooled stages 3 and 4 only for modeling because stage 4 contained 6 studies, and adjusted for study design, publication format, and publication period. Acceptance and Trust were modeled separately and were not included in the F1-F8 multiplicity family.

[b]Direction of findings was coded at study level for the extracted finding set. Negative or impeding (n=8) and neutral (n=2) classifications are not shown, so the 2 columns do not sum to the stage total.

[c]Participant groups are multilabel: a study contributing more than 1 group is counted in each.

[d]95% CIs for Acceptance and Trust are Wilson score intervals.

Measurement and Critical Appraisal

Measurement was heterogeneous (Table 4). The codebook classified 247 (79.9%) studies as study-specific or unvalidated, a category spanning bespoke questionnaires and study-specific qualitative guides and not implying psychometric inadequacy in every case. Adapted validated instruments appeared in 34 studies, validated instruments in 16, and 12 were not applicable. Validated or adapted measures included the System Usability Scale [65], System Causability Scale [66], and Patient Activation Measure [67]; studies drawing on the technology acceptance model or unified theory of acceptance and use of technology constructs [6,7] without formal validation were classified as study-specific. Among quantitative studies, 153 of 196 (78.1%) were study-specific/unvalidated; among mixed methods studies, 41 of 59 (69.5%). Study-level instruments and validation status are in Multimedia Appendix 5.

"Empirically assessed" indicates that Acceptance or Trust was measured quantitatively, directly elicited through qualitative data, or both. For Trust, direct assessment of trust, confidence, or reliance, including observed reliance behavior, also qualified. Studies addressing these constructs only through proxy indicators, inference, or author discussion were classified separately.

Among the 307 studies appraised with MMAT, both screening questions were satisfied for all studies (614/614 ratings). Across 2125 applicable design-specific MMAT methodological criteria, 1322 (62.2%) were rated Yes, 353 (16.6%) No, and 450 (21.2%) Can't tell. Including the 614 MMAT screening-question ratings, there were 2739 applicable consensus ratings overall (Yes: 1936, No: 353, and Can't tell: 450). For MMAT criterion 5.5, which asks whether the different components of a mixed methods study adhere to the quality criteria of their respective traditions, 0 of 59 mixed methods studies were rated Yes, 43 were rated No, and 16 Can't tell. In nonrandomized studies, confounding was rated No for 49 of 68 (72.1%) applicable assessments. In quantitative descriptive studies, representativeness was rated No in 116 of 177 (65.5%), nonresponse risk was rated No in 74 of 177 (41.8%), and measurement appropriateness was Can't tell in 109 of 177 (61.6%). These patterns informed interpretation but were not converted into a global quality score. Across the 3 post hoc appraisal operationalizations (n=90-187), every F2-F7 contrast significant after false discovery rate correction retained its direction, and the stage 2 contrasts for F2, F5, F6, and F7 remained nominally significant in all 3. Stages 3+4 contrasts were less stable in the smaller subsets: F3 lost significance in 1 of 3 (the no-criterion-rated-No subset, n=117, which contained no stage 4 studies), F5 in 1 of 3, and F7 in 2 of 3. The quality-restricted analyses therefore supported the direction of the stage-contingent pattern but not uniform significance for later-stage contrasts.

XSL•FO
RenderX

Table 4. Construct assessment and measurement status in the included studies (N=309).

Domain and category	Values, n (%)[a]
Acceptance	
Empirically assessed (any mode)	267 (86.4)
Quantitatively measured	178 (57.6)
Qualitatively elicited	65 (21)
Both quantitative and qualitative	24 (7.8)
Proxy, inferred, or discussed only	29 (9.4)
Not assessed	13 (4.2)
Trust	
Empirically assessed (any mode)	127 (41.1)
Quantitatively measured	57 (18.4)
Qualitatively elicited	65 (21)
Both quantitative and qualitative	5 (1.6)
Proxy, inferred, or discussed only	150 (48.5)
Not assessed	32 (10.4)
Instrument status	
Study-specific or unvalidated	247 (79.9)
Adapted validated	34 (11)
Validated	16 (5.2)
Not applicable	12 (3.9)
Reporting	
Response rate available	121 (39.2)

[a]Percentages are of the 309 included studies and may not sum to 100 because of rounding. The study-specific or unvalidated category is an extraction label spanning study-developed questionnaires and qualitative interview or focus-group guides; it is not a claim that all instruments in this category are psychometrically inadequate.

Convergent Integrated Synthesis

Overview

Eight integrated findings were retained after independent human verification and reconciliation, and one additional emergent finding was identified through the separate completeness adjudication. The 8 integrated findings retained 4 convergent and 4 mixed or context-dependent classifications; convergence denotes consistency across designs and stakeholder groups, not study quality. F9 was not convergence-classified. In the final 309-study matrix, human-confirmed evidence coverage was F1 180 (58.3%), F2 183 (59.2%), F3 78 (25.2%), F4 175 (56.6%), F5 182 (58.9%), F6 127 (41.1%), F7 138 (44.7%), F8 150 (48.5%), and F9 5 (1.6%). These are study-level coverage counts, not prevalence estimates of stakeholder attitudes or effect sizes. Table 5 combines the 4-stage descriptive coverage with adjusted categorical stage contrasts. After Benjamini-Hochberg false discovery rate correction across the 16 F1-F8 contrasts, significant patterns included higher F2 coverage in stage 2, lower F3 coverage in stages 3+4, higher F5 coverage in stage 2 and stages 3+4, lower F6 coverage in stage 2, and lower F7 coverage in stage 2 and stages 3+4. The very large F8 contrast for stages 3+4 was expected from the implementation/adaptation content of F8 and was treated as a definition-aligned association rather than an independent stage effect. The findings remain evidence-anchored, conditional statements (Table 1).

Robustness analyses supported the main stage-contingent interpretation. Excluding 71 conference abstracts preserved the adjusted Acceptance contrasts and lower F3 coverage in stages 3+4, while Trust remained nonsignificant. Restricting the corpus to the 229 included studies already retained at the original title/abstract pass preserved every Acceptance and F2-F7 stage contrast that was significant in the full corpus. Acceptance remained less frequent in stage 2 (aOR 0.07, 95% CI 0.01-0.30; $P<.001$) and stages 3+4 (aOR 0.13, 95% CI 0.03-0.62; $P=.01$); F3 remained lower in stages 3+4 (aOR 0.17, 95% CI 0.06-0.48; $q=0.003$). Trust remained nonsignificant, and F8 could not be modeled because all 67 stage 3+4 studies addressed it (quasi-complete separation). Within stage 3, excluding 35 quantitative descriptive studies left 41 studies: F3 was addressed in 2 of 41 (4.9%), F5 in 35 of 41 (85.4%), and Trust in 14 of 41 (34.1%). Full sensitivity results are provided in Multimedia Appendix 5.

XSL•FO
RenderX

Table 5. Study-level coverage of final findings by stage, with adjusted categorical stage contrasts (N=309).

Finding (F)	Coverage by stage, n/N (%)	Stage 2 vs stage 1		Stages 3+4 vs stage 1		Adjusted pattern[a]
		aOR[b] (95% CI)	*q* value	aOR (95% CI)	*q* value	
F1. Clinical value and task fit	• Stage 1: 76/133 (57.1) • Stage 2: 56/94 (59.6) • Stage 3: 43/76 (56.6) • Stage 4: 5/6 (83.3)	1.44 (0.81-2.55)	0.29	1.25 (0.69-2.26)	0.53	No adjusted stage difference
F2. Performance, reliability, validation, and calibrated reliance	• Stage 1: 74/133 (55.6) • Stage 2: 68/94 (72.3) • Stage 3: 38/76 (50) • Stage 4: 3/6 (50)	2.44 (1.32-4.51)	0.009	0.95 (0.52-1.71)	0.91	Higher in stage 2
F3. Transparency, explainability, and calibrated trust	• Stage 1: 46/133 (34.6) • Stage 2: 24/94 (25.5) • Stage 3: 8/76 (10.5) • Stage 4: 0/6 (0)	0.60 (0.31-1.14)	0.19	0.20 (0.08-0.47)	<0.001	Lower in stages 3+4
F4. Human oversight, autonomy, and relationships	• Stage 1: 87/133 (65.4) • Stage 2: 45/94 (47.9) • Stage 3: 38/76 (50) • Stage 4: 5/6 (83.3)	0.58 (0.33-1.04)	0.12	0.68 (0.38-1.23)	0.29	No adjusted stage difference
F5. Usability, workflow, and interoperability	• Stage 1: 49/133 (36.8) • Stage 2: 75/94 (79.8) • Stage 3: 53/76 (69.7) • Stage 4: 5/6 (83.3)	9.13 (4.63-17.99)	<0.001	5.52 (2.86-10.64)	<0.001	Higher in stage 2 and stages 3+4
F6. Training, AI literacy, and organizational readiness	• Stage 1: 71/133 (53.4) • Stage 2: 15/94 (16) • Stage 3: 38/76 (50) • Stage 4: 3/6 (50)	0.14 (0.07-0.28)	<0.001	0.80 (0.44-1.44)	0.53	Lower in stage 2
F7. Governance, privacy, fairness, and accountability	• Stage 1: 93/133 (69.9) • Stage 2: 13/94 (13.8) • Stage 3: 32/76 (42.1) • Stage 4: 0/6 (0)	0.07 (0.03-0.14)	<0.001	0.31 (0.16-0.60)	0.001	Lower in stage 2 and stages 3+4
F8. Implementation, adaptation, and sustained adoption	• Stage 1: 42/133 (31.6) • Stage 2: 27/94 (28.7) • Stage 3: 75/76 (98.7) • Stage 4: 6/6 (100)	0.98 (0.49-1.94)	0.94	333.53 (42.41-2622.67)	<0.001	Near-structural, definition-aligned stage 3+4 association
F9. Cost, affordability, and economic trade-offs	• Stage 1: 5/133 (3.8) • Stage 2: 0/94 (0) • Stage 3: 0/76 (0) • Stage 4: 0/6 (0)	Not modeled	Not modeled	Not modeled	Not modeled	Emergent stage 1 finding

[a]Coverage is shown for stage 1 (n=133), stage 2 (n=94), stage 3 (n=76), and stage 4 (n=6); addressed comprises supports, qualifies, or contradicts in the final consensus matrix. aORs compare stage 2 and stages 3+4 with stage 1 and control for study design, publication format, and publication period. *q* values are Benjamini-Hochberg false discovery rate adjusted across the 16 F1-F8 contrasts. Stage 4 remained separate descriptively but was pooled with stage 3 for modeling because n=6. F9 was not modeled because all 5 instances occurred in stage 1. Coverage indicates what the literature addressed and does not indicate stakeholder prevalence, effect size, or the importance of a finding. The F8 stages 3+4 estimate is near-structural because F8 concerns implementation, adaptation, and sustained adoption and should not be interpreted as an independent stage effect.

[b]aOR: adjusted odds ratio.

Clinical Value and Task Fit

Perceived clinical usefulness and fit with the specific oncology task were linked to greater acceptance or willingness to use AI, but favorable value judgments remained conditional on demonstrated performance, workflow compatibility, human oversight, and governance; some tools and use cases showed only modest or context-specific benefit [32-35]. The finding was therefore classified as mixed/context-dependent, not as a general efficacy claim.

Performance, Reliability, Validation, and Calibrated Reliance

Stronger performance, reliability, and credible validation were linked to greater confidence and willingness to use AI, whereas errors, false positives, hallucinations, uncertain generalizability,

XSL•FO
RenderX

or weak validation undermined trust and reliance [36-39]. A controlled study additionally showed that high displayed AI confidence could increase trust while increasing overreliance and reducing diagnostic accuracy [40]. After independent second-reviewer verification, this was incorporated as a boundary refinement: greater trust was not necessarily calibrated reliance. Technical performance alone did not resolve governance, responsibility, or workflow concerns.

Transparency, Explainability, and Calibrated Trust

Transparency, understandable explanations, and uncertainty information could support calibrated trust, but their effects depended on explanation quality, representation, user experience, and clinical context. In some studies, explanation or uncertainty displays introduced confusion or interpretive burden, and explainability alone did not reliably increase trust [38,41-43]. This finding was mixed/context-dependent.

Human Oversight, Autonomy, and Relationships

Patient, public, and professional evidence favored AI as decision support with retained human oversight and professional authority, not autonomous replacement; preservation of human interaction remained an important condition of acceptable use [33,44-46]. This was one of the clearest convergent findings.

Usability, Workflow, and Interoperability

Workflow fit and usability distinguished favorable attitudes from implementable use: interoperability, time savings, and reduced burden facilitated uptake, whereas manual entry, poor interfaces, added work, and workflow disruption impeded it [47-50]. The convergence reflected both positive implementations and cases in which deployment failed to produce the anticipated efficiency gain.

Training, AI Literacy, and Organizational Readiness

Training, AI literacy, local support, infrastructure, and organizational capacity were linked to readiness and intention to use AI and were identified as modifiable implementation conditions; limited knowledge, access, resources, or implementation capacity constrained readiness [51-54]. The evidence was strongest for professional readiness and preadoption contexts.

Governance, Privacy, Fairness, and Accountability

Privacy, data security, fairness, accountability, liability, consent, and regulatory clarity shaped trust, acceptability, and deployment choices. These governance requirements functioned chiefly as enabling or limiting conditions, but their salience and preferred arrangements varied by stakeholder and context [55-58]. The finding was therefore mixed/context-dependent; it did not represent a single universal governance preference.

Implementation, Adaptation, and Sustained Adoption

Implementation and preimplementation evidence indicated that readiness for and embedding of AI were supported by local workflow adaptation or configurability, stakeholder involvement, training and information technology support, quality assurance or monitoring, and iterative feedback and refinement [33,50,59,60]. Sustained normalization remained sparsely evidenced: only 6 of the 309 included studies were coded as routine use, and real-world deployment did not consistently produce expected efficiency gains. This finding was mixed/context-dependent.

Cost, Affordability, and Economic Trade-offs

Five stage 1 studies supported a distinct economic finding. Cost was an explicit preference dimension, not simply an incidental study characteristic: anticipated higher health care costs were an explicit patient concern; willingness-to-pay and provider-cost trade-offs affected preferences for AI tumor diagnostics and skin-cancer screening; and radiation-therapy preferences involved heterogeneous trade-offs among cost, speed, autonomy, privacy, and clinician oversight [55,61-64]. All 5 studies were anticipatory; none evaluated interaction or implementation. This finding therefore concerns stated economic preferences and affordability trade-offs, not realized cost-effectiveness or implementation economics. The supporting studies were S00631, S01114, S01974, S05152, and S19803 (Multimedia Appendix 5).

Coverage Sensitivity Analysis

The sensitivity analysis comprised 80 theme-study decisions from 72 unique studies outside the 47-study purposive synthesis set and was applied to the F1-F8 findings only. In total, 40 decisions supported the tested finding, 4 qualified it, none contradicted it, and 36 did not address it. The number of sampled studies that substantively addressed each finding was clinical value and task fit 6/10, performance and reliability 6/10, transparency and explainability 3/10, human oversight 7/10, workflow and usability 6/10, training and readiness 4/10, governance 5/10, and implementation and sustained adoption 7/10. The 4 qualifying decisions comprised 1 clinical-value case and 3 performance/reliability cases. The calibrated-reliance boundary condition incorporated into the performance finding was independently checked by the second reviewer. No F1-F8 finding was overturned, and classification counts were treated as robustness evidence only, not as theme prevalence. F9 emerged from a separate blind completeness adjudication and was not retrospectively inserted into this random sensitivity sample. Tables 6 and 7 provide, respectively, the interpretive stage-specific model for clinical evaluation and the comparison with recent evidence syntheses discussed below.

XSL•FO
RenderX

Table 6. Interpretive stage-specific model for evaluating AI in oncology practice.

Stage	Clinical question a service might ask	What might be assessed	Oncology emphasis	Evidence caution
1. Anticipated use (n=133)	Might this tool help this cancer task, and under what conditions would it be acceptable?	Clinical value and task fit; support versus replacement; explanation and uncertainty needs; privacy, liability, fairness, consent; stated cost or willingness-to-pay	Preference-sensitive choices; tumor-board authority; communication before an irreversible recommendation; screening false-positive burden.	Acceptance in 130/133 studies, so later contrasts partly reflect a ceiling. Cost evidence is only here (5 studies) and is stated preference, not cost-effectiveness.
2. Evaluated interaction (n=94)	When people use a functioning system, does reliance stay calibrated and the task workable?	Performance for the oncology end point; usability and added burden; overrides and overreliance, including displayed confidence; retain transparency and governance	Contouring uncertainty; diagnostic sign-out; one controlled breast-cancer study in which higher displayed confidence increased trust and overreliance and reduced accuracy.	Acceptance less frequent than stage 1 (aOR[a] 0.07, 95% CI 0.02-0.25). Training and governance coverage also lower. These constructs should not be dropped only because interaction studies emphasize workflow.
3. Local implementation (n=76)	Can the tool be embedded without leaving accountability or workflow unresolved?	Interoperability with imaging, pathology, radiotherapy, or the oncology record; training and local support; adaptation; continued Acceptance and Trust; operational governance	Lung-cancer imaging triage, prostate computer-aided detection, ambient documentation, diagnostic tools in lower-resource clinics. Allied and technical staff are more visible here.	Workflow coverage higher (aOR 5.52, 95% CI 2.86-10.64 vs stage 1). Transparency coverage 8/76. Governance lower (aOR 0.31, 95% CI 0.16-0.60). Early deployment is not routine use.
4. Attempted routine use (n=6)	Does use persist after support recedes, and do safeguards still operate?	Sustained use, nonuse, abandonment; workload and overrides; governance performance; economic consequences; recalibration if case mix or the model changes	A successful pilot in screening, radiotherapy, or clinic should not be treated as evidence that the tool has become normal work.	Only 6 studies. Trust assessed in 1/6. Stage 4 was pooled with stage 3 for modeling and should not support its own effect estimate.

[a]aOR: adjusted odds ratio.

XSL•FO
RenderX

Table 7. Positioning of this review relative to recent evidence syntheses.

Review	Design	Included studies, n[a]	Search end[a]	Scope
Young et al (2021) [14]	Mixed methods systematic review	23	September 2020	Health care, patients and public
Hogg et al (2023) [17]	Qualitative systematic review	111	April 2021	Clinical AI, multiple stakeholders
Lambert et al (2023) [15]	Integrative review	42	June 2022	Hospitals, health professionals
Kuo et al (2024) [18]	Qualitative evidence synthesis	44	February 2024	Diagnostic AI, multiple stakeholders
Goh et al (2025) [21]	Systematic review and framework	20	January 2024	Breast cancer screening programs
Tun et al (2025) [19]	Systematic review	27	November 2024	AI decision support, health care workers
Wenderott et al (2025) [24]	Systematic review and qualitative analysis	38	November 2024	Routine AI in medical imaging
Hilbers et al (2025) [20]	Scoping review	19	February 2025	Cancer care, patients
Kikuchi et al (2026) [22]	Scoping review	199	September 2024	Palliative and supportive cancer care
White et al (2026) [23]	Scoping review	44	January 2025	Empathic AI in cancer care
Shi et al (2026) [16]	Systematic review	61	December 2025	Health care, patients
Menz et al (2026) [68]	Scoping review	32	July 2026[b]	Generative AI, people affected by cancer
This review	Mixed methods systematic review (JBI[c] convergent integrated)	309 primary studies	August 2026	All oncology tasks, 10 stakeholder groups, exposure by implementation maturity

[a]Numbers of included studies and search end dates are as reported in the respective reviews.

[b]This review was published after the search for the present review had been completed and could not inform its conduct.

[c]JBI: Joanna Briggs Institute.

Discussion

Principal Findings and Contribution

This review suggests that oncology AI acceptance, trust, and implementation are better understood as a stage-contingent clinical problem than as a single attitude or a direct consequence of model performance. The oncology-specific contribution is not that cancer stakeholders hold attitudes unseen in other specialties. It is that the same sociotechnical conditions may have different consequences when decisions are preference-sensitive, multidisciplinary, and difficult to reverse. A screening algorithm that increases recalls, a contouring tool that shifts target volumes, a pathology aid that anchors sign-out, or a treatment-response display that influences a neoadjuvant recommendation can each alter care before a service has established who is accountable, how uncertainty is communicated, or whether clinicians override erroneous output. In that setting, it may matter that the literature does not appear to study the same questions before and after deployment.

The main empirical signal is a shift in measurement. Acceptance was almost universally assessed at the anticipatory stage but substantially less often later, even after adjustment. Workflow and interoperability became more prominent once participants interacted with a system or a service deployed one. Transparency, explainability, and governance were examined less often at later stages. Trust showed no independent stage association after adjustment. These are differences in what studies examined, not estimates of how important each domain is to patients or clinicians. Only 6 studies described routine use despite 82 real-world implementations, so sustained normalization remains thinly evidenced.

The 8 integrated findings, together with 1 additional emergent economic finding, organize the conditions that may shape movement from anticipated value to sustainable use without implying a causal sequence. Clinical usefulness and reliable performance operate alongside explanation quality, retained human authority, workflow compatibility, training and organizational capacity, governance safeguards, and local adaptation. The very large stage 3+4 association for implementation and sustained adoption (F8) is partly definition-aligned and is therefore not interpreted as an independent stage effect. The emergent economic finding is narrower: in 5 anticipatory studies, cost, affordability, or willingness-to-pay materially shaped stated preferences, but no distinct later-stage implementation-economics finding was identified. These distinctions are especially relevant in oncology, where high-stakes decisions, longitudinal care relationships, preference-sensitive choices, multidisciplinary workflows, and accountability make poorly calibrated reliance and weak implementation safeguards consequential.

Acceptance Is Not the Same as Calibrated Trust

A central distinction is between acceptance, subjective trust, and calibrated reliance. Acceptance was empirically assessed in 267 studies, and Trust in 127, but the stage pattern differed

XSL•FO
RenderX

between the 2 constructs. Compared with stage 1, Acceptance was markedly less often assessed in stage 2 (aOR 0.07, 95% CI 0.02-0.25) and stages 3+4 (aOR 0.12, 95% CI 0.03-0.45), whereas the adjusted Trust contrasts were nonsignificant (Table 3). Because Acceptance was assessed in almost all anticipatory studies (130/133, 97.7%), these odds ratios partly reflect a stage 1 ceiling; absolute differences were 24 and 15 percentage points. The descriptive decrease in Trust assessment across stages should therefore not be interpreted as an independent stage effect. This asymmetry matters because willingness to use can be studied before deployment, while calibrated reliance requires observation of behavior in relation to system reliability. A controlled breast-cancer decision-support study identified during sensitivity testing illustrated that distinction: higher displayed AI confidence increased trust but also increased overreliance and reduced diagnostic accuracy [40]. Future evaluations should therefore measure whether reliance changes appropriately with reliability, whether clinicians override erroneous recommendations, and whether confidence or explanation displays alter diagnostic behavior. Self-reported willingness or trust alone is not sufficient evidence of safe adoption.

From Anticipated Acceptability to Sustained Use

Prior reviews have highlighted many of the same broad domains, including transparency, reliability, usability, human control, workflow, and governance [17,19,20,24]. The added value of the present synthesis is to show that their empirical coverage changes substantially with exposure and maturity (Table 5). After adjustment, transparency/explainability and calibrated trust (F3) were less often addressed in stages 3+4 than in stage 1 (aOR 0.20, 95% CI 0.08-0.47; q<0.001). By contrast, usability, workflow, and interoperability (F5) were much more often addressed in stage 2 (aOR 9.13, 95% CI 4.63-17.99; q<0.001) and stages 3+4 (aOR 5.52, 95% CI 2.86-10.64; q<0.001). Governance, privacy, fairness, and accountability (F7) showed the opposite pattern, with lower coverage in stage 2 (aOR 0.07, 95% CI 0.03-0.14; q<0.001) and stages 3+4 (aOR 0.31, 95% CI 0.16-0.60; q=0.001). These are differences in what studies examined, not estimates of stakeholder importance or causal effects. Later-stage studies measured practical conditions of use much more often while retaining transparency and governance less consistently as empirical targets.

Consistent with the protocol, the 8 integrated findings were anchored to NASSS domains (Table 1); the emergent economic finding was reported separately because the human-confirmed completeness adjudication assigned it no NASSS anchor. Figure 2 adds the prespecified exposure and implementation-maturity axes. The central Acceptance, F3, F5, and F7 patterns persisted when the analysis was restricted to the 229 studies already retained at the original title/abstract pass. Under post hoc appraisal restrictions, these contrasts kept their direction, although several stages 3+4 contrasts lost significance in smaller subsets. The interpretation is therefore robust to screening provenance and directionally robust to appraisal restriction, but it should not be read as a fixed developmental pathway.

An Interpretive Stage-Specific Model for Clinical Evaluation

The stage contrasts may be more useful to oncology services if they are restated as questions to ask, rather than only as differences in literature coverage. Table 6 proposes such a model. It is anchored to the prespecified exposure and maturity axes, the adjusted coverage contrasts in Tables 3 and 5, and the integrated findings. It is not a validated pathway, and a tool need not pass through every stage. Stages 3 and 4 were pooled for modeling because only 6 studies met the routine-use definition; stage 4 is therefore shown separately for clinical planning but should not be read as a stable estimate.

Three clinical readings follow, each suggestive. First, anticipatory acceptability may be a poor proxy for safe use. Although Acceptance was empirically assessed in 130 of 133 (97.7%) stage 1 studies, these studies examined hypothetical opinions or vignettes rather than actual clinical use. Such findings cannot establish whether clinicians would recognize and appropriately challenge inaccurate AI outputs, such as an incorrect radiotherapy contour, screening recall recommendation, or treatment-response assessment. Second, contact with a working system appears to move attention toward workflow, which is appropriate, but transparency and governance coverage were lower later. A radiotherapy or diagnostic service might therefore deliberately retain explanation, uncertainty display, consent, liability, and fairness checks after go-live, because the later literature examined them less consistently. Third, local implementation should not be treated as sustained use. Real-world deployment was compatible with pilot or bounded implementation in most studies, and evidence on abandonment and workload after support is withdrawn, and realized cost is sparse. The 5 economic studies were all anticipatory, so willingness-to-pay should not be read as cost-effectiveness.

Across stages, 6 conditions recurred and may be worth preserving rather than assuming they are settled once a tool is deployed: a plausible oncology use case, performance credible for that task, retained human authority in the tumor board or responsible clinician, workflow fit without hidden clerical burden, training and local capacity, and governance that names who is accountable if the recommendation is wrong. Their coverage was not stable. Workflow was addressed more often after interaction. Transparency and governance were addressed less often later. A cautious clinical implication is to measure the constructs that become harder to see, not only the frictions that become obvious. For treatment-response prediction specifically, Table 6 should be read as a prospective checklist rather than as evidence that any neoadjuvant system will be accepted.

Human Oversight and the Relational Context of Oncology

The integrated evidence supported AI as decision support with retained human authority, not autonomous replacement [33,44-46]. In oncology, uncertainty, preference-sensitive decisions, longitudinal relationships, and high-stakes communication are integral to care. Retained human oversight is more than a safety backstop; it is part of the clinical and

XSL•FO
RenderX

relational context in which AI recommendations acquire meaning, responsibility, and legitimacy. This also helps explain why technically strong systems may remain unacceptable when role boundaries, accountability, or patient-clinician interaction are poorly specified.

Workflow, Readiness, and Governance as Implementation Conditions

Usability, interoperability, time burden, manual data entry, interface design, and fit with existing workflows conditioned whether favorable perceptions could translate into feasible use [47-50]. Training and AI literacy were embedded within organizational readiness rather than treated as individual attributes in isolation [51-54]. Governance concerns were similarly operational: privacy, fairness, consent, accountability, liability, and regulatory clarity shaped deployment choices [55-58]. The adjusted stage analysis adds an important qualification: workflow coverage increased sharply with direct interaction and implementation, whereas governance coverage was lower than at the anticipatory stage. The literature therefore appears to follow operational friction into deployment more consistently than governance questions. These domains should be evaluated together because a system can perform well and fit a workflow while accountability, fairness, consent, or oversight remain unresolved.

Transferability to Treatment-Response Prediction

The review was motivated in part by translational questions surrounding AI-based treatment-response prediction, but the evidence base should not be read as direct evidence that a particular neoadjuvant chemotherapy response-prediction system will be accepted. The corpus spans heterogeneous oncology tasks, stakeholder groups, and exposure conditions and was not designed to estimate use-case–specific acceptance. The 8 integrated findings and the additional emergent economic finding are therefore better interpreted as conditions to test prospectively in that workflow: demonstrable clinical value and credible validation; uncertainty communication and calibrated reliance; retained clinician authority; workflow integration; training and organizational readiness; explicit accountability; and, at the anticipatory stage, cost and affordability trade-offs. Direct evaluation with the clinicians and patients who would use or be affected by a treatment-response system remains necessary.

Comparison With Prior Work

Table 7 positions this review against 11 recent syntheses and 1 concurrent review. Earlier syntheses addressed patient/public attitudes [14,16], health professionals' acceptance [15], diagnostic AI [18], implementation factors [17], health care workers' trust [19], oncology patient perspectives [20], and stage-specific implementation in medical imaging [24]. Oncology-specific frameworks by Hantel et al [4] and Stetson et al [5] likewise emphasize ethical deployment, accountability, and lifecycle governance. Other oncology syntheses have focused on breast screening [21], palliative/supportive care [22], empathic AI [23], and, in a scoping review published after our search, generative AI perspectives among people affected by cancer [68]. The present review is oncology-wide, includes 10 stakeholder groups and 309 primary studies, 187 of them published in 2025 or 2026, and cross-classifies evidence by exposure type and implementation maturity.

Our synthesis complements these contributions by integrating quantitative, qualitative, and mixed methods primary evidence across multiple oncology stakeholders; explicitly separating exposure type from implementation maturity; distinguishing Acceptance from empirically assessed Trust and calibrated reliance; and combining construct measurement, critical appraisal, independent human synthesis verification, and random coverage sensitivity testing. The resulting contribution is a cross-stakeholder, stage-contingent account of oncology AI adoption, not a generic catalogue of attitudes or barriers.

Measurement and Methodological Implications

Measurement remained heterogeneous (Table 4): 247 of 309 studies used study-specific or unvalidated instruments. Later-stage evidence also measured Acceptance less often, while Trust remained sparse despite settings that permit direct observation of reliance, overrides, workload, and nonuse. Conference abstracts (n=71 studies) often provided less methodological detail, and no mixed methods study met MMAT criterion 5.5. These features limit comparability and reinforce the need to distinguish recency from external validation or implementation maturity.

Future studies would benefit from clearer construct definitions, validated or transparently adapted measures, and prespecified assessment of trust calibration. Longitudinal or repeated-stage designs could test whether anticipated acceptance persists after direct interaction, whether explanation needs change with experience, and whether governance concerns are resolved or simply stop being measured after deployment. Reports should also distinguish participants actually sampled from intended users and other affected stakeholders.

Strengths and Limitations

Strengths include an 8-database search, reproducible deduplication, calibrated balanced-allocation title/abstract screening with joint bidirectional readjudication and blinded verification, dual independent full-text assessment, 2-reviewer extraction and appraisal, study-level report linkage, and independently verified convergent integrated synthesis. Random coverage sensitivity extended direct human review to 72 additional unique studies outside the 47-study purposive synthesis set; secondary reviews and frameworks were excluded from empirical denominators.

Several limitations require explicit consideration. First, primary title/abstract decisions were made by 1 reviewer within 2 balanced allocations rather than by 2 independent reviewers for all 31,802 records. However, every primary record forwarded to full-text assessment received a second human review through joint readjudication, targeted QC, an independent pre–full-text reassessment, or dual remediation rescreening; single-reviewer decisions therefore stood only for exclusions, which blinded verification sampled at 10%. The original screening master contained 1124 E8 exclusions. Of the 80 included studies first excluded at title/abstract, 34 carried code E8 and were recovered through joint readjudication (n=17, including the 2 reinstated

XSL•FO
RenderX

from the 611-record missing/insufficient-abstract subgroup), targeted QC (n=13), or remediation (n=4). E8 exclusions outside that subgroup were not comprehensively rescreened beyond these documented quality-control layers. In total, 4 records from the targeted QC set also appeared in the 609-record remediation worklists; all 4 were rescreened blind, 3 were recognized at reconciliation as previously resolved and excluded from the net-new denominator, and the fourth was recognized only at identity reconciliation and remained in the 606-record denominator. The remediation added 4 studies from 606 net-new records and left 7 reports unresolved; these cases do not bound all possible residual screening error. The initial E8 use remains a documented departure from the prespecified uncertainty rule. Second, reference-list screening, forward citation searching, and the planned final update search were not completed before synthesis despite being specified in the protocol. Dedicated gray-literature databases or repositories were also not searched, although eligible records indexed within the prespecified sources were retained. Third, 2 full-text records required nonstandard but documented evidence handling, one through official retraction/report-status evidence and one through a validated parallel open-access source-language report.

Fourth, stakeholder representation was still based on primary coding with a blinded 50-study verification sample for the primary workflow, although the 4 remediation studies were dual-coded. Acceptance/Trust modes, however, were subsequently independently coded across all 309 studies, and every initial disagreement was readjudicated blind by both reviewers before residual joint consensus. The κ values of 0.781 for Acceptance and 0.775 for Trust therefore quantify agreement within the deliberately difficult disagreement subsets (119 and 81 cases), not whole-corpus reliability; initial full-corpus κ was 0.406 and 0.589, respectively. Fifth, F1-F8 were ultimately independently coded across all 309 studies and reconciled, strengthening coverage beyond the 47-study purposive evidence set. The 80-decision random sensitivity analysis nevertheless remained restricted to the 8 integrated findings. That design provides a robustness check, not an exhaustive falsification test: the absence of contradicts classifications does not establish that contradictory evidence is absent from the corpus, particularly because the full-corpus coverage matrix was coded against finding formulations that had incorporated qualifying evidence during synthesis development. F9 emerged from a completeness audit of 14 residual signals: both reviewers agreed independently on all 14 primary decisions, but the distinct finding is supported by only 5 stage 1 studies and was not subjected to that random sensitivity procedure. It should therefore be regarded as an emergent, early-stage economic finding and does not establish realized implementation cost-effectiveness.

Sixth, publication and indexing bias cannot be excluded. Studies reporting failed implementation, neutral stakeholder responses, or abandonment may be less likely to enter the published literature. Although no language filter was applied in the database searches, language and indexing asymmetries may still affect which reports could be retrieved and assessed reliably. The evidence base was concentrated in high-income countries (225/270 single-country studies, 83.3%), and no study was conducted solely in a low-income country, which limits transferability to lower-resource oncology services. Seventh, the review intentionally spans oncology and was not designed to estimate acceptance for a single AI use case such as neoadjuvant chemotherapy response prediction. The findings should therefore be treated as cross-oncology implementation conditions to test prospectively, not as direct evidence that any particular response-prediction system will be accepted. Eighth, no formal certainty assessment such as Grading of Recommendations Assessment, Development and Evaluation–Confidence in the Evidence from Reviews of Qualitative Research or JBI ConQual was performed; confidence in each finding is conveyed through its convergence status, the methodological profile of its key evidence, and the coverage sensitivity analysis. Ninth, 3 of the 32 key evidence pairs derived from conference abstracts and 1 representative study reported a patient and public involvement event, although every finding was supported by at least 3 full journal reports. Adjusted stage models were post hoc and exploratory; stage 4 had only 6 studies, and sparse or near-structural cells, especially F8, limit precision. They are interpreted as coverage associations, not causal effects. Finally, heterogeneity in design, stakeholder population, AI task, exposure, implementation maturity, publication format, and measurement instruments limits cross-study comparability and precludes interpreting thematic support counts as prevalence.

Implications for Research and Practice

Future oncology-AI evaluations should move beyond asking whether stakeholders are willing to use AI and should preserve key constructs across implementation stages. Prospective studies should test whether anticipated acceptance persists after interaction and deployment, whether trust is calibrated to reliability, how clinicians override or defer to AI, how workload changes, and whether adoption persists after implementation support diminishes. Reports should document local adaptation, training, IT support, monitoring, quality assurance, feedback loops, and reasons for nonuse or abandonment. Governance evaluation should continue after deployment and be linked to concrete decisions and accountability. For high-stakes decision support, acceptance and reliance should be measured directly in the target clinical workflow.

For practice, Figure 2 provides a pragmatic evaluation sequence. Before deployment, organizations can test clinical value, role boundaries, affordability, and explanation or uncertainty information; during evaluation, reliability, usability, and calibrated reliance; during implementation, workflow, training, infrastructure, governance, resources, Acceptance, and Trust; and after deployment, sustained use, workload, overrides, nonuse, abandonment, governance performance, and economic consequences. This is an interpretive framework, not a validated implementation score.

Conclusions

This review identifies a stage-contingent evidence gap that may be clinically important in oncology. The questions studied before deployment are not the same questions studied once systems enter evaluation and practice. Empirical assessment of Acceptance was concentrated in anticipatory research, workflow coverage was more prominent once participants interacted with

XSL•FO
RenderX

a system, and transparency, explainability, and governance were examined less often later. Trust showed no independent adjusted stage association, and self-reported trust remains conceptually distinct from calibrated reliance. Across stages, clinical value, credible performance, human oversight, workflow fit, organizational readiness, and governance remain recurrent conditions, while evidence on sustained routine use and realized economic consequences is sparse. Table 6 restates these patterns as an interpretive model for what an oncology service might assess at each stage. The translational challenge is not simply favorable attitudes or early deployment. It is whether acceptance, reliance, governance, and performance remain aligned as AI is embedded in screening, diagnostics, radiotherapy, treatment recommendation, and longitudinal cancer care.

Acknowledgments

The authors thank the AI-PredictChemo project team for supporting the development of this review.

Funding

This review was conducted within the AI-PredictChemo project at Danube Private University. Funding for AI-PredictChemo was provided by the EU Structural Fund IBW/EFRE and the Lower Austrian Economic and Tourism Fund (project: WST3-F-5035462/005-2024). No separate grant was awarded specifically for this review. The funders had no role in the study design, data collection, analysis or interpretation, preparation of the manuscript, or the decision to submit the manuscript for publication.

Data Availability

The data supporting the findings of this review are included in the article and its multimedia appendices. Complete database search strategies and execution records are provided in Multimedia Appendix 3; the full-text exclusion audit in Multimedia Appendix 4; the study-level extraction dataset, codebook, critical-appraisal results, and synthesis and sensitivity audit materials in Multimedia Appendix 5; and the list of included studies in Multimedia Appendix 6. The registered protocol and supporting methodological materials are available through the Open Science Framework registration [28]. The completed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 and PRISMA-S (PRISMA Extension for Literature Searches) checklists are provided in Multimedia Appendices 1 and 2, respectively. The analytic code used for the reported analyses is available from the corresponding author on reasonable request.

Generative AI Use

OpenAI Codex was used as a programming assistant to develop and refine Python scripts executed locally in PyCharm. In the revised review, scripts supported deterministic deduplication and data handling, reproducible sampling and allocation, blinded reviewer materials, quality-control and integrity checks, and, after reviewer coding, analytic tables, statistics, and figures. No revised-review extraction field was populated programmatically. In the original narrative review, Codex-assisted scripts prepopulated selected fields; reviewers verified and corrected them against source PDFs, and that dataset was discarded. Codex did not make or suggest eligibility, extraction, appraisal, coding, adjudication, synthesis, sensitivity, or scientific interpretation decisions. No generative AI tool was used to draft or revise manuscript prose. All scientific judgments were made by the human reviewers, and the authors take full responsibility for the manuscript.

Authors' Contributions

Conceptualization: SF, AM, SH
Data curation: SF, BP
Formal analysis: SF
Investigation: SF, BP
Methodology: SF, KEAC
Project administration: SF
Supervision: KEAC, SH
Validation: BP, AAA
Writing – original draft: SF
Writing – review & editing: SF, BP, AM, AAA, KEAC, SH

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA 2020 checklist for systematic reviews.

XSL•FO
RenderX

[DOCX File , 13 KB-Multimedia Appendix 1]

Multimedia Appendix 2

PRISMA-S checklist for reporting literature searches in systematic reviews.
[DOCX File , 10 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Complete database search strategies and execution record, including Peer Review of Electronic Search Strategies documentation.
[DOCX File , 59 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Complete full-text exclusion table (n=22) with reasons, and the 7 unresolved reports identified during screening remediation.
[DOCX File , 14 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Study-level extraction dataset, construct and instrument coding, codebook, critical-appraisal outputs, synthesis and sensitivity audit materials, and the study-level classifications used for the descriptive analyses.
[XLSX File (Microsoft Excel File), 1738 KB-Multimedia Appendix 5]

Multimedia Appendix 6

Reference list of the 309 included studies.
[DOCX File , 43 KB-Multimedia Appendix 6]

References

1. Kann BH, Hosny A, Aerts H. Artificial intelligence for clinical oncology. Cancer Cell. 2021;39(7):916-927. [doi: 10.1016/j.ccell.2021.04.002]
2. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195. [FREE Full text] [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]
3. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res. 2017;19(11):e367. [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
4. Hantel A, Clancy DD, Kehl KL, Marron JM, Van Allen EM, Abel GA. A process framework for ethically deploying artificial intelligence in oncology. J Clin Oncol. 2022;40(34):3907-3911. [doi: 10.1200/jco.22.01113]
5. Stetson PD, Choy J, Summerville N, Baldwin-Medsker A, Mak J, Chatterjee A, et al. Responsible artificial intelligence governance in oncology. NPJ Digit Med. 2025;8(1):407. [FREE Full text] [doi: 10.1038/s41746-025-01794-w] [Medline: 40615544]
6. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. 1989;13(3):319. [doi: 10.2307/249008]
7. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q. 2003;27(3):425. [doi: 10.2307/30036540]
8. Sheeran P. Intention-behavior relations: a conceptual and empirical review. Eur Rev Soc Psychol. 2011;12(1):1-36. [doi: 10.1080/14792772143000003]
9. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health. 2011;38(2):65-76. [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
10. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors. 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50.30392]
11. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: an attentional integration. Hum Factors. 2010;52(3):381-410. [doi: 10.1177/0018720810376055] [Medline: 21077562]
12. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc. 2012;19(1):121-127. [FREE Full text] [doi: 10.1136/amiajnl-2011-000089] [Medline: 21685142]
13. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res. 2020;22(6):e15154. [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
14. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health. 2021;3(9):e599-e611. [doi: 10.1016/s2589-7500(21)00132-1]

XSL•FO
RenderX

15. Lambert SI, Madi M, Sopka S, Lenes A, Stange H, Buszello C, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. NPJ Digit Med. 2023;6(1):111. [FREE Full text] [doi: 10.1038/s41746-023-00852-5] [Medline: 37301946]
16. Shi H, Huang J, Yang J, Han M. Barriers and facilitators to patient acceptance of artificial intelligence in health care: systematic review. J Med Internet Res. 2026;28:e80581. [FREE Full text] [doi: 10.2196/80581] [Medline: 42101346]
17. Hogg HDJ, Al-Zubaidy M, Talks J, Denniston AK, Kelly CJ, Malawana J, et al. Stakeholder perspectives of clinical artificial intelligence implementation: systematic review of qualitative evidence. J Med Internet Res. 2023;25:e39742. [FREE Full text] [doi: 10.2196/39742] [Medline: 36626192]
18. Kuo RYL, Freethy A, Smith J, Hill R, Jerome D, Harriss E, et al. Stakeholder perspectives towards diagnostic artificial intelligence: a co-produced qualitative evidence synthesis. EClinicalMedicine. 2024;71:102555. [FREE Full text] [doi: 10.1016/j.eclinm.2024.102555] [Medline: 38549586]
19. Tun HM, Rahman HA, Naing L, Malik OA. Trust in artificial intelligence-based clinical decision support systems among health care workers: systematic review. J Med Internet Res. 2025;27:e69678. [FREE Full text] [doi: 10.2196/69678] [Medline: 40772775]
20. Hilbers D, Nekain N, Bates A, Nunez J. Patient attitudes toward artificial intelligence in cancer care: scoping review. JMIR Cancer. 2025;11:e74010-e74010. [FREE Full text] [doi: 10.2196/74010] [Medline: 40845395]
21. Goh S, Goh RSJ, Chong B, Ng QX, Koh GCH, Ngiam KY, et al. Challenges in implementing artificial intelligence in breast cancer screening programs: systematic review and framework for safe adoption. J Med Internet Res. 2025;27:e62941. [doi: 10.2196/62941] [Medline: 40373301]
22. Kikuchi S, Sakata M, Hasegawa T, Wada S, Funada S, Makishi M, et al. Implementation of artificial intelligence in palliative and supportive care for people with cancer: a scoping review. Palliat Med. 2026;40(6):692-704. [FREE Full text] [doi: 10.1177/02692163261416261] [Medline: 41645881]
23. White BM, Aitken GL, Zink JA, Kheirkhah Rahimabad P, Kumsa FA, Hashtarkhani S, et al. Empathic AI for patient-centered cancer care: a scoping review of patient navigation, support, and clinical practice. JMIR Cancer. 2026;12:e82336-e82336. [FREE Full text] [doi: 10.2196/82336] [Medline: 41955525]
24. Wenderott K, Krups J, Weigl M, Wooldridge AR. Facilitators and barriers to implementing AI in routine medical imaging: systematic review and qualitative analysis. J Med Internet Res. 2025;27:e63649. [doi: 10.2196/63649] [Medline: 40690758]
25. Stern C, Lizarondo L, Carrier J, Godfrey C, Rieger K, Salmond S, et al. Methodological guidance for the conduct of mixed methods systematic reviews. JBI Evid Synth. 2020;18(10):2108-2118. [doi: 10.11124/JBISRIR-D-19-00169] [Medline: 32813460]
26. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
27. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Koffel JB, et al. PRISMA-S Group. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst Rev. 2021;10(1):39. [FREE Full text] [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
28. Fitzek S, Pohn B, Abbasian Ardakani A, Hatamikia S, Choi KE. Stakeholder acceptance, trust, and implementation of artificial intelligence in oncology: a mixed methods systematic review. Open Science Framework. 2026. URL: https://osf.io/nmep8 [accessed 2026-11-08]
29. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol. 2016;75:40-46. [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
30. Hong QN, Pluye P, Fabregues S, Bartlett G, Boardman F, Cargo M. Mixed Methods Appraisal Tool (MMAT), version 2018. Mixed Methods Appraisal Tool. Montreal, QC. McGill University; 2018. URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf [accessed 2026-11-08]
31. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1171-1179. [FREE Full text] [doi: 10.1007/s11136-017-1765-4] [Medline: 29260445]
32. Carstensen FV, Irankunda BB, Batista E, Van den Bongard D, Spanic T, Gernaat SA, et al. Perspectives on artificial intelligence-generated chronic disease risk predictions in early breast cancer: an international survey among radiation oncologists. Acta Oncol. 2026;65:553-556. [doi: 10.2340/1651-226X.2026.45901] [Medline: 42312821]
33. Togher D, Dean G, Moon J, Mayola R, Medina A, Repec J, et al. Evolution of radiology staff perspectives during artificial intelligence (AI) implementation for expedited lung cancer triage. Clin Radiol. 2025;81:106704. [doi: 10.1016/j.crad.2024.09.010] [Medline: 39443240]
34. Chan J, Parker L, Carter S, Nickel B, Carroll S. Radiation oncology patients' perceptions of artificial intelligence and machine learning in cancer care: a multi-centre cross-sectional study. Radiother Oncol. 2025;207:110891. [FREE Full text] [doi: 10.1016/j.radonc.2025.110891] [Medline: 40233873]

XSL•FO
RenderX

35. Anjara SG, Janik A, Dunford-Stenger A, Mc Kenzie K, Collazo-Lorduy A, Torrente M, et al. Examining explainable clinical decision support systems with think aloud protocols. PLoS One. 2023;18(9):e0291443. [FREE Full text] [doi: 10.1371/journal.pone.0291443] [Medline: 37708135]
36. Sarchosoglou A, Genitsarios I, Silvis-Cividjian N, Papavasileiou P, Bakas A, Papanikolaou N, et al. An AI-assisted, failure mode-based toolkit for proactive risk management in radiotherapy: a feasibility study. Tech Innov Patient Support Radiat Oncol. 2026;38:100404. [FREE Full text] [doi: 10.1016/j.tipsro.2026.100404] [Medline: 42099501]
37. Browning L, Ghosh A, Dolton M, Hutton K, Birks J, Scheffer R, et al. Exploring patient and clinician opinions, perspectives and acceptance of the use of artificial intelligence in the histological diagnosis of prostate cancer. BJUI Compass. 2025;6(11):e70108. [doi: 10.1002/bco2.70108] [Medline: 41221432]
38. Rezaeian O, Ghorbanichemazkati E, Bayrak AE, Asan O. Clinician perspectives on trust and adoption of AI in breast cancer diagnosis. IISE Trans Healthc Syst Eng. 2026;16(1):32-48. [doi: 10.1080/24725579.2025.2611435]
39. Thavanesan N, Naiseh M, Terol M, Rahman SA, Hill SL, Parfitt C, et al. Oesophageal cancer multi-disciplinary tool: a co-designed, externally validated, machine learning tool for oesophageal cancer decision making. EClinicalMedicine. 2025;89:103527. [FREE Full text] [doi: 10.1016/j.eclinm.2025.103527] [Medline: 41079022]
40. Rezaeian O, Bayrak AE, Asan O. Explainability and AI confidence in clinical decision support systems: effects on trust, diagnostic performance, and cognitive load in breast cancer care. Int J Hum Comput Interact. 2025;42(6):4477-4497. [doi: 10.1080/10447318.2025.2539458]
41. Heising LM, Ou CX, Verhaegen F, Wolfs CJA, Hoebers F, Jacobs MJ. 2723: trust in artificial intelligence in radiotherapy: a survey. Radiother Oncol. 2024;194:S2857-S2860. [doi: 10.1016/s0167-8140(24)02882-2]
42. Huet-Dastarac M, van Acht NMC, Maruccio FC, van Aalst JE, van Oorschodt JCJ, Cnossen F, et al. Quantifying and visualising uncertainty in deep learning-based segmentation for radiation therapy treatment planning: what do radiation oncologists and therapists want? Radiother Oncol. 2024;201:110545. [doi: 10.1016/j.radonc.2024.110545] [Medline: 39326521]
43. Rogowski V, Svalkvist A, Maspero M, Janssen T, Maruccio FC, Gorgisyan J, et al. Impact of deep learning model uncertainty on manual corrections to MRI-based auto-segmentation in prostate cancer radiotherapy. J Appl Clin Med Phys. 2025;26(9):e70221. [doi: 10.1002/acm2.70221] [Medline: 40849835]
44. Kenig N, Muntaner Vives A, Monton Echeverria J. Is my doctor human? Acceptance of AI among patients with breast cancer. Plast Reconstr Surg Glob Open. 2024;12(10):e6257. [doi: 10.1097/GOX.0000000000006257] [Medline: 39410970]
45. de Vries CF, Morrissey BE, Duggan D, Staff RT, Lip G. Screening participants' attitudes to the introduction of artificial intelligence in breast screening. J Med Screen. 2021;28(3):221-222. [doi: 10.1177/09691413211001405] [Medline: 33715512]
46. Khairat S, Mehraby H, Masoumi S, Coffel M, Rockey-Bartlett C, Huang A, et al. Evaluation of cancer survivors' experience of using AI-based conversational tools: qualitative study. JMIR Cancer. 2025;11:e77390-e77390. [FREE Full text] [doi: 10.2196/77390] [Medline: 41237331]
47. Zacharias L, Hesso I, Kayyali R, Charalambous A, Triumbari EKA, Chouvarda I, et al. Evaluation of the INCISIVE services in cancer imaging: a feasibility study. Semin Oncol Nurs. 2026;42(1):152088. [doi: 10.1016/j.soncn.2025.152088] [Medline: 41500909]
48. Talman M, Alam L, Mize D, Kumah-Crystal Y, Nall C, McCoy AB, et al. Ambient artificial intelligence scribes in pediatric hematology-oncology: early implementation of DAX copilot. Appl Clin Inform. 2026;17(2):374-379. [FREE Full text] [doi: 10.1055/a-2873-3894] [Medline: 42114818]
49. Toussi N, Zhang C, Kang J, Licitra EJ. Impact of AI scribes on physician productivity and workflow in medical oncology. JCO Oncol Pract. 2026;22(10):1797-1804. [doi: 10.1200/OP-25-00612] [Medline: 41671532]
50. Wenderott K, Krups J, Luetkens JA, Gambashidze N, Weigl M. Prospective effects of an artificial intelligence-based computer-aided detection system for prostate imaging on routine workflow and radiologists' outcomes. Eur J Radiol. 2024;170:111252. [FREE Full text] [doi: 10.1016/j.ejrad.2023.111252] [Medline: 38096741]
51. Jonnalagedda-Cattin M, Moukam Datchoua AM, Yakam VF, Kenfack B, Petignat P, Thiran J, et al. Barriers and facilitators to the preadoption of a computer-aided diagnosis tool for cervical cancer: qualitative study on health care providers' perspectives in Western cameroon. JMIR Cancer. 2025;11:e50124. [FREE Full text] [doi: 10.2196/50124] [Medline: 39908553]
52. van Heerden J, Sandri L, Geel J. Artificial intelligence in pediatric oncology in Africa: a survey of awareness, use, and readiness amongst healthcare workers. Pediatr Blood Cancer. 2026;73(6):e70253. [doi: 10.1002/1545-5017.70253] [Medline: 41852264]
53. Kritz M, Holawatsch E, Behrens DA, Hyll W. Barriers, facilitators, and intention to use AI for breast cancer diagnosis: mixed methods study among Austrian physicians with and without AI experience. J Med Internet Res. 2026;28:e80274. [FREE Full text] [doi: 10.2196/80274] [Medline: 42262843]
54. Dommaraju S, Ramesh K, Mendenhall M, MacAndrew A, Tubadeza E, Vangala SS, et al. Cancer patients' comfort with artificial intelligence in oncology care and first-round "AI chemo-teach" quality improvement initiatives. J Clin Oncol. 2026;44(16_suppl):e13652. [doi: 10.1200/jco.2026.44.16_suppl.e13652]
55. Gaube S, Biebl I, Engelmann MKM, Kleine A, Lermer E. Comparing preferences for skin cancer screening: AI-enabled app vs dermatologist. Soc Sci Med. 2024;349:116871. [doi: 10.1016/j.socscimed.2024.116871]

XSL•FO
RenderX

56. Hantel A, Walsh TP, Marron JM, Kehl KL, Sharp R, Van Allen E, et al. Perspectives of oncologists on the ethical implications of using artificial intelligence for cancer care. JAMA Netw Open. 2024;7(3):e244077. [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.4077] [Medline: 38546644]
57. Stavropoulou C, Scarbrough H, Rawlinson J, Cui M, Baldwin D, Woznitza N. Implementing AI innovation in radiology departments in the English NHS: a qualitative study on the experiences of professionals, patient groups and innovators. Front Digit Health. 2026;8:1736911. [FREE Full text] [doi: 10.3389/fdgth.2026.1736911] [Medline: 41924179]
58. Oyewusi W, Vasquez Osorio EM, Nenadic G, MacGregor I, Price G. Data, dialogue, and design: patient and public involvement and engagement for natural language processing with real-world cancer data. Front Digit Health. 2025;7:1560757. [doi: 10.3389/fdgth.2025.1560757] [Medline: 40444182]
59. Swart R, Boersma L, Fijten R, van Elmpt W, Cremers P, Jacobs MJG. Implementation strategy for artificial intelligence in radiotherapy: can implementation science help? JCO Clin Cancer Inform. 2024;8:e2400101. [doi: 10.1200/CCI.24.00101] [Medline: 39705640]
60. Vroegindeweij M, Heising L, Ou CX, van Elmpt W, Jacobs M. 3864 decoding AI-adoption in radiotherapy: a multi-case qualitative comparative analysis. Radiother Oncol. 2025;206:S3437-S3438. [doi: 10.1016/s0167-8140(25)02014-6]
61. Erul E, Aktekin Y, Danışman FB, Gümüştaş Ş, Aktekin BS, Yekedüz E, et al. Perceptions, attitudes, and concerns on artificial intelligence applications in patients with cancer. Cancer Control. 2025;32:10732748251343245. [FREE Full text] [doi: 10.1177/10732748251343245] [Medline: 40407404]
62. Hölgyesi Á, Zrubka Z, Gulácsi L, Baji P, Haidegger T, Kozlovszky M, et al. Robot-assisted surgery and artificial intelligence-based tumour diagnostics: social preferences with a representative cross-sectional survey. BMC Med Inform Decis Mak. 2024;24(1):87. [FREE Full text] [doi: 10.1186/s12911-024-02470-x] [Medline: 38553703]
63. Jagemann I, Wensing O, Stegemann M, Hirschfeld G. Acceptance of medical artificial intelligence in skin cancer screening: choice-based conjoint survey. JMIR Form Res. 2024;8:e46402. [FREE Full text] [doi: 10.2196/46402] [Medline: 38214959]
64. Lewandowska M, Street D, Yim J, Jones S, Viney R. Preferences for adopting artificial intelligence in radiation therapy treatment: a discrete choice experiment. Value Health. 2026;29(2):267-276. [FREE Full text] [doi: 10.1016/j.jval.2025.09.013] [Medline: 41015332]
65. Brooke J. SUS: a "quick and dirty" usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry. London. Taylor & Francis; 1996:189-194.
66. Holzinger A, Carrington A, Müller H. Measuring the quality of explanations: the system causability scale (SCS): comparing human and machine explanations. Kunstliche Intell (Oldenbourg). 2020;34(2):193-198. [FREE Full text] [doi: 10.1007/s13218-020-00636-z] [Medline: 32549653]
67. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the patient activation measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res. 2004;39(4 Pt 1):1005-1026. [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
68. Menz BD, Scarfo NL, Cornelisse E, Shahnam A, Chin-Yee B, Rickard C, et al. Perspectives of people affected by cancer toward generative AI in oncology: a scoping review of applications, evidence gaps, and future priorities. Eur J Cancer. 2026;246:116990. [FREE Full text] [doi: 10.1016/j.ejca.2026.116990] [Medline: 42628140]

Abbreviations

aOR: adjusted odds ratio
COSMIN: Consensus-Based Standards for the Selection of Health Measurement Instruments
JBI: Joanna Briggs Institute
MMAT: Mixed Methods Appraisal Tool
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability
PRESS: Peer Review of Electronic Search Strategies
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-S: PRISMA Extension for Literature Searches
QC: quality control

Edited by M Balcarras; submitted 15.May.2026; peer-reviewed by S Podder, W Yu Jen ; comments to author 29.Jul.2026; revised version received 06.Oct.2026; accepted 09.Oct.2026; published 24.Nov.2026

Please cite as:
Fitzek S, Pohn B, Mohebbi A, Ardakani AA, Choi K-E(, Hatamikia S
Stakeholder Acceptance, Trust, and Implementation of AI in Oncology: Mixed Methods Systematic Review
J Med Internet Res 2026;28:e101445
URL: https://www.jmir.org/2026/1/e101445
doi: 10.2196/101445
PMID:

XSL•FO
RenderX

©Sebastian Fitzek, Birgit Pohn, Alisa Mohebbi, Ali Abbasian Ardakani, Kyung-Eun (Anna) Choi, Sepideh Hatamikia. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 24.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX