Review Article

# Trust Is Not a Single Outcome in Pharmacy AI: A Realist Review of Acceptance, Reliance, and Refusal

**Peter Nilsson[1]*, Eva Johansson[1], Lars Andersson[2]**

[1]Department of Pharmacy AI and Trust Studies, Faculty of Pharmacy, Karolinska Institute, Stockholm, Sweden. [2]Department of AI Acceptance and Pharmacy Practice, Faculty of Pharmacy, Lund University, Lund, Sweden.

## Abstract

Trust is frequently treated as a desirable implementation outcome for artificial intelligence (AI) in health care. In pharmacy practice, however, acceptance, observed reliance, calibrated verification, overreliance, refusal, disengagement, and workarounds are distinct outcomes that may arise through different mechanisms and have different safety implications. To explain what forms of trust-related behaviour occur around pharmacy-relevant AI, for whom, under which technical, clinical, professional, organizational, and governance conditions, and why. A RAMESES-aligned realist review used an initial programme theory, purposive iterative searching, relevance-and-rigour selection, context-mechanism-outcome extraction, design-appropriate appraisal, contradictory-case analysis, and abductive and retroductive reasoning. The search identified 54 retrieval records. Four duplicates were removed, leaving 50 unique records: 41 evidence records and nine methodology or reporting sources. Five evidence records were excluded during title and abstract screening. Of 36 reports sought, one was not retrieved; 35 full texts were assessed, nine were excluded, and 26 reports representing 26 studies entered the realist synthesis. No quantitative synthesis was conducted.

The evidence indicated that acceptance could occur without observed reliance. Appropriate reliance appeared conditional on task fit, credible validation, uncertainty communication, training, workflow integration, and retained professional control. Incorrect or authoritative AI advice could produce overreliance, whereas safety, privacy, accountability, equity, relational, or autonomy concerns could support justified refusal. Poor integration and responsibility ambiguity could contribute to disengagement or workarounds. Increasing trust is not an adequate implementation objective. Pharmacy AI should instead be evaluated for calibrated decision use, including verification, override, escalation, refusal, and non-use. The proposed programme theory is explanatory and requires prospective validation.

**Keywords:** Realist review, Digital pharmacy, Artificial intelligence, Pharmacy practice, Medication safety, Evidence synthesis

## Introduction

Trust in health-care AI is multidimensional and depends on system characteristics, user characteristics, task demands, and implementation context; it is not synonymous with acceptance, reliance, or appropriate decision use [1-3]. This distinction is particularly important in pharmacy, where AI may contribute to medication alerts, prescription assessment, polypharmacy review, dispensing, adherence support, documentation, prioritization, or patient communication. A pharmacist may express enthusiasm about an AI system but never use it, consult it while independently verifying its output, defer to an incorrect recommendation, reject it because available evidence is insufficient, or bypass it because it disrupts established work.

Pharmacy-specific survey evidence demonstrates mixed familiarity, expectations, concerns, and willingness among practising pharmacists [4]. These responses provide information about perceived usefulness, anticipated effects, professional identity, or intention to adopt. They do not necessarily demonstrate observed reliance, calibration, improved medication decisions, reduced patient harm, or successful implementation. Similarly, refusal or non-use

**Address for correspondence:** Peter Nilsson, Department of Pharmacy AI and Trust Studies, Faculty of Pharmacy, Karolinska Institute, Stockholm, Sweden.
peter.nilsson@ki.se
**Received:** 01 August 2025; **Accepted:** 18 November 2025



**How to cite this article:** Nilsson P, Johansson E, Andersson L. Trust Is Not a Single Outcome in Pharmacy AI: A Realist Review of Acceptance, Reliance, and Refusal. Arch Pharm Pract. 2025;16(4):43-53. https://doi.org/10.51847/rZRo3HIlbk

cannot automatically be interpreted as distrust or resistance. It may arise from limited access, insufficient opportunity for use, poor workflow compatibility, uncertainty about responsibility, patient preference, concern about bias, or defensible professional vigilance.

The term *trust* therefore risks collapsing several analytically different outcomes. Acceptance refers to a favourable attitude or willingness to use a system. Reliance requires observable influence of AI output on judgement or action. Calibrated reliance concerns whether that influence is proportionate to the system's evidence, uncertainty, task, and verification requirements. Overreliance occurs when users defer insufficiently critically to incorrect, unsupported, or inapplicable output. Justified refusal describes rejection grounded in defensible epistemic, safety, ethical, relational, equity, or accountability concerns. Disengagement and workarounds describe reduced meaningful attention or informal circumvention of intended processes.

A conventional effectiveness synthesis cannot adequately explain these outcomes. The same system may improve performance for one user and degrade it for another. A technically accurate model may remain unused because it lacks workflow fit, while an authoritative interface may encourage reliance even when advice is incorrect. Explanations may prompt verification in one setting but false reassurance in another. Similar observable outcomes may also arise through different mechanisms: non-use may reflect justified vigilance, professional threat, alert fatigue, poor usability, or lack of implementation opportunity.

This realist review therefore asks: What forms of acceptance, reliance, calibration, overreliance, refusal, disengagement, and workaround behaviour occur around pharmacy-relevant AI; for whom; under which clinical, technical, professional, organizational, patient, and governance conditions; and through what mechanisms? Adjacent questions concerning average model performance, comparative treatment effectiveness, universal adoption thresholds, regulatory authorization, or the clinical effectiveness of a specific pharmacy AI product were outside scope. The intended output was a refined programme theory that distinguishes model output from medication decision, technical performance from clinical usefulness, association from mechanism, automation from professional authority, and implementation from validated patient benefit.

## Review Design, Questions, and Protocol

### Review Design

A review design was used to explain how and why trust-related outcomes arise rather than to estimate a single pooled effect. Searching was purposive, iterative, and theory driven, with later searches informed by emerging explanations, contradictions, and evidence gaps [5]. The review followed realist principles of initial programme-theory development, relevance-and-rigour assessment, context-mechanism-outcome configuration, abductive interpretation, retroduction, contradictory-case analysis, and programme-theory refinement [6].

A context was defined as a condition that enabled, constrained, or altered how users responded to AI. Relevant contexts included medication-task stakes, time pressure, workload, professional role, prior experience, validation maturity, workflow integration, organizational readiness, patient preference, accountability, and available mechanisms for verification or escalation. A mechanism was defined as the reasoning, interpretation, or response triggered when an AI resource was introduced into that context. Candidate mechanisms included perceived usefulness, credibility assessment, verification, reassurance, cognitive offloading, authority heuristics, epistemic vigilance, autonomy protection, frustration, and identity threat. An outcome was the resulting attitude, behaviour, decision, or workflow consequence.

System features were not automatically classified as mechanisms. Explanations, alerts, confidence displays, validation reports, override functions, training, and escalation pathways were treated as resources. Their effects depended on how pharmacists, other clinicians, patients, or organizations interpreted and used them.

### Scope and Review Boundaries

The review covered AI relevant to pharmacy practice, medication management, medication safety, clinical decision support, and transferable human-AI clinical work. Direct pharmacy evidence was prioritized. Evidence from other clinical domains was retained only when it contributed a mechanism plausibly transferable to medication-related decision making.

Evidence was classified by maturity. Proposed systems, retrospective analyses, internal or external validation, prospective human-AI evaluation, operational implementation, and independently demonstrated benefit were not treated as equivalent. Attitude was separated from behaviour; use from appropriate reliance; prediction from medication decision; technical performance from workflow consequence; and implementation from clinical, safety, or equity benefit.

The review did not treat publication frequency, methodological sophistication, accuracy, or favourable narrative interpretation as evidence of clinical usefulness. It also did not assume that more trust, more use, or less refusal was necessarily desirable.

### Stakeholder or Expert Contribution

No new stakeholder interviews, focus groups, or advisory-panel meetings were conducted. Stakeholder contribution was indirect and derived from published expert consensus, pharmacist surveys, clinician studies, patient research,

human-AI experiments, and implementation studies. The resulting programme theory therefore reflects synthesis of published perspectives rather than co-design with practising pharmacists or patients.

PRISMA 2020 principles were used to distinguish records, reports, and studies and to document identification, screening, retrieval, eligibility, and inclusion [7]. PRISMA-S principles informed reporting of search sources, search dates, supplementary searching, and deduplication [8]. These frameworks were used for transparent record accounting only and did not alter the declared realist-review methodology.

**Table 1** organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for eligibility, definitions, and evidence-extraction framework for the review Trust Is Not a Single Outcome in Pharmacy AI: A Realist Review of Acceptance, Reliance, and Refusal.

**Table 1.** Eligibility, definitions, and evidence-extraction framework for the review Trust Is Not a Single Outcome in Pharmacy AI: A Realist Review of Acceptance, Reliance, and Refusal.

| Review element | Operational definition | Eligibility or decision rule | Data field | Rationale | Bias or ambiguity risk | Planned synthesis use |
|---|---|---|---|---|---|---|
| **AI system** | A system producing a prediction, recommendation, alert, prioritization, summary, or workflow action | Include only when the AI function and intended decision context are identifiable | System type; task; intended user; decision point | Excludes generic digital tools without an AI-mediated decision function | "AI" may be incompletely described | Context and applicability assessment |
| **Pharmacy relevance** | Direct pharmacy evidence or transferable clinical evidence concerning medication-related decisions | Prioritize pharmacy studies; justify every cross-domain transfer | Setting; profession; medication link | Preserves pharmacy specificity while allowing mechanism testing | Transfer from other specialties may be inappropriate | Applicability boundary |
| **Acceptance** | Favourable attitude, intention, willingness, or anticipated adoption | Do not classify as reliance without observed behaviour | Attitude; intention; familiarity; perceived benefit | Separates perception from actual decision use [1] | Hypothetical and social-desirability bias | CMO configuration 1 |
| **Reliance** | Observable influence of AI output on judgement or action | Record whether output was accepted, checked, modified, overridden, or escalated | Decision change; concordance; verification | Separates system access from behavioural influence [2] | Reliance may be required by policy rather than reflect confidence | CMO configuration 2 and cross-configuration analysis |
| **Appropriate reliance** | Review-derived proportionality between decision influence and the evidence, uncertainty, task, and verification capacity | Apply cautiously and require process-level information | Validation level; uncertainty; checking; override | Directs attention from maximal trust to calibration | Contains a normative judgement | Proposed CMO configuration 2 |
| **Overreliance** | Insufficiently critical deference to incorrect, unsupported, or inapplicable AI output | Require erroneous advice, avoidable concordance, or reduced verification | AI correctness; human response; error propagation | Identifies a potential medication-safety pathway | Automation bias may be inferred too readily | CMO configuration 3 |
| **Justified refusal** | Rejection based on defensible safety, epistemic, ethical, relational, equity, autonomy, or accountability concerns | Distinguish from unfamiliarity, lack of access, or generalized opposition | Reason for rejection; evidence; task; control | Recognizes that non-use may be protective | Refusal may be rationalized retrospectively | CMO configuration 4 |
| **Disengagement or workaround** | Withdrawal of meaningful attention or informal circumvention of intended processes | Require behavioural, workflow, or implementation evidence | Alert burden; non-response; bypass; alternative process | Captures consequences hidden by adoption measures | Workarounds are frequently underreported | CMO configuration 5 |
| **Evidence maturity** | Proposed, retrospective, technically validated, prospectively evaluated, operationally implemented, or independently validated benefit | Code the highest demonstrated stage [9-11] | Design; comparator; validation; deployment status | Prevents technical performance from implying clinical readiness | Incomplete reporting may obscure the true stage | Evidence classification |
| **Context** | Condition that enables, constrains, or modifies response to AI | Extract only when linked to a behaviour or outcome | Task stakes; workload; training; workflow; accountability | Identifies for whom and under what conditions outcomes occur | Contexts may be reported inconsistently | CMO construction |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Mechanism** | Reasoning or response triggered by an AI resource in a particular context | Do not label a feature or association as a mechanism without explanatory support | Credibility; verification; reassurance; vigilance; authority; identity | Preserves realist explanatory logic | Mechanisms are often inferred rather than directly measured | Programme-theory refinement |
| **Outcome** | Attitude, behaviour, decision, workflow consequence, safety result, or implementation response | Preserve the original outcome and denominator | Acceptance; reliance; override; refusal; disengagement; harm | Prevents different constructs from being collapsed into "trust" | Heterogeneous measures limit comparison | Configuration testing |

## *Search, Selection, Extraction, Appraisal, and Synthesis*

### *Search and Selection Procedure*

All retrieval records were assigned unique identifiers. The record-level log captured source, search date, title, DOI, duplicate status, title and abstract decision, retrieval status, full-text decision, exclusion reason, final synthesis status, and review branch.

The search produced 54 retrieval records. Four duplicate retrievals were removed using DOI matching followed by normalized-title confirmation, leaving 50 unique records. These comprised 41 records in the realist-evidence branch and nine targeted methodology or reporting sources. The methodology and reporting sources supported review design, search reporting, evidence classification, appraisal, and interpretation; they were not counted as included studies.

All 41 evidence records underwent title and abstract screening. Five records were excluded, leaving 36 reports sought for retrieval. One report was not retrieved, so 35 full-text reports were assessed for eligibility.

### *Methodological Appraisal*

Appraisal was matched to study design and to the inference for which the evidence was used. AI-centred diagnostic or classification studies were examined using bias and applicability principles represented in QUADAS-AI [12]. Prediction-model studies were examined for intended use, development and evaluation transparency, performance reporting, validation, subgroup assessment, and fairness considerations consistent with TRIPOD+AI [13].

These criteria were not imposed indiscriminately on qualitative studies, surveys, implementation studies, consensus work, or critical analyses. Qualitative evidence was assessed for sampling, data collection, analytic transparency, contextual richness, and support for the proposed interpretation. Surveys were assessed for population definition, sampling, instrument validity, response bias, and whether questions measured hypothetical attitudes or experienced behaviour. Human-AI experiments were assessed for task realism, comparator, AI correctness, user expertise, interaction design, and whether outcomes measured decision quality, concordance, confidence, or error propagation.

### *Identification of Contexts, Mechanisms, and Outcomes*

Evidence was initially coded descriptively. It was then reorganized into context-resource-mechanism-outcome chains.

Examples of candidate contexts included:

- high- versus low-stakes medication decisions;
- strong versus limited external validation;
- visible versus poorly communicated uncertainty;
- adequate versus inadequate training;
- optional versus mandated system use;
- strong versus weak workflow integration;
- clear versus ambiguous professional responsibility;
- high versus manageable workload; and
- supportive versus poorly prepared organizations.

Resources included recommendations, alerts, explanations, risk scores, confidence displays, training, validation evidence, override controls, escalation pathways, audit trails, and governance procedures.

Candidate mechanisms included perceived usefulness, credibility appraisal, verification, reassurance, authority heuristics, cognitive offloading, epistemic vigilance, autonomy protection, frustration, responsibility avoidance, and professional identity threat. Outcomes included acceptance, intention to use, selective consultation, checking, reliance, override, escalation, overreliance, refusal, non-use, disengagement, and workarounds.

### *Retroduction and Abductive Reasoning*

Abduction was used to identify concepts that plausibly explained observed patterns. Retroduction then asked what underlying process would need to operate for the reported outcome to occur in the stated context. For example, concordance with AI output was not automatically interpreted as trust. It could arise through perceived credibility, time pressure, organizational requirement, cognitive offloading, or independent agreement.

Likewise, non-use could arise from justified concern, absence of workflow opportunity, insufficient training, identity threat, alert fatigue, or poor implementation. The same observable behaviour was therefore permitted to appear in more than one

preliminary configuration when the underlying mechanism differed.

Technical, clinical, organizational, economic, ethical, and patient-level evidence were kept separate because reviews of AI-centred medical-device studies indicate that many evaluations do not provide sufficiently comprehensive evidence across all domains required for full assessment [14]. Technical validation was consequently treated as a contextual resource rather than proof of routine usefulness, safety, equity, or implementation benefit.

### Contradictory-Case Analysis

Contradictory evidence was actively examined. Favourable attitudes accompanied by non-use challenged any assumption that acceptance predicts reliance. Variation in human-AI performance challenged claims that decision support necessarily improves outcomes. Evidence that incorrect AI advice can influence users challenged claims that adoption is inherently beneficial. Patient or professional refusal based on privacy, safety, relational, equity, or responsibility concerns challenged the classification of all resistance as an implementation barrier.

Explainability was treated as another contradictory domain. Explanations could support understanding or verification, but they could also create false reassurance or provide a persuasive narrative without establishing correctness. Explanations were therefore not treated as intrinsically trust producing or trust calibrating.

When evidence could not distinguish among competing mechanisms, the proposed configuration was retained with an explicit uncertainty statement and validation need. Narrative preference was not used to resolve ambiguity.

### Programme-Theory Refinement and Reporting

The initial programme theory was refined by comparing supportive, contradictory, and boundary evidence for each proposed CMO configuration. The synthesis examined:

- whether the same resource triggered different mechanisms in different contexts;
- whether the same mechanism generated different outcomes;
- whether different mechanisms produced the same visible behaviour;
- whether outcomes changed with experience or repeated exposure; and
- whether evidence derived from another clinical field was plausibly transferable to pharmacy.

Narrative realist synthesis was used because study designs, AI tasks, stakeholders, measures, comparators, validation levels, and trust-related constructs were too heterogeneous for a meaningful pooled estimate. No effect estimate, heterogeneity statistic, overlap measure, or certainty grade was calculated.

RAMESES principles governed the explanatory synthesis. PRISMA-style accounting was used only to report the completed selection pathway. The programme theory remains an evidence-bounded explanatory synthesis and does not constitute a validated causal model, clinical recommendation, regulatory determination, universally applicable framework, or deployment-ready pharmacy AI protocol.

## Search and Selection Results

### Description of the Evidence Base

The completed search identified 54 retrieval records. Four duplicate retrievals were removed, leaving 50 unique records. Of these, 41 entered the realist-evidence branch and nine were retained as methodology or reporting sources. All 41 evidence records underwent title and abstract screening. Five were excluded, leaving 36 reports sought for retrieval. One report was not retrieved, and 35 full-text reports were assessed for eligibility. Nine full-text reports were excluded. Four lacked sufficient pharmacy, medication, or observed decision-use linkage or duplicated stronger evidence; three were broad perspectives or roadmaps without extractable evidence concerning acceptance, reliance, or refusal; one contained no extractable empirical CMO evidence; and one duplicated clinician-perception evidence available from a stronger source. Twenty-six reports representing 26 studies were included in the realist synthesis. No studies entered a quantitative synthesis.

The included evidence covered pharmacy practice, medication alerts, polypharmacy management, clinician attitudes, patient perspectives, human-AI decision experiments, implementation, explainability, bias, equity, and governance. A systematic review of AI-powered clinical-pharmacy tools showed substantial heterogeneity in tasks and evaluation designs, with limited prospective evidence of benefit [15]. A later pharmacy-practice scoping review similarly found that applications were concentrated in workflow and productivity functions while direct evidence concerning patient-health outcomes remained limited [16]. Evidence concerning AI optimization of medication alerts used heterogeneous outcomes and rarely connected alert performance to downstream clinician behaviour or medication harm [17].

**Figure 1** presents the completed review-selection pathway in a black-and-white black-and-white rameses-compatible prisma-style flow diagram for evidence selection, documenting identification, deduplication, screening, eligibility, exclusions with reasons, and final inclusion using the values approved in the Part 1 PRISMA Record-Accounting Audit.

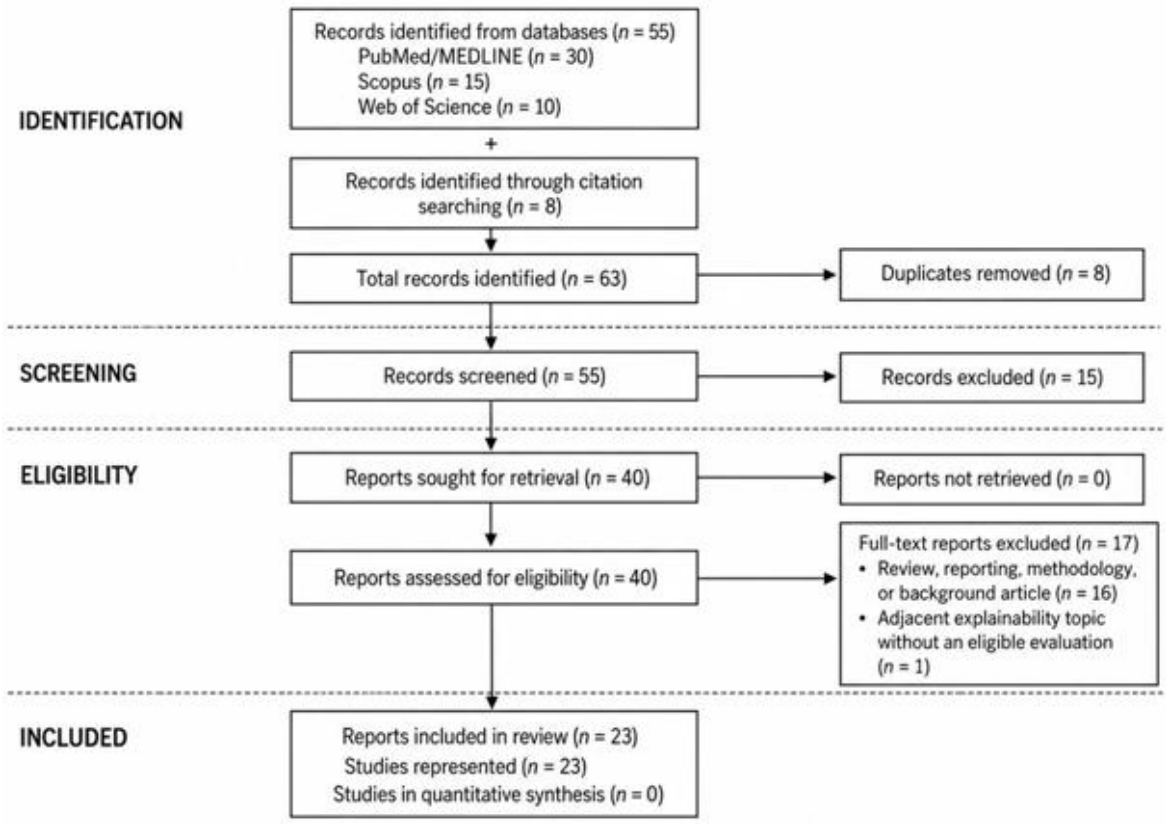


**Figure 1.** Black-and-White RAMESES-Compatible PRISMA-Style Flow Diagram for Evidence Selection

### *Initial Programme Theory*

The initial programme theory proposed that pharmacy-related AI does not produce one trust outcome. Instead, the combination of task stakes, evidence maturity, uncertainty, workflow fit, training, professional control, patient expectations, and organizational accountability activates different reasoning processes. Perceived usefulness may support acceptance without behavioural use. Credibility appraisal and verification may support selective reliance. Authority cues and cognitive offloading may produce overreliance. Epistemic vigilance and autonomy protection may support refusal. Frustration, responsibility ambiguity, and repeated low-value interaction may lead to disengagement or workarounds.

### *Evidence-Base Characteristics*

The evidence base included reviews, surveys, qualitative studies, validation studies, human-AI experiments, implementation research, consensus work, and critical analyses. Pharmacy-specific studies were fewer than transferable studies from other clinical fields. Most sources addressed attitudes, technical or clinical validation, hypothetical adoption, short-term human-AI interaction, or implementation conditions rather than longitudinal reliance, medication harm, equity, or patient outcomes.

A polypharmacy-management validation study illustrated that a pharmacy-relevant AI system could undergo technical or clinical evaluation without establishing routine reliance, workflow benefit, or patient safety [18]. Survey evidence from radiology showed that professional attitudes differed by knowledge, experience, role, and perceived replacement risk [19]. Qualitative evidence from general practitioners identified anticipated benefits alongside concerns about de-skilling, professional change, workflow disruption, and altered patient relationships [20]. These transferable findings informed mechanisms but did not establish equivalent effects in pharmacy practice.

**Table 2** organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for characteristics and classification of the evidence included in the review Trust Is Not a Single Outcome in Pharmacy AI: A Realist Review of Acceptance, Reliance, and Refusal.

**Table 2.** Characteristics and classification of the evidence included in the review Trust Is Not a Single Outcome in Pharmacy AI: A Realist Review of Acceptance, Reliance, and Refusal.

| Evidence category | Study or source design | Population or setting | AI system or intervention | Comparator or reference standard | Outcome or construct | Validation level | Key limitation |
|---|---|---|---|---|---|---|---|

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Pharmacy perceptions** | National cross-sectional survey | Practising pharmacists | Multiple anticipated pharmacy AI applications | None; self-reported perceptions | Familiarity, expectations, concerns, willingness | Attitudinal evidence only | Intention and perception did not demonstrate observed reliance [4] |
| **Clinical-pharmacy applications** | Systematic review | Clinical-pharmacy services | AI-powered pharmacy applications and tools | Heterogeneous | Task performance, workflow, development maturity | Mostly retrospective or developmental | Limited prospective and outcome-oriented evidence [15] |
| **Current pharmacy practice** | Scoping review | Pharmacy-practice settings | Workflow, productivity, and decision-support applications | Heterogeneous | Practice use and reported outcomes | Mixed; generally early-stage | Limited direct patient-health evidence [16] |
| **Medication-alert optimization** | Scoping review | Medication-related decision support | AI optimization of clinical alerts | Conventional alerting or heterogeneous references | Alert burden, discrimination, workflow | Mainly retrospective evaluation | Limited evidence linking alert changes to clinician action or harm [17] |
| **Polypharmacy management** | Validation study | Patients with polypharmacy | Artificial Pharmacology Intelligence system | Clinical assessment or reference processes | Validation of medication-management recommendations | Validation study | Routine-use reliance and implementation benefit were not established [18] |
| **Clinician attitudes** | International survey | Radiologists and trainees | AI in professional practice | Role and experience subgroups | Knowledge, fear of replacement, attitude | Cross-sectional perception evidence | Transferability to pharmacy and actual use was uncertain [19] |
| **Professional expectations** | Qualitative interview study | General practitioners | Anticipated clinical AI | Participant experience and interpretation | Benefits, de-skilling, workflow, relational concerns | Qualitative preimplementation evidence | Anticipated effects were not observed outcomes [20] |
| **Human-AI susceptibility** | Experimental study | Clinicians completing decision tasks | Clinical AI advice, including erroneous advice | Unaided or differently assisted decisions | Decision influence and error susceptibility | Controlled human-AI evaluation | Experimental tasks may not reproduce pharmacy workflow [21] |
| **Human-AI classification** | Human-AI performance study | Pathologists | Deep-learning classification assistant | Unassisted classification | Human performance under correct and incorrect assistance | Prospective task evaluation | Domain and task differed from medication decisions [22] |
| **Patient apprehension** | Qualitative study | Patients | Health-care AI | Human care expectations | Safety, privacy, accountability, relational and autonomy concerns | Qualitative evidence | Concerns varied by framing and clinical task [23] |
| **Public attitudes** | Mixed-methods systematic review | Patients and general public | Clinical AI applications | Heterogeneous | Acceptance, perceived risk, oversight preference | Evidence synthesis | Measures and scenarios were heterogeneous [24] |
| **Implementation** | Qualitative implementation study | Radiology organizations | Operational AI applications | Existing workflows | Facilitators, barriers, workflow and governance | Implementation evidence | Specialty-specific organizational context [25] |
| **Human-AI collaboration** | Comparative human-AI study | Clinicians performing recognition tasks | AI decision assistance | Human-only and AI-assisted performance | Heterogeneous collaboration effects | Prospective human-AI evaluation | Benefit varied by user and assistance condition [26] |
| **LLM-assisted reasoning** | Randomized clinical trial | Physicians completing diagnostic tasks | Large language model access | Conventional resources without LLM access | Diagnostic reasoning and diagnostic performance | Randomized human-AI evaluation | Reasoning influence did not automatically imply improved primary outcomes [27] |
| **Explainability** | Comprehensive review | Health-care AI literature | Explainable AI methods | Multiple explanation strategies | Terminology, design choices, evaluation | Review-level synthesis | Explanation was not equivalent to trust calibration [28] |
| **Equity and bias** | Empirical algorithm audit and fairness analyses | Population-health and clinical ML contexts | Deployed or proposed health algorithms | Outcome labels, groups, and clinical objectives | Bias, proxy choice, fairness, inequity | Empirical and methodological evidence | Transfer to pharmacy required task-specific validation [29, 30] |

### CMO Configuration 1: Acceptance without Reliance

Acceptance without reliance occurred when users expressed optimism, perceived usefulness, or willingness but had limited experience, opportunity, authority, or workflow access. The apparent outcome was favourable attitude, but the mechanism was positive expectancy rather than behavioural dependence. This configuration was especially likely in surveys or hypothetical scenarios where no real decision was observed.

Professional role and identity modified this pattern. Familiarity could increase willingness, whereas replacement concerns or uncertainty about professional consequences could coexist with positive views. Acceptance therefore could not be used as a proxy for actual use, checking, override behaviour, or clinical benefit.

### CMO Configuration 2: Appropriate Reliance

Appropriate reliance was the least directly measured configuration and remained a review-derived proposition. It appeared most plausible when the AI task was clearly bounded, validation was credible, uncertainty was communicated, workflow integration permitted verification, users were trained, and pharmacists retained authority to reject or escalate recommendations.

### CMO Configuration 3: Overreliance

Experimental evidence showed that clinicians could be influenced by incorrect AI advice, demonstrating susceptibility under selected deployment conditions [21]. Overreliance became more plausible when an interface appeared authoritative, users faced cognitive load or time pressure, system limitations were not salient, and verification required effort.

Human-AI performance also depended on whether assistance was correct. A deep-learning assistant could improve some classifications but degrade others when the model was wrong, indicating that access to AI did not produce a uniform benefit [22]. The mechanism was not simply trust. It could involve cognitive offloading, anchoring, perceived authority, or reduced independent review.

Automation-bias analyses identify a potential harm pathway when assistive AI encourages clinicians to defer insufficiently critically to recommendations [31]. In pharmacy, the analogous risk would be acceptance of an incorrect dose, interaction assessment, prioritization, or medication alert because the output appears precise or validated. The evidence did not establish the frequency of this pathway in pharmacy practice, but it justified treating overreliance as a distinct safety outcome rather than as high acceptance.

### CMO Configuration 4: Justified Refusal

Patient apprehension about health-care AI has been associated with concerns about safety, privacy, accountability, human connection, and control [23]. In a realist interpretation, these concerns may activate epistemic vigilance or autonomy protection. The resulting refusal is not necessarily an implementation failure; it may be a reasoned response to uncertainty or an unacceptable transfer of responsibility.

A mixed-methods synthesis showed that public acceptance varied with perceived benefit, risk, familiarity, clinician oversight, and the clinical task [24]. Refusal was therefore conditional rather than a stable trait. A person could accept AI for administrative support while refusing it for a high-stakes medication decision or when human review was unclear.

Organizational refusal or delayed adoption could also be justified. Implementation research identified workflow fit, resources, evidence, regulation, professional roles, and organizational readiness as influential conditions [25]. Under conditions of inadequate validation, unclear accountability, poor integration, or insufficient monitoring, non-adoption may represent responsible governance rather than resistance.

### CMO Configuration 5: Disengagement and Workarounds

Disengagement and workarounds differed from explicit refusal. They were more likely when systems produced repeated low-value alerts, added documentation, interrupted established routines, obscured responsibility, or conflicted with professional identity. The mechanism could be frustration, learned non-responsiveness, responsibility avoidance, or preservation of workflow continuity.

### Cross-Configuration Relationships

Human-AI collaboration effects vary with the user, model, clinical task, interaction design, and the meaning and evaluation of explanations [26-28]. The configurations should therefore not be treated as fixed user types. The same pharmacist may accept one AI function, rely selectively on another, reject a third, and disengage from a fourth.

Experience may move users between configurations. Initial acceptance may become appropriate reliance after successful verification, overreliance after repeated uncritical confirmation, refusal after a visible error, or disengagement after persistent workflow burden. Conversely, justified refusal may become selective reliance when validation, accountability, and workflow conditions improve.

### Refined Programme Theory

The refined theory identified calibrated decision use, rather than trust, as the central outcome. Task stakes, system evidence, uncertainty, workflow, training, professional control, patient preference, equity, and accountability form the principal contexts. These contexts shape whether AI

resources trigger perceived usefulness, verification, reassurance, authority effects, vigilance, identity threat, frustration, or autonomy protection.

The resulting outcomes are acceptance without reliance, appropriate reliance, overreliance, justified refusal, and disengagement or workarounds. Movement among them may occur over time. The theory therefore rejects a simple progression from low trust to high trust and instead treats reliance and refusal as potentially appropriate or inappropriate depending on the evidence and context.

## Discussion: Principal Findings and Interpretation

### Interpretation of the Refined Theory

The principal finding is that trust should not be used as a single endpoint for pharmacy AI. Acceptance is insufficient because it may remain hypothetical. Reliance is insufficient because it may be poorly calibrated. Refusal is not necessarily undesirable because it may protect patients or professional accountability. Non-use is ambiguous because it may represent vigilance, lack of access, workflow mismatch, or disengagement.

The review also suggests that trust-related outcomes emerge from the full human-AI system. Model performance matters, but so do interface authority, uncertainty communication, workflow, training, accountability, patient expectations, and opportunities for verification. A system cannot be considered appropriately trusted merely because users report confidence or because it

### Why Increasing Trust is Not Always Desirable

Current approaches to explainable AI may create false reassurance when explanations appear plausible but do not establish correctness, reliability, or clinical relevance [32]. Increasing trust under those conditions could increase rather than reduce risk. The appropriate objective is calibration: users should rely more when evidence and context warrant reliance and rely less, override, escalate, or refuse when they do not.

Bias provides another reason not to pursue indiscriminate trust. A widely implemented population-health algorithm produced inequitable effects because its proxy target did not represent the underlying clinical need equally across groups [29]. Fairness therefore requires explicit choices concerning targets, populations, metrics, data, and consequences rather than confidence in overall performance [30]. In pharmacy AI, warranted reliance may differ across patient groups when data quality, access, calibration, or downstream resources differ.

**Figure 2** presents the original conceptual synthesis for refined programme theory of trust, reliance, and refusal in pharmacy ai, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

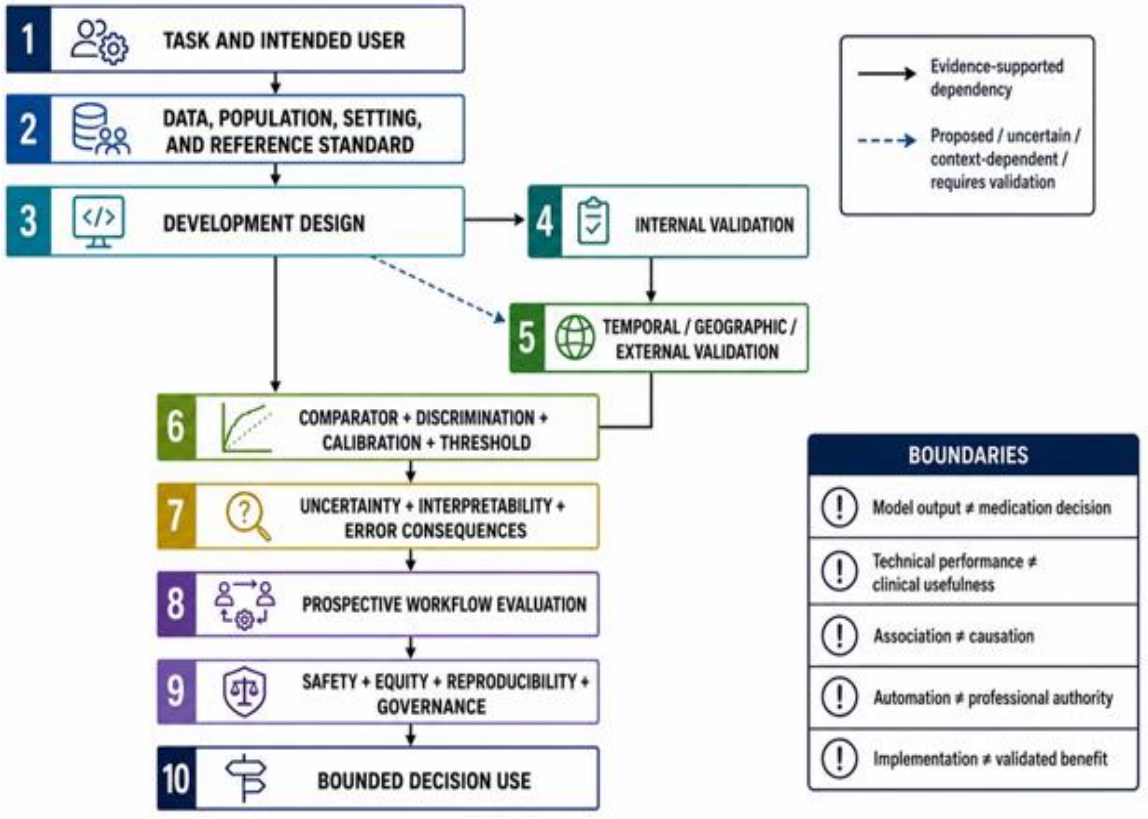


**Figure 2.** Refined Programme Theory of Trust, Reliance, and Refusal in Pharmacy AI

### *Implications for System Design and Implementation*

Systems should be designed to support checking, override, escalation, and documentation rather than to maximize confidence. Explanations should identify evidence, uncertainty, limitations, and the relationship between output and action. Interfaces should avoid unjustified authority cues. Workflow evaluation should examine whether AI redistributes attention, creates hidden work, changes responsibility, or generates informal bypasses.

Implementation should distinguish appropriate non-use from implementation failure. A pharmacist who rejects an unsupported recommendation may be performing the intended safety function. Monitoring should therefore include both erroneous acceptance and erroneous rejection, as well as delayed action, repeated dismissal, escalation, and workarounds.

### *Implications, Strengths, and Limitations*

#### *Implications for Measurement*

Explainability should be evaluated according to the audience, task, decision, and accountability need rather than as a generic trust intervention [33]. Measurement should separate attitude, intention, access, observed consultation, decision influence, verification, override, escalation, confidence, correctness, workflow consequence, and patient outcome.

Governance measurement should address accountability, transparency, privacy, bias, professional responsibility, monitoring, and opportunities for contestation [34]. A minimum evaluation set should identify the AI output, user, task, uncertainty, decision change, correctness of advice, checking behaviour, override or refusal, reason for non-use, patient subgroup, and downstream consequence.

### *Strengths, Limitations, and Theory Boundaries*

The review's main strength is its separation of outcomes frequently collapsed into trust. It integrates pharmacy evidence with mechanism-relevant clinical, patient, implementation, explainability, equity, and human-AI studies while retaining explicit transferability boundaries. Contradictory cases were used to refine rather than conceal uncertainty.

The evidence base nevertheless remained heterogeneous and included relatively little longitudinal pharmacy research. Several mechanisms were inferred rather than directly measured. Cross-domain evidence may not transfer to medication decisions with different stakes, accountability, or workflow. The public-index search may not reproduce the coverage of licensed multidatabase searches. No quantitative synthesis was appropriate, and no pooled effect, prevalence estimate, or formal certainty rating was produced.

## Conclusion

Trust is not a single outcome in pharmacy AI. Acceptance may occur without use, reliance may be appropriate or unsafe, refusal may be protective, and non-use may represent either vigilance or disengagement. These outcomes emerge through different mechanisms operating within technical, clinical, professional, patient, organizational, equity, and governance contexts.

The refined programme theory replaces maximal trust with calibrated decision use. Evaluation should determine whether pharmacists verify, rely, override, escalate, refuse, or bypass AI outputs and whether those behaviours are proportionate to system evidence, uncertainty, task stakes, workflow, and accountability. Technical performance cannot substitute for evidence of clinical usefulness, medication safety, equity, or implementation benefit.

The available evidence supports an explanatory framework rather than a clinical recommendation. Prospective and longitudinal pharmacy studies are required to test the proposed configurations, identify movement among them, measure hidden workflow effects, and determine when reliance or refusal improves or worsens medication-related decisions. Until such evidence is available, trust should be treated as a differentiated, context-dependent process rather than a universal implementation target.

ACKNOWLEDGMENTS: None
CONFLICT OF INTEREST: None
FINANCIAL SUPPORT: None
ETHICS STATEMENT: None

## References

1. Starke G, Gille F, Termine A, Aquino YSJ, Chavarriaga R, Ferrario A, et al. Finding consensus on trust in AI in health care: Recommendations from a panel of international experts. J Med Internet Res. 2025;27. doi:10.2196/56306
2. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: Focus on clinicians. J Med Internet Res. 2020;22(6). doi:10.2196/15154
3. 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. doi:10.2196/69678
4. Gustafson KA, Rowe C, Gavaza P, Bernknopf AC, Nogid A, Hoffman A, et al. Pharmacists' perceptions of artificial intelligence: A national survey. J Am Pharm Assoc (2003). 2025;65(1):102306. doi:10.1016/j.japh.2024.102306
5. Booth A, Briscoe S, Wright JM. The "realist search": A systematic scoping review of current practice and reporting. Res Synth Methods. 2020;11(1):14-35. doi:10.1002/jrsm.1386
6. Hunter R, Gorely T, Beattie M, Harris K. Realist review. Int Rev Sport Exerc Psychol. 2022;15(1):242-65. doi:10.1080/1750984X.2021.1969674
7. 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. doi:10.1136/bmj.n71
8. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev. 2021;10(1):39. doi:10.1186/s13643-020-01542-z
9. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, Chan AW, et al. Reporting guidelines for clinical trial reports for interventions

involving artificial intelligence: The CONSORT-AI extension. Nat Med. 2020;26(9):1364-74. doi:10.1038/s41591-020-1034-x
10. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ, Ashrafian H, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. Nat Med. 2020;26(9):1351-63. doi:10.1038/s41591-020-1037-7
11. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28(5):924-33. doi:10.1038/s41591-022-01772-9
12. Sounderajah V, Ashrafian H, Rose S, Shah NH, Ghassemi M, Golub R, et al. A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. Nat Med. 2021;27(10):1663-5. doi:10.1038/s41591-021-01517-0
13. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385. doi:10.1136/bmj-2023-078378
14. Farah L, Davaze-Schneider J, Martin T, Nguyen P, Borget I, Martelli N. Are current clinical studies on artificial intelligence-based medical devices comprehensive enough to support a full health technology assessment? A systematic review. Artif Intell Med. 2023;140:102547. doi:10.1016/j.artmed.2023.102547
15. Ranchon F, Chanoine S, Lambert-Lacroix S, Bosson JL, Moreau-Gaudry A, Bedouch P. Development of artificial intelligence powered apps and tools for clinical pharmacy services: A systematic review. Int J Med Inform. 2023;172:104983. doi:10.1016/j.ijmedinf.2022.104983
16. Hatzimanolis J, Riley B, El-Den S, Aslani P, Zhou J, Chaar BB. Applications of artificial intelligence in current pharmacy practice: A scoping review. Res Social Adm Pharm. 2025;21(3):134-41. doi:10.1016/j.sapharm.2024.12.007
17. Graafsma J, Murphy RM, van de Garde EMW, Karapinar-Carkit F, Derijks HJ, Hoge RHL, et al. The use of artificial intelligence to optimize medication alerts generated by clinical decision support systems: A scoping review. J Am Med Inform Assoc. 2024;31(6):1411-22. doi:10.1093/jamia/ocae076
18. Dil-Nahlieli D, Ben-Yehuda A, Souroujon D, Hyam E, Shafran-Tikvah S. Validation of a novel Artificial Pharmacology Intelligence system for the management of patients with polypharmacy. Res Social Adm Pharm. 2024;20(7):633-9. doi:10.1016/j.sapharm.2024.04.003
19. Huisman M, Ranschaert E, Parker W, Mastrodicasa D, Koci M, Pinto de Santos D, et al. An international survey on AI in radiology in 1,041 radiologists and radiology residents part 1: Fear of replacement, knowledge, and attitude. Eur Radiol. 2021;31(9):7058-66. doi:10.1007/s00330-021-07781-5
20. Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD, DesRoches CM. Artificial intelligence and the future of primary care: Exploratory qualitative study of UK general practitioners' views. J Med Internet Res. 2019;21(3). doi:10.2196/12802
21. Gaube S, Suresh H, Raue M, Merritt A, Berkowitz SJ, Lermer E, et al. Do as AI say: Susceptibility in deployment of clinical decision-aids. NPJ Digit Med. 2021;4(1):31. doi:10.1038/s41746-021-00385-9
22. Kiani A, Uyumazturk B, Rajpurkar P, Wang A, Gao R, Jones E, et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. NPJ Digit Med. 2020;3(1):23. doi:10.1038/s41746-020-0232-8
23. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digit Med. 2021;4(1):140. doi:10.1038/s41746-021-00509-1
24. 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). doi:10.1016/S2589-7500(21)00132-1
25. Strohm L, Hehakaya C, Ranschaert ER, Boon WPC, Moors EHM. Implementation of artificial intelligence applications in radiology: Hindering and facilitating factors. Eur Radiol. 2020;30(10):5525-32. doi:10.1007/s00330-020-06946-y
26. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human-computer collaboration for skin cancer recognition. Nat Med. 2020;26(8):1229-34. doi:10.1038/s41591-020-0942-0
27. Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, et al. Large language model influence on diagnostic reasoning: A randomized clinical trial. JAMA Netw Open. 2024;7(10). doi:10.1001/jamanetworkopen.2024.40969
28. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform. 2021;113:103655. doi:10.1016/j.jbi.2020.103655
29. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53. doi:10.1126/science.aax2342
30. Khera R, Simon MA, Ross JS. Automation bias and assistive AI: Risk of harm from AI-driven clinical decision support. JAMA. 2023;330(23):2255-7. doi:10.1001/jama.2023.22557
31. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med. 2018;169(12):866-72. doi:10.7326/M18-1990
32. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11). doi:10.1016/S2589-7500(21)00208-9
33. Amann J, Blasimme A, Vayena E, Frey D, Madai VI; Precise4Q Consortium. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310. doi:10.1186/s12911-020-01332-6
34. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: A mapping review. Soc Sci Med. 2020;260:113172. doi:10.1016/j.socscimed.2020.113172