Randa M. Perkins, MD, MBA, FAMIA

- Chief Health Information Officer (CHIO), Dartmouth Health
- Associate Professor of Community & Family Medicine, Geisel School of Medicine, Dartmouth College
- Member, Health Information Technology Advisory Committee (HITAC), Office of the National Coordinator of Health IT (ONC-HIT)

Why Clinicians Must Lead

Professional Responsibility in the Age of Healthcare AI

Geisel Academy of Educators and Scholars
AI in Medicine Series:
What Every Clinician Needs to Know
November 12, 2026

Views expressed are my own.

Disclosure Statement

- Content, recommendations, and opinions are my own (RP).
  Views expressed are personal, not institutional policy.
- Generative AI was used to draft slides and organize speaker notes.
- Evidence and regulatory claims were checked against primary and official sources on November 5, 2026.
- Mistakes are my own.

“But they [computers] are useless. They can only give you answers.”

— **Pablo Picasso**, speaking about calculating machines, reported in The Paris Review, 1964.

OPENING CASE

The Summary Is “Functional.” Is the Care Safe?

AI-GENERATED ENCOUNTER SUMMARY · HYPOTHETICAL

68-year-old admitted with community-acquired pneumonia. Improving on IV antibiotics; afebrile 24 hours. Oxygen weaned to room air. Plan: transition to oral therapy and discharge in 24–48 hours.

Medications: see medication list.
Allergies: no known drug allergies.

Not captured: yesterday’s nursing note documents facial swelling and hives after the first antibiotic dose. The drug was switched; the reaction was never added to the allergy list.

You are the student on the team. Handoff is in ten minutes, and the night resident asks you to “just read the summary.”

What must happen before this summary influences care?

Hold your answer. We will return to this patient at the end.

Roughly - Four Kinds of Health AI

1

Perception and classification

Recognizes patterns in images, text, or signals.

Examples
Image finding detection, coding support, document classification

2

Prediction

Estimates the chance of a future event.

Examples
Deterioration risk, readmission risk, demand forecasting

3

Generation

Produces new text or images.

Examples
Visit summaries, draft messages, patient education

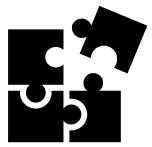
4

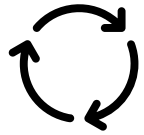
Agentic action

Uses tools and takes steps in other systems.

Examples
Calling tools, updating records, placing items in workflows, starting multistep tasks

Recognizes and estimates → Greater capacity to act on its own



Risk follows intended use, consequence, reach, autonomy, and recoverability, not the label "administrative."

AI Already Shapes the Patient Journey

1	2	3	4	5	6
Appointment access	**Symptom triage**	**Chart summary**	**Ambient visit note**	**Deterioration alert**	**Follow-up message**
Decision created	**Decision created**	**Decision created**	**Decision created**	**Decision created**	**Decision created**
Who is seen first	How urgent this is	What the team believes	What the record says	Who gets evaluated tonight	What the patient is told
Assists	*Recommends*	*Generates*	*Generates*	*Recommends*	*Generates / Acts*

Every touchpoint creates a clinical decision that someone must own.

Illustrative composite journey. Not every organization uses every tool shown; level of use depends on configuration.

Clinical Leadership Means Shaping the Decision

1 INTENDED USE

Which patients, which decision, and what it must never be used for

2 ACCEPTABLE ERROR

Which mistakes we can tolerate, at what rate, and who will absorb them

3 WORKFLOW

Who sees the output, when, and what they are expected to do with it

4 ESCALATION

What happens when the output is wrong, late, or ignored

5 OUTCOMES

What "working" means for patients, not just for the model

6 STOPPING

Who can pause it, on what evidence, and how fast

A physician signature does not replace safety architecture.

Clinical Expertise Is Necessary and Complimentary

CLINICAL TEAMS

Physicians, nurses, pharmacists, therapists, and others who deliver care

- Clinical context and what a wrong answer costs
- How the work actually happens at 3 a.m.
- Which errors are recoverable and which are not

PATIENTS AND FAMILIES

The people who live with the result

- Values and acceptable trade-offs
- Lived experience of access, language, and trust
- What an explanation needs to contain

TECHNICAL AND OPERATIONAL TEAMS

Informatics, data science, engineering, quality and safety, privacy and security, operations, legal

- Data limits, drift, and validation methods
- Human factors and interface design
- Monitoring, security, contracts, and incident response

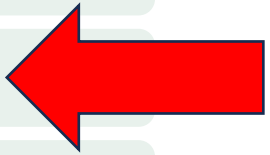
Shared work with clear decision authority. No one approves a model alone.

**NOT a comprehensive list! More on multidisciplinary review and governance later.*

A Correct Score Can Fail the Patient

- Patient population and access
- Source data
- **Model or rules**
- User interface
- Clinician interpretation
- Staffing and workflow
- Downstream response capacity
- Feedback and correction

Where most evaluations look



A useful score still fails *if no one can respond to it.*

Example:
A deterioration alert (e.g. sepsis alert) can be correct and still help no one if:

- it arrives after the team has acted,
- reaches the wrong clinical team,
- or flags a need the unit cannot meet that night.

Ask two questions of every alert: who receives it, and what can they actually do?

Strong Model Metrics Can Still Create Difficult Work

A Sepsis Model AUROC 0.82-0.92

- How well the score ranks patients who go on to develop sepsis above those who do not.
- (It says nothing about how many alerts you will chase.)

AT A 4-HOUR HORIZON

24–69

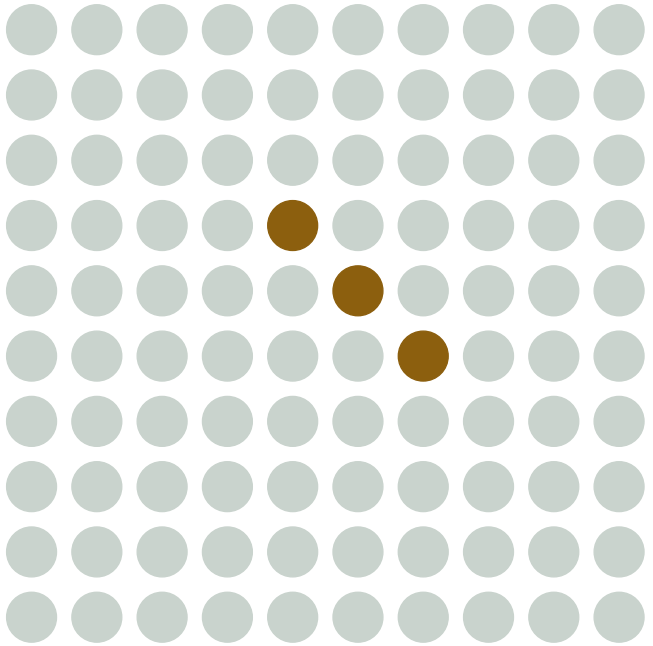
alerts evaluated per true case, by site (NNE)

PER ENCOUNTER

13–26%

positive predictive value at the 60% sensitivity threshold

Of every 100 alerts that fire 4 hours ahead:



About 2 to 4 are followed by sepsis in that window

All still need a clinician to evaluate

Choose your metrics wisely.

Source: Wong et al., JAMA Network Open 2026.

Scope & Scale Must Be Evidence-Based



7	Sustained monitoring after deployment	*Does it keep working as things change?*
6	Patient or clinical outcomes	*Are patients better off?*
5	Human and workflow testing	*Can real people use it under pressure?*
4	Silent prospective evaluation	*Does it work live, before anyone sees it?*
3	Local and subgroup validation	*Does it work here, for our patients?*
2	Technical performance	*Does it work anywhere?*
1	Clear intended use	*What is it for, and for whom?*

Accuracy is an early rung, not the finish line.

Deciding how much evidence is enough… is a clinical leadership decision.

Not a validated instrument. Consistent with the CHAI Responsible AI Guide and NIST AI RMF Measure/Manage functions.

WHAT THE EVIDENCE SHOWS

Ambient AI Makes the Leadership Question Visible

CLINICIAN RELIEF

52% → 39%

reporting burnout after 30 days (pre-post, no control). Two randomized trials: less work exhaustion and task load.

●●● *Most consistent*

TIME SAVED

~16 min

less documentation per 8 scheduled hours (about 10%). Half a visit more per week. After-hours EHR time unchanged.

●●○ *Modest; varies by use*

NOTE QUALITY

"Occasional"

clinically significant inaccuracies reported by physicians in a randomized trial; one mild adverse event.

●○○ *Depends on review*

PATIENT OUTCOMES

Not yet

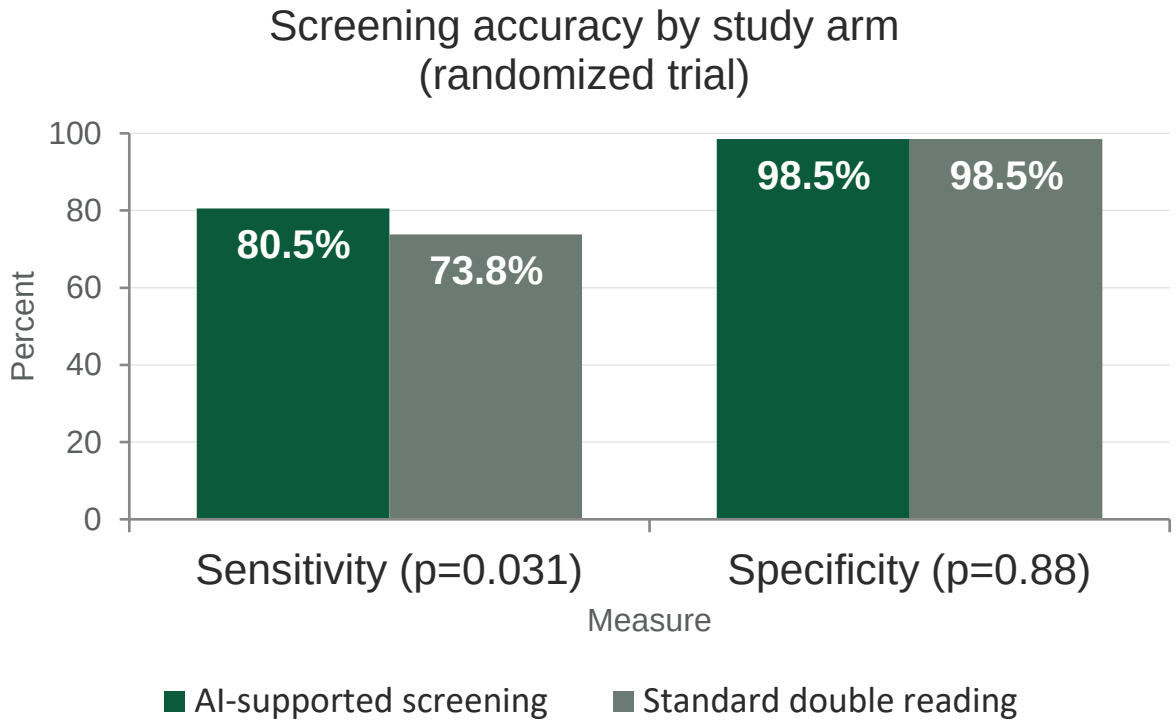
the endpoint of the randomized trials published so far. Patient experience and safety remain the least studied.

○○○ *Least studied*

Who decides how saved time is used? Who hears about the errors?

Sources: Olson et al., JAMA Netw Open 2025 (pre-post); Lukac et al., Afshar et al., NEJM AI 2025 (randomized); Rotenstein et al., JAMA 2026 (observational, n=8,581).

Clinical AI Can Earn a Place in Care. The MASAI Example.



1.55 vs 1.76

interval cancers per 1,000 women, AI-supported vs standard. Met the test for noninferiority. (p=0.41)

How AI was used

AI scores each mammogram

↓

Exam routed to one or two radiologists

↓

Radiologists decide on the interpretation

What earned trust can look like: a workflow trial, in the setting where the tool will be used.

Source: MASAI trial, Lancet, 2026 (randomized controlled trial)

Someone *Chooses* What the Model Should Predict

Example: A CLINICAL GOAL

Find the patients who most need extra care-management support.

→

THE MEASURED TARGET (THE LABEL)

Predict who will generate the highest health care costs next year.

IF THE LABEL WERE FIXED

17.7% → 46.5%

share of patients flagged for extra help who would be Black (simulated; Obermeyer et al., 2019)

Problem selection and label definition are clinical decisions. If clinicians are not in that room, the room will choose what is easy to measure.

Source: Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Science. 2019;366:447–453 (retrospective analysis of one commercial algorithm).

A Human Reviewer Needs a Workable Job

1 INFORMATION

The underlying data, not just the output

2 COMPETENCE

Relevant expertise for this decision

3 TIME

Enough of it to actually look

4 AUTHORITY

Permission to say no and be backed

5 ESCALATION

Somewhere to take a concern, fast

AUTOMATION BIAS, MEASURED

−11 points

in clinician diagnostic accuracy when shown a systematically biased model's predictions; explanations did not undo the harm (457 clinicians, vignette RCT).



AUTOMATION BIAS

Sources: Jabbour et al., JAMA 2023 (randomized vignette survey); Goh et al., JAMA Netw Open 2024 (randomized trial).
Image source: https://www.cloud-science.de/automation-bias/

Students Are Learning Judgment Alongside AI

WHAT IS AT RISK FOR A LEARNER

Premature closure

A fluent answer arrives before you have finished thinking, and the thinking stops.

Unchecked sources

Does the output tell you which note, lab, or guideline it leaned on, or whether it invented one?

Borrowed certainty

Confident prose hides the uncertainty you are supposed to be learning to feel.

A COGNITIVE EXERCISE

1 Commit first

Write your own assessment and plan before you open the AI output.

2 Then read critically

Check every claim against the chart. Mark what is missing, what is wrong, what is unsupported.

3 Discuss the discrepancies

Bring differences to your supervisor/leadership.

Protect the reasoning. The tool can only draft.

PATIENTS

What Builds Patient Trust

51%

say AI makes them trust health care less

12%

say AI increases their trust

93%

report at least one concern about health AI

80%+

say clear accountability would increase trust

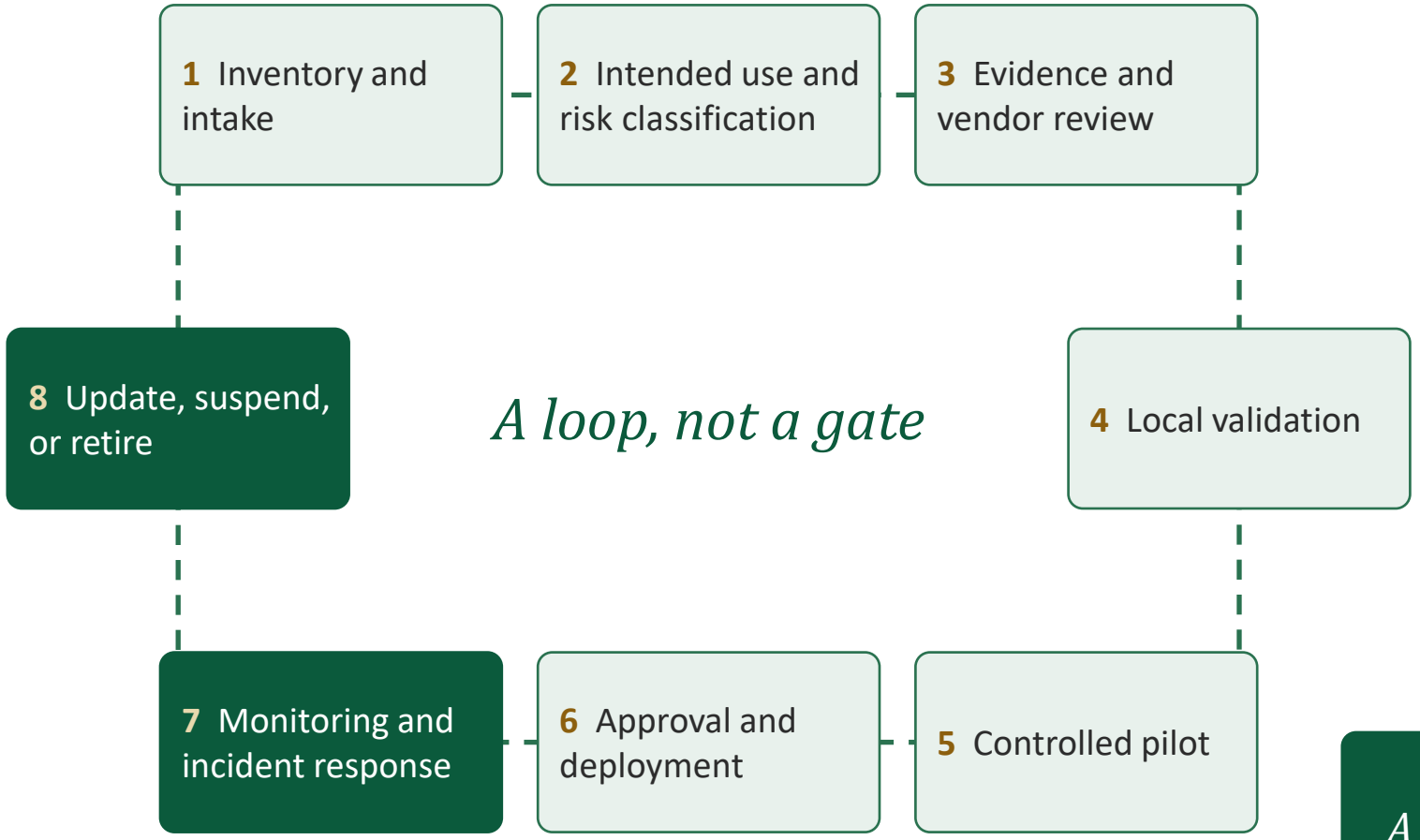
What disclosure should come with

- A clear purpose
- Human oversight
- Named accountability
- A way to question or appeal
- Privacy protection
- Evidence it helps people like me

Disclosure matters, but disclosure alone is not enough.

Source: Coalition for Health AI (CHAI)/NORC national patient survey, 2026 (probability-based AmeriSpeak panel). No claim is made about consent law or opt-out rights.

The Healthcare AI Lifecycle



Applies to *every* source of AI

- Purchased vendor products
- Internally developed models
- AI embedded in existing software
- Generative and agentic tools adopted informally by staff

Monitoring covers
Performance, safety events, overrides, disparities, workflow burden, feedback, and model or data changes

A free feature can create a very expensive clinical obligation.

Sources: CHAI Responsible AI Guide, CHAI governance playbooks, 2026

GOVERNANCE

Every Clinical Use Needs an Owner and a Response Plan

Role	Who, typically	Approves	Monitors	Escalates	Can suspend

The test: for every live tool, can you name the person who would get the call at 2 a.m., and does that person have the authority to turn it off?

A proposed model for discussion. Not an existing Dartmouth Health policy, structure, or deployment. Roles scale with organizational size and may be combined.

External Oversight Leaves Local Decisions

REGULATION (BINDING, WITHIN SCOPE)

FDA

Software meeting the medical-device definition. Some clinical decision support is excluded by statute; FDA's Jan 2026 guidance draws the line.

ONC certification

Certified health IT (e.g., EHRs), not every AI tool. HTI-5 (final Oct 2026) removed HTI-1's predictive decision support "model card" and risk-management requirements.

State and general law

Privacy, consumer protection, professional practice, civil rights. Varies by jurisdiction.

VOLUNTARY FRAMEWORKS

NIST AI RMF

Cross-sector risk vocabulary; voluntary; version 1.0 (2023) under revision.

Coalition for Health AI (CHAI)

Consensus guidance and governance playbooks (2026). Not a regulator.

Joint Commission RUAIH

Certifies an organization's AI governance (launched June 2026). Does not certify products.

LOCAL ACCOUNTABILITY (ALWAYS)

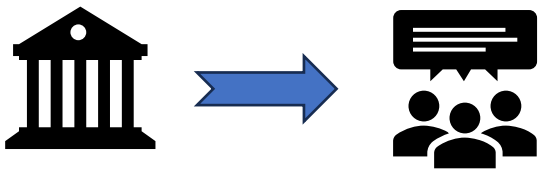
Your health system

Decides intended use, validates locally, designs the workflow, monitors, and stops. No external body does this for you.

Your clinical team

Notices when the tool and the patient disagree, and says so through a channel that works.

As federal transparency requirements narrow, more of the evaluation burden sits here.



Whether a tool helps your patients, in your workflow, is a local finding.

Orientation only; not legal advice. Status checked Nov 5, 2026. HTI-5 provisions take effect on staggered dates from Dec 2026.

Acting on an Answer Changes the Risk

A wrong answer can become a wrong action.

Refill example (hypothetical)

Generative
Produces a draft

Reads chart → Drafts refill request → **Clinician reviews and signs**

Agentic
Takes action

Reads chart → Checks pharmacy data → Submits refill → Updates medication list



CONTROLS WHEN A SYSTEM CAN ACT

Strict permissions read vs write, per task

Its own identity not a borrowed login

Independent check before consequential actions

Audit trail provenance, tamper-evident

Escalation and safe failure when unsure

Tested recovery stop and roll back, rehearsed

Illustrative scenario; autonomous refills are not presented as routine practice.
Controls: presenter synthesis informed by NIST NCCoE agent-identity concept work (2026).

YOUR DECISION

What Would You Require Before a Pilot?

Proposal: the same AI encounter summary from our opening case would now auto-populate the discharge handoff document sent to the primary care clinic. The vendor reports high clinician satisfaction and a 95% “accuracy” score on an internal review.

A Proceed as proposed

The feature is helpful; clinicians will review before sending.

B Redesign and test first

Define what must never be omitted, test it here, then pilot with measurement.

C Defer

The evidence is not there yet for a handoff that leaves the building.

0:20 Choose.

0:40 Explain in the chat.

1:00 Add: what evidence or safeguard would change your answer?

“What you permit is what you promote.”

Hypothetical scenario for teaching. The “95% accuracy” figure is invented to illustrate an unexamined vendor claim.

YOUR DECISION

A Pilot Needs an Endpoint and a Stopping Rule

Element	Illustrative pilot (hypothetical)	Why it matters

Teaching example. Numerical thresholds are illustrative and would be set locally with quality and safety. Not a Dartmouth Health protocol.

Leadership Can Begin During Training

STUDENTS

Under supervision

- Ask what a tool is for and what it was tested on
- Check AI output against the chart before using it
- Notice and report failures through the proper channel
- Join a supervised QI or informatics project

RESIDENTS

Within the team

- Verify before you sign; teach students to do the same
- Join supervised improvement work on a live tool
- Bring near-misses to safety reporting, not just to the hallway

PRACTICING PHYSICIANS

In the workflow

- Help define intended use and acceptable error
- Co-design and evaluate the workflow, not just approve it
- Serve on review groups; insist on local data

LEADERS

With authority and budget

- Ongoing monitoring must be part of the purchase price
- Assign named owners and decision rights
- Protect time; back the person who says stop

Nobody needs to become a programmer or a committee chair. Everybody needs to be able to say "show me."

Students should not independently approve systems, enter patient data into unapproved tools, change workflows, or act outside supervision.

Clinical Leadership Needs Institutional Support

1

PROTECTED TIME

Review and monitoring are work, not volunteering

2

EDUCATION

Evidence appraisal, human factors, data limits, implementation

3

REPORTING

A route for AI-related concerns that someone actually reads

4

DATA ACCESS

Local performance, subgroup results, override and error rates

5

POWER TO PAUSE

A named person can stop an unsafe use without a committee cycle



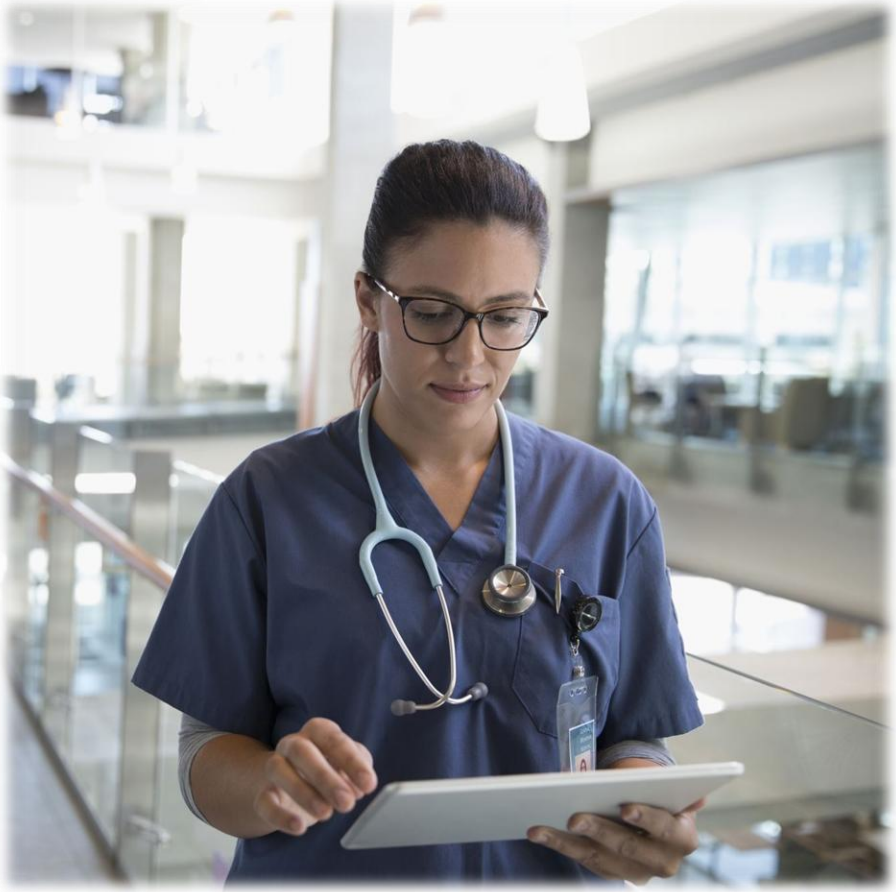
TWO PATHWAYS, BOTH REAL

1. **Clinical informatics** is a board-certified subspecialty and a career.
2. **Quality and safety leadership** is another other path: the clinician on the unit who insists on local data and a working feedback loop.

Responsibility without resources = liability transfer with a nicer name.

Four Questions Clinicians Should Bring

1. What problem are we solving, and for whom?
2. What evidence shows benefit here, in this setting?
3. Who owns the decision, and who catches failure?
4. How will we monitor, correct, and stop it?



CLOSING

Caring for the Patient Remains the Purpose

BACK TO OUR PATIENT: A PROPOSED SAFE RESPONSE



1 **Check the record**

Read the nursing notes and the medication administration record yourself. Find the reaction.

2 **Bring it to the supervisor**

Before handoff: “The summary is missing a drug reaction. I think the allergy list needs updating before we hand off.”

3 **Report the system issue**

Through the safety reporting channel: the summary omitted a documented reaction. If it happened once, it will happen again.

“…the secret of the care of the patient is in caring for the patient.”

Francis W. Peabody, MD. The Care of the Patient. JAMA, 1927.

Peabody FW. JAMA. 1927;88(12):877–882.

DISCUSSION

Questions and Discussion

1

Where in your training have you already been asked to trust a system you could not evaluate?

2

What would make it possible for you to raise a concern about an AI tool, and to whom?

3

You have ten minutes with the people buying an AI tool for your clinic. What do you ask first?

4

What would it take to give a resident the authority to pause use of a tool on their unit?

APPENDIX A1

Regulatory Scope and Status

Status checked November 5, 2026. Labels only; not legal advice. Full citations in notes.

Item	Type	Status (Nov 5, 2026)	Scope and limits
FDA device oversight	Statute and regulations	In effect	Software meeting the medical-device definition. Certain clinical decision support functions are excluded by statute (21st Century Cures Act).
FDA Clinical Decision Support Software guidance	Final guidance (nonbinding)	Revised final issued Jan 6, 2026; reissued Jan 29, 2026	Explains which CDS functions FDA considers non-device, including conditions under which LLM-enabled functions may qualify.
FDA PCCP guidance for AI-enabled device software	Final guidance (nonbinding)	Final Dec 2024; reissued Aug 2025	Lets planned modifications to authorized AI devices be pre-specified. Devices only.
ONC HTI-1	Final rule	Published Jan 2024	Certified health IT. Added the decision support interventions (DSI) criterion, incl. predictive DSI transparency (“source attributes”).
ONC HTI-5	Final rule	Released Oct 26, 2026; Fed. Reg. Oct 28, 2026; most provisions effective Dec 28, 2026	Removes 14 of 60 certification criteria, revises 7. Removes source-attribute (“model card”) and intervention risk-management requirements for predictive DSI. Information-blocking proposals not finalized.
HITAC	Federal advisory committee	Active	Recommendations to ONC. No regulatory authority.
NIST AI RMF 1.0	Voluntary framework	Released Jan 2023; revision directed July 2025, in progress	Cross-sector risk management. Not a health care regulation. Reconfirm revision status at nist.gov.
Joint Commission RUAIH certification	Voluntary organizational certification	Launched June 1, 2026	Certifies an organization’s AI governance. Does not evaluate or certify AI products.
CHAI guidance and playbooks	Consensus guidance	Playbooks released May 27, 2026	Voluntary; aligned with Joint Commission certification. Not a regulator.
State and general law	Statutes and regulations	Varies	Privacy, consumer protection, professional practice, insurance, civil rights, employment.

APPENDIX A2

Governance Framework Crosswalk

Dimension	CHAI Responsible AI Guide	CHAI governance playbooks (2026)	NIST AI RMF 1.0	Joint Commission RUAIH certification
Structure	Five themes: usefulness/usability/efficacy; fairness; safety and reliability; transparency, intelligibility, accountability; security and privacy	Eight domains incl. AI policy, organizational structures, resources, lifecycle management, risk and impact assessment, data management, third-party management, education and feedback	Four functions: Govern, Map, Measure, Manage; seven trustworthiness characteristics	Standard areas incl. governance, data management, risk and bias reduction, monitoring and validation, transparency and education
Audience	Developers and deployers of health AI	Health delivery organizations of all sizes	Any organization, any sector	US hospitals and health systems
Who owns clinical decisions	Deployer accountability; clinical input expected	Organizational structures domain names roles	Govern function assigns accountability	Governance standard requires accountable leadership
Lifecycle monitoring	Across lifecycle stages	Lifecycle management domain	Measure and Manage	Monitoring and validation area
External certification	No; guidance	No; aligned with certification	No; voluntary framework	Yes; organization-level, voluntary, not product-level

Restrained crosswalk. Mappings are the presenter's summary, not official crosswalks by these organizations. None of these frameworks evaluates whether a specific tool helps a specific patient population; that remains a local finding.

Sources: CHAI RAIG; CHAI governance playbooks (May 2026); NIST AI RMF 1.0 (Jan 2023); Joint Commission RUAIH (June 2026). Not an officially validated instrument.

APPENDIX A3

Clinical Evidence: Details and Limitations

Study	Design and population	Key findings (as reported)	Limitations to state when teaching
Wong et al., JAMA Netw Open 2026 (Epic Sepsis Model v2)	Prospective multicenter validation; 227,091 adult inpatient encounters; 4 US systems; Aug 2023–Mar 2025	Encounter-level AUROC 0.82–0.92. At 60%-sensitivity threshold: PPV 0.13–0.26. 4-h horizon: PPV 0.02–0.04, NNE 24–69 by site; 12-h NNE 21–35; hospitalization NNE 4–8	Automated Sepsis-3 labels; model live during data collection; all sites fine-tuned; not designed for outcomes
Gommers et al., Lancet 2026 (MASAI)	Randomized noninferiority screening-accuracy trial; ~105,900 women; Swedish double-reading program	Interval cancers 1.55 vs 1.76/1,000 (ratio 0.88; noninferior at 20% margin; between-group P=.41). Sensitivity 80.5% vs 73.8% (P=.031). Specificity 98.5% both	One country, one AI, double-reading model; noninferior ≠ superior; later rounds and cost-effectiveness pending
Lukac et al.; Afshar et al., NEJM AI 2025 (ambient scribes)	Randomized trials: 238 physicians (3-arm, UCLA); 66 practitioners (stepped-wedge, UW Health)	Lower work exhaustion and task load; one tool reduced time-in-note; “occasional” clinically significant inaccuracies; one mild adverse event	Short follow-up; products change; clinician-reported measures; patient outcomes not primary endpoints
Rotenstein et al., JAMA 2026 (ambient scribes)	Multisite longitudinal cohort, EHR audit logs; 8,581 clinicians (1,809 adopters), 5 academic systems	−16.0 min documentation and −13.4 min EHR time per 8 scheduled h; +0.49 visits/wk; after-hours time unchanged	Observational association; opt-in adoption; academic centers; documentation time ≠ total work
Obermeyer et al., Science 2019	Retrospective analysis of one commercial population-health algorithm	Cost label underestimated Black patients’ need; correcting label raised share of Black patients flagged from 17.7% to 46.5% (simulated)	One algorithm, one proxy; simulation not deployment
Jabbour et al., JAMA 2023; Goh et al., JAMA Netw Open 2024	Randomized vignette survey (457 clinicians); RCT (50 physicians)	Biased model −11.3 points accuracy, explanations did not rescue; LLM access +2 points n.s. while LLM alone +16	Constructed cases, not live care; task-dependent

Full citations with DOIs in Appendix A6. All accessed November 5, 2026.

APPENDIX A4

Sample AI Pilot Evaluation Checklist (Teaching Example)

BEFORE THE PILOT

- ☐ Intended use, population, and exclusions written down
- ☐ Clinical goal vs. measured target reconciled
- ☐ Local validation and subgroup results reviewed
- ☐ Workflow mapped: who sees output, when, what they do
- ☐ Primary safety measure and comparator named
- ☐ Stop and escalation rules prespecified
- ☐ Named clinical, technical, and operational owners
- ☐ Patient explanation and feedback route drafted

DURING THE PILOT

- ☐ Safety events and near-misses reviewed weekly
- ☐ Review time and substantive-edit rate tracked
- ☐ Override and alert-burden rates tracked
- ☐ Subgroup performance checked, not just averages
- ☐ User and patient feedback collected and read
- ☐ Any model or data change logged and assessed

BEFORE EXPANSION

- ☐ Prespecified endpoints met, and by how much
- ☐ Unexpected harms or burdens described honestly
- ☐ Equity findings addressed, not footnoted
- ☐ Monitoring plan and owners for steady state
- ☐ Decision recorded: expand, modify, or stop, and why

A teaching checklist, not a validated instrument or an institutional form. Adapt locally with quality and safety.

Presenter synthesis; consistent with the pilot and monitoring stages described in the CHAI Responsible AI Guide and governance playbooks.

APPENDIX A5

Ways Students and Trainees Can Participate, Under Supervision

Supervised quality improvement

Join a QI project that evaluates a live tool: chart review for omissions, override patterns, or alert burden. Most programs require QI participation anyway; choose one that matters.

Informatics electives and rotations

Many schools and health systems offer clinical informatics electives; ask your clerkship director. Clinical informatics is also a board-certified subspecialty for physicians, with parallel paths in nursing and pharmacy.

Governance committees as observers

Ask whether your institution's AI or clinical decision support committee accepts student or resident observers. Watching a real intake decision is the fastest education available.

Safety reporting

Use the institution's safety reporting system for AI-related near-misses and errors, with your supervisor's knowledge. A pattern becomes visible only when individual reports exist.

Patient and family advisory councils

Attend or support a council meeting where an AI tool is discussed. Hear how patients describe trust, explanation, and recourse in their own words.

Scholarship

Evaluation studies, usability studies, and implementation case reports are publishable and needed. Partner with informatics, quality, or health services research faculty.

Boundaries: do not enter patient data into tools your institution has not approved; do not change clinical workflows independently; do not approve or disable systems; act within supervision and your institution's policies.

General pathways; availability varies by institution and program. Confirm local options with your clerkship or program director.

References

Clinical evidence

Wong A, Currey D, Schwinne M, et al. Multicenter Prospective Validation of an Updated Proprietary Sepsis Prediction Model. JAMA Netw Open. 2026;9(2):e260181. doi:10.1001/jamanetworkopen.2026.0181. Prospective validation.

Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Intern Med. 2021;181(8):1065–1070. doi:10.1001/jamainternmed.2021.2626.

Gommers J, Hernström V, Josefsson V, et al. Interval cancer, sensitivity, and specificity comparing AI-supported mammography screening with standard double reading without AI in the MASAI study. Lancet. 2026;407(10527):505–514. doi:10.1016/S0140-6736(25)02464-X. Randomized noninferiority trial.

Lång K, Josefsson V, Larsson A-M, et al. AI-supported screen reading versus standard double reading in MASAI: a clinical safety analysis. Lancet Oncol. 2023;24(8):936–944. doi:10.1016/S1470-2045(23)00298-X.

Lukac PJ, Turner W, Vangala S, et al. Ambient AI Scribes in Clinical Practice: A Randomized Trial. NEJM AI. 2025;2(12). doi:10.1056/AIoa2501000.

Afshar M, Baumann MR, Resnik F, et al. A Pragmatic Randomized Controlled Trial of Ambient AI to Improve Health Practitioner Well-Being. NEJM AI. 2025;2(12). doi:10.1056/AIoa2500945.

Olson KD, Meeker D, Troup M, et al. Use of Ambient AI Scribes to Reduce Administrative Burden and Professional Burnout. JAMA Netw Open. 2025;8(10):e2534976. doi:10.1001/jamanetworkopen.2025.34976. Pre-post QI study.

Rotenstein LS, Holmgren AJ, Thombley R, et al. Changes in Clinician Time Expenditure and Visit Quantity With Adoption of AI-Powered Scribes. JAMA. 2026. doi:10.1001/jama.2026.2253. Observational cohort.

Human–AI interaction and equity

Jabbour S, Fouhey D, Shepard S, et al. Measuring the Impact of AI in the Diagnosis of Hospitalized Patients: A Randomized Clinical Vignette Survey Study. JAMA. 2023;330(23):2275–2284. doi:10.1001/jama.2023.22295.

Goh E, Gallo R, Hom J, et al. Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial. JAMA Netw Open. 2024;7(10):e2440969. doi:10.1001/jamanetworkopen.2024.40969.

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):447–453. doi:10.1126/science.aax2342.

Coalition for Health AI; NORC at the University of Chicago. Patient Survey Report on Health AI and Transparency. Jan 28, 2026. chai.org/news.

Governance and regulation

ONC/HHS. HTI-5 Final Rule. Released Oct 26, 2026; Federal Register Oct 28, 2026. healthit.gov/resources/…hti-5-final-rule/

ONC. HTI-1 Final Rule. Jan 2024. healthit.gov/regulations/hti-rules/hti-1-final-rule/

FDA. Clinical Decision Support Software: Guidance. Jan 6, 2026; reissued Jan 29, 2026. FDA-2017-D-6569.

FDA. Marketing Submission Recommendations for a PCCP for AI-Enabled Device Software Functions. Final Dec 2024; reissued Aug 2025.

NIST. AI Risk Management Framework 1.0. Jan 2023 (revision in progress). nist.gov/itl/ai-risk-management-framework

NIST NCCoE. Software and AI Agent Identity and Authorization. Draft concept paper, Feb 2026.

Coalition for Health AI. Responsible AI Guide; Governance Playbooks (May 27, 2026). chai.org

Joint Commission. Responsible Use of AI in Healthcare Certification. June 1, 2026. jointcommission.org

Peabody FW. The Care of the Patient. JAMA. 1927;88(12):877–882. Reprinted JAMA. 1984;252(6):813–818.

All web sources accessed November 5, 2026.