NURSING AND AI INNOVATION CONSORTIUM (NAIIC)

Governance and Legal Considerations for Responsible AI Implementation

Rebecca Love
Jing Wang Brenton Hill

NURSIFYING AI

NAIIC Membership:

NURSING.FSU.EDU/NAIIC

Responsible AI in Nursing Micro-Credential:

CHAI

NURSING.FSU.EDU/RESPONSILBEAI

The Governance Gap

Rapid Adoption

AI is transforming clinical decision, documentation, workflow redesign, and patient monitoring - faster than governance structures can keep pace.

Alowais et al., 2023

Policy Momentum

The White House National AI Legislative Framework (Mar 2026) and HHS AI Strategy both call for creative innovation with appropriate guardrails and public trust.

White House, 2026; HHS, 2026

Governance Lag

Rapid AI adoption has outpaced organizational structures needed for safe, ethical, and sustainable implementation.

Kim et al., 2025

WHO Implementation: Nursing Governance in Practice

WHO's Implementation Tools for Governance

- **Ethical Impact Assessments (EIAs)** — conducted before deployment and at regular intervals; include bias and privacy risk review
- **Human Rights Impact Assessments (HRIAs)** — identify potential violations for marginalized patient groups
- **Public Engagement** — communities most affected by AI (especially patients) must participate in design and governance
- **Developer & Ministerial Checklists** — structured tools for Ministries of Health to verify ethical compliance

Nursing Leadership Imperatives from WHO

WHO requires all stakeholders — including nurses — to integrate ethical norms at every stage of technology's development, deployment, and use.

- Champion AI oversight committees ensuring nursing is represented at governance tables
- Require EIAs that include equity audits for race, gender, age disparities in AI outputs
- Ensure patients are engaged in AI design and informed when AI directly impacts their care
- Embed WHO's 6 principles into AI literacy curricula and nursing orientation

Source: WHO Ethics & Governance of AI for Health, 2021 · WHO Guidance on Large Multimodal Models, 2024

NIST AI Risk Management Framework (AI RMF 1.0)

Voluntary, lifecycle-oriented framework from National Institute of Standards and Technology (NIST) Defines AI as socio-technical — risk emerges from how people build, deploy, and use systems.

GOVERN	Cross-cutting culture of AI risk management. Establishes accountability, defines policies, ensures oversight across the full AI lifecycle. GOVERN 1.1: understand legal/regulatory requirements. Applies at all lifecycle stages.
MAP	Scoping function. Establishes context, identifies stakeholders and categories of potential harm across the AI system lifecycle. MAP 5.1: document likelihood and magnitude of impacts. Feeds MEASURE and MANAGE.
MEASURE	Choose metrics and evaluate trustworthiness characteristics over time. MEASURE 2.11: evaluate fairness and bias. Tracks 7 characteristics: valid/reliable, safe, secure, accountable, explainable, privacy-enhanced, bias-managed.
MANAGE	Prioritize risks, define response treatments (mitigate, transfer, avoid, accept). MANAGE 4.1: post-deployment monitoring, appeal and override mechanisms, decommissioning, and change management.

Source: NIST AI 100-1 — AI Risk Management Framework (RMF 1.0), January 2023

CHAI & The Joint Commission RUAIH (2025)

Responsible Use of AI in Healthcare (RUAIH)

- **AI Policy & Governance Structures** — formal oversight with board-level accountability
- **Patient Privacy & Transparency** — disclose AI use; notify patients when AI directly impacts care
- **Data Security & Data Use Protections** — HIPAA compliance; encryption; BAAs with vendors
- **Ongoing Quality Monitoring** — risk-based, continuous performance validation post-deployment
- **Voluntary, Blinded Safety Event Reporting** — confidential reporting to PSOs; link to sentinel event process
- **Risk & Bias Assessment** — CHAI Applied Model Card; pre- and post-deployment audits
- **Education & Training** — AI literacy, role-specific training, and change management

Joint Commission Context

Released September 17, 2025 - the first formal framework from a U.S. accrediting body for safe, ethical AI integration. Currently voluntary; will inform future accreditation and certification pathways.

- Governance structure must include nurses, physicians, C-suite, compliance, IT, and cybersecurity
- Fiduciary board receives regular updates on AI outcomes and adverse events
- Tools that inform clinical decisions require more frequent monitoring than administrative AI
- Survey readiness: document AI oversight, validation, and training protocols now

Source: Joint Commission & CHAI — Guidance on Responsible Use of AI in Healthcare (RUAIH), September 2025

Nursing-Centered AI Governance Framework :

WHO establishes why. NIST defines how. CHAI/JC operationalizes where. Nursing leadership decides when and governs throughout.

WHO

Ethics & values foundation

- 6 ethical principles (autonomy, wellbeing, transparency, accountability, equity, sustainability)
- EIAs + HRIAs as required governance tools
- 2024: LMM governance across design, provision, liability

NIST

Risk management structure

- GOVERN – MAP – MEASURE – MANAGE across lifecycle
- 7 trustworthiness characteristics with explicit trade-offs
- Current/Target profiles for gap analysis and audit readiness

CHAI / JC

Operational & accreditation

- 7 RUAIH elements; voluntary but accreditation-shaping
- CHAI Applied Model Card for bias documentation
- Sentinel event integration for AI adverse event reporting

Nursing's Role

Lead governance - not just implement

- Serve on AI governance committees across all frameworks
- Drive EIAs, bias audits, and patient transparency requirements
- Align AI safety reporting with sentinel event processes

Nursing leaders must not wait to be governed — they must lead governance. Claim decisional authority at every stage of the AI lifecycle

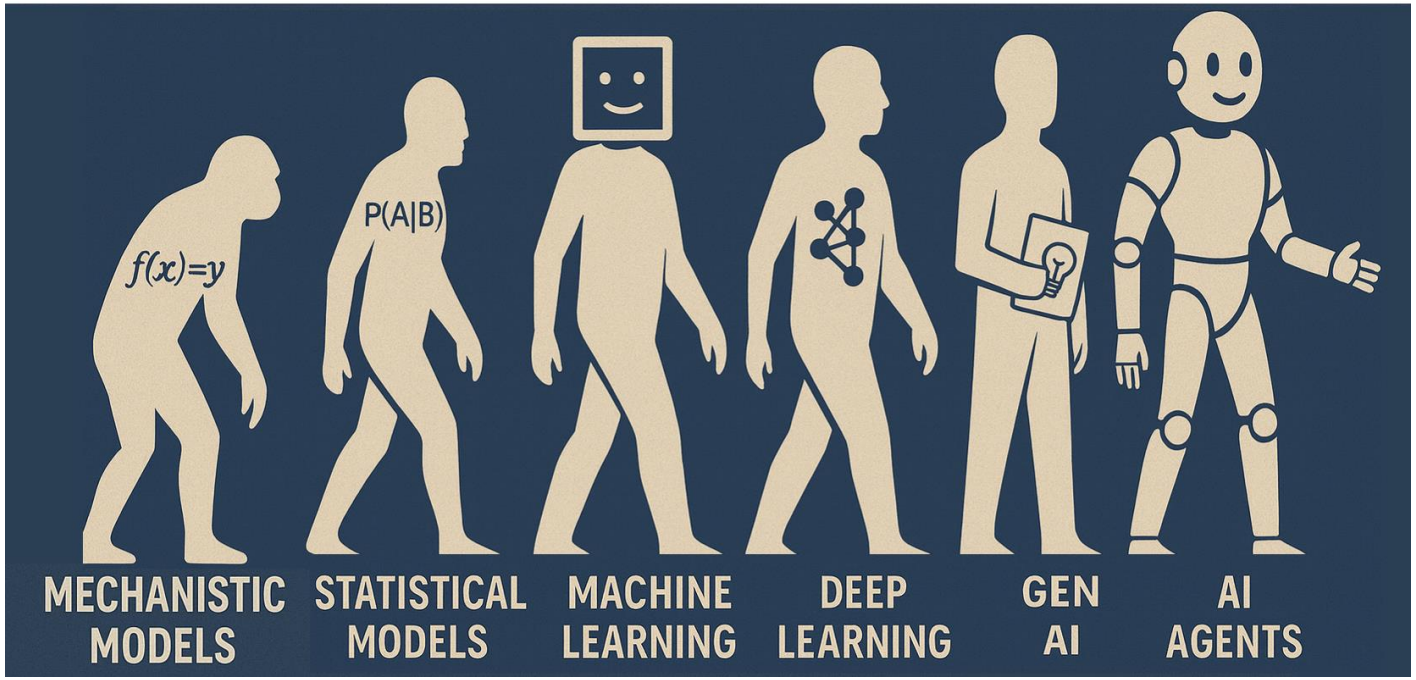
Source: WHO 2021/2024 · NIST AI 100-1, 2023 · Joint Commission & CHAI RUAIH, 2025 · Cary et al., 2026

Thank you nurses!!

MountainStar
Ogden Regional Medical Center

onvida Health™

AI 101: What You Need to Know



Less Data
Single Systems

More Data
More Systems

- Drug Prediction
- Hospital Readmission
- Sepsis Risk Prediction
- Radiology tool that flags tumors
- Ambient Scribe
- Automatic decisions and insights

How clinical information sourcing has evolved

Historical Clinical Evidence Sourcing

A clinical question

Traditional sources
Books, journals, guidelines

Internet search
Google, PubMed, websites

AI

TODAY: Clinical Evidence Generated Instantly

General-purpose LLMs

ChatGPT · Claude · Gemini

Ask in natural language. Broad, fast, and fluent across topics.

Needs source grounding — citations not guaranteed

OR

Clinical AI point solutions

OpenEvidence · Doximity · UpToDate

LLM connected to trusted clinical content, retrieval, and workflows.

Answers come with citations & clinical workflow fit

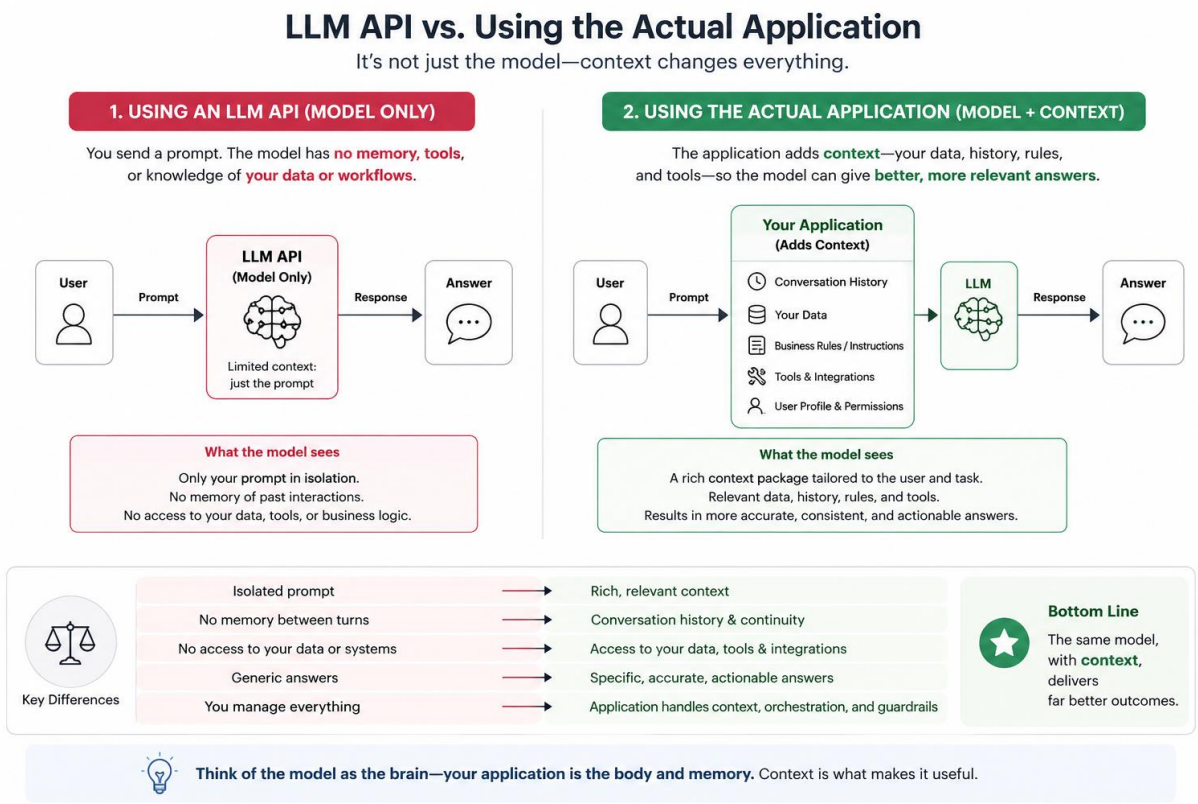
Clinical AI point solution = LLM + trusted clinical content + retrieval + citations + workflow design

The LLM is the communicator; the clinical knowledge base is the library; citations let clinicians verify the answer.

How a clinical AI point solution reaches an answer:

Clinical question → **Retrieve evidence** → **LLM summarizes** → **Cited answer**

Context, Context, Context…..



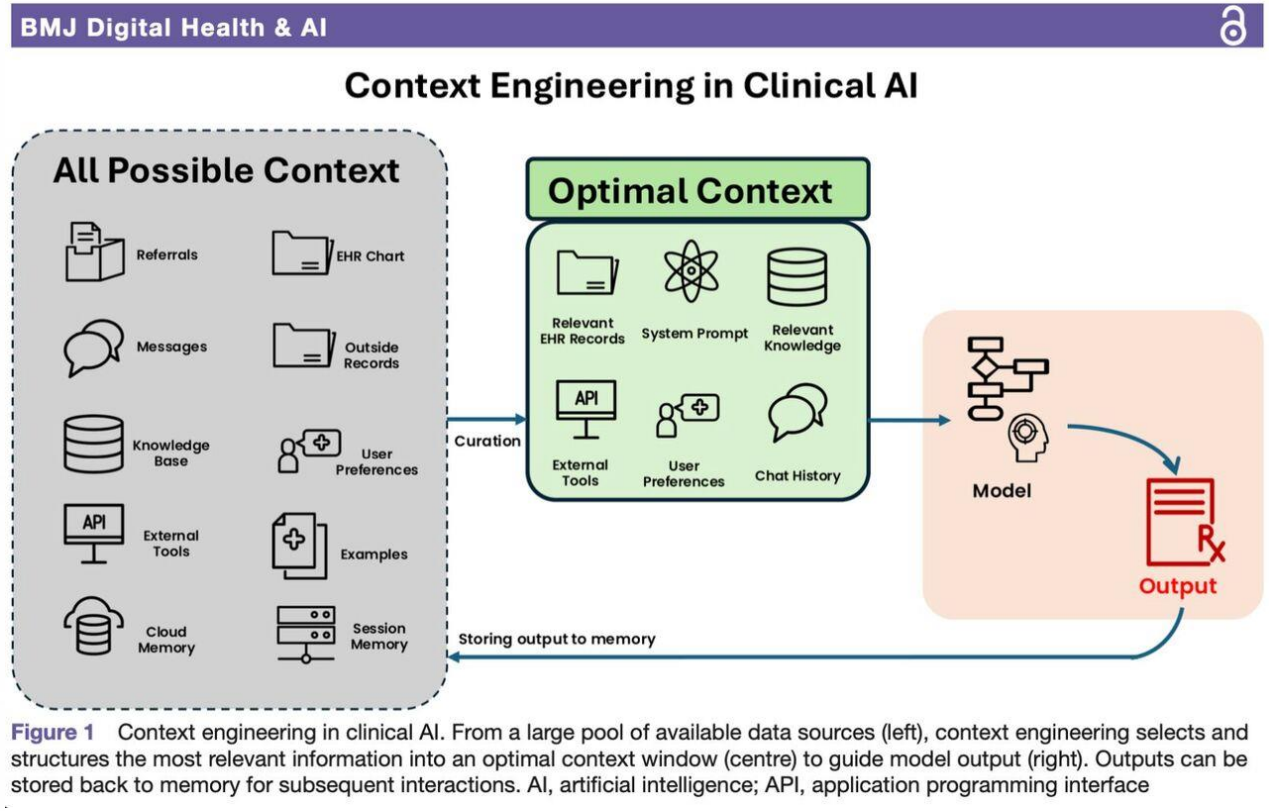


Figure 1 Context engineering in clinical AI. From a large pool of available data sources (left), context engineering selects and structures the most relevant information into an optimal context window (centre) to guide model output (right). Outputs can be stored back to memory for subsequent interactions. AI, artificial intelligence; API, application programming interface

"Giving an AI system 'more context' is not always better. It should receive just enough information to support the task, but not so much that the signal gets buried."

– Ethan Goh, Stanford ARISE

AI Architecture in Healthcare: Data & Systems

Gen AI vs. Agentic AI — nursing use cases

GEN AI

Ambient Nursing Documentation · Dragon Copilot / Abridge

3 systems · linear flow · nurse signs off every time

DATA LAYER — WHAT FLOWS IN & OUT

Bedside assessment audio → Transcript (raw text) → *** Draft note + flowsheet rows** → Signed note in EHR

SYSTEMS LAYER

MIC — **Recording Device / App**
Microphone captures nurse–patient bedside encounter
Nurse activates

↓

LLM — **LLM / Transcription Engine**
Converts audio → draft narrative note + structured flowsheet entries
AI generates

↓

EHR — **EHR (Epic / Cerner)**
Draft surfaces for nurse review, edit & validation
Nurse must validate

The key legal distinction:

Every output passes through a licensed nurse before it enters the medical record. The RN reviews, edits, and validates within their scope of practice.

VS

AGENTIC AI

Early Warning — Sepsis Detection · Epic Sepsis Model / Rothman

7 systems · parallel · autonomous actions · continuous 24/7

DATA LAYER — WHAT FLOWS IN & OUT

Vitals + labs (live) → Med list + allergies → *** Agent scores sepsis risk** → RN alerted + bundle initiated

SYSTEMS LAYER — AGENT CONNECTS TO ALL SIMULTANEOUSLY

AI Agent — Continuous Monitor
Runs autonomously 24/7 · reads & writes across all systems below
Always on

PULLS DATA FROM

EHR	**EHR (Epic)** — History, nursing assessments, current orders	Auto-read
LAB	**Lab Information System** — Live CBC, lactate, blood culture flags	Auto-read
MON	**Vitals / Monitoring System** — Nurse-charted HR, BP, temp, SpO2 — every 5 min	Auto-read
RX	**Pharmacy System** — Checks med list, allergies, dosing protocols	Auto-read

ACTS ON / WRITES TO

PG	**Nurse Alert System** — Notifies bedside RN + charge nurse simultaneously	Auto-fires
TSK	**Nurse-Driven Protocol (standing orders)** — Queues sepsis bundle tasks — lactate, cultures, fluids — for RN to initiate	AI drafts
DSH	**Care Management Dashboard** — Updates risk score & case status unit-wide	Auto-writes

MedAgentBench – testing tasks in a real EHR

“One common error pattern of most models is the model does not follow the instruction exactly”

300 physician-written tasks in a FHIR-compliant EHR · 100 patients · 785K records · Stanford ML Group, 2025.

THE TASK MIX

Two task families, ten clinical categories

QUERY · read from the EHR

- Lab results lookup
- Vital sign retrieval
- Procedure & diagnosis history
- Current medication list
- Risk-score computation from chart data

ACTION · write back to the EHR

- Place medication orders
- Order labs and imaging
- Schedule follow-up & referrals
- Update problem list / documentation

Table 3: **Success rate (SR) of state-of-the-art LLMs on MedAgentBench.** This table presents the performance of various state-of-the-art large language models (LLMs) on MedAgentBench, measured by overall success rate (SR), query SR, and action SR. The best-performing SR values in each column are highlighted in bold.

Model	Size	Form	Overall SR	Query SR	Action SR
Claude 3.5 Sonnet v2	N/A	API	**69.67%**	**85.33%**	54.00%
GPT-4o	N/A	API	64.00%	72.00%	56.00%
DeepSeek-V3	685B	open	62.67%	70.67%	54.67%
Gemini-1.5 Pro	N/A	API	62.00%	52.67%	**71.33%**
GPT-4o-mini	N/A	API	56.33%	59.33%	53.33%
o3-mini	N/A	API	51.67%	54.67%	48.67%
Qwen2.5	72B	open	51.33%	38.67%	64.00%
Llama 3.3	70B	open	46.33%	50.00%	42.67%
Gemini 2.0 Flash	N/A	API	38.33%	34.00%	42.67%
Gemma2	27B	open	19.33%	38.67%	0.00%
Gemini 2.0 Pro	N/A	API	18.00%	25.33%	10.67%
Mistral v0.3	7B	open	4.00%	8.00%	0.00%

The action gap: *Claude 3.5 Sonnet scores 85% on reading the chart, only 54% on acting on it. Models can find, yet acting on the information is not yet sophisticated.*

Source: arxiv.org/abs/2501.14654 · stanfordmlgroup.github.io/projects/medagentbench

CLINICAL AI BRIEFING

The Compliance Landscape for Nurses Using AI

Every AI-assisted task in the daily workflow maps to an existing legal duty — the tool is new, the obligation is not.

1	2	3	4	5	6
Triage & Intake	**Assessment & Monitoring**	**Clinical Decision Support**	**Documentation & EHR**	**Medication Administration**	**Patient Communication**
AI acuity / risk scoring	*Early-warning & sepsis alerts*	*AI care recommendations*	*Ambient scribe / auto-charting*	*Dose & interaction checks*	*Chatbots & translation*
NIST RISK ATTRIBUTE	NIST RISK ATTRIBUTE	NIST RISK ATTRIBUTE	NIST RISK ATTRIBUTE	NIST RISK ATTRIBUTE	NIST RISK ATTRIBUTE
Fair – Bias Managed	**Safe**	**Explainable & Accountable**	**Private & Secure**	**Valid & Reliable**	**Fair & Explainable**
maps to	*maps to*	*maps to*	*maps to*	*maps to*	*maps to*
LEGAL OBLIGATION	LEGAL OBLIGATION	LEGAL OBLIGATION	LEGAL OBLIGATION	LEGAL OBLIGATION	LEGAL OBLIGATION
ADA · §1557 · §504	**Malpractice / Negligence**	**Informed Consent · Standard of Care**	**HIPAA Privacy & Security Rules**	**Negligence**	**§1557 Language Access · Informed Consent**

Existing law already governs AI in nursing practice: HIPAA, civil-rights statutes, informed consent, and the standard of care all apply.

Texas BON Proposed New Position Statement (2024)

Board Recommendations for Safe AI Use

- **Be prepared & stay competent.** Pursue foundational and ongoing education in informatics, data science, and AI. Nurses are lifelong learners, responsible for their own continuing competence.
- **Build governance.** Employers and nurse leaders should form an ethics committee with guidelines to ensure responsible, transparent, and accountable use of AI.
- **Collaborate & speak up.** Coordinate AI use with the healthcare team; raise concerns and clarify any order when an AI tool seems inappropriate or unsafe.

The bottom line

AI is never the final decision maker in patient care. The nurse relies on nursing judgment and expertise and remains accountable to uphold the Nursing Practice Act and Board Rules whether or not AI is used.

Professional Responsibility Obligations

Standards of Nursing Practice — Board Rule 217.11

- **Know the law & policies —** comply with the NPA, Board rules, and facility AI policies [217.11(1)(A)]
- **Keep patients safe —** implement measures to maintain a safe environment [(1)(B)]
- **Protect privacy —** safeguard confidential patient data used with AI [(1)(E)]
- **Educate patients —** explain AI's role so patients can make informed decisions [(1)(F)]
- **Maintain competency —** obtain orientation and training on new technology [(1)(H),(R)]
- **Guard against bias —** provide nondiscriminatory care; watch for biased data [(1)(L)]
- **Question & clarify —** challenge inaccurate AI-driven orders [(1)(N)] and collaborate with the team [(1)(P)]

Quick Hitters

Some questions to get you started about understanding the tool you are evaluating

Gen AI

1. What Large Language Model are you using?
2. How is institutional information processed (recordings, transcriptions, draft copies)?
3. What were the training data AND what contextual information is needed to provide optimal responses?

Agentic AI

1. Does it respond when a clinician asks, or continuously monitor data and act on its own?
2. How many separate systems does it connect to or draw from when producing an output?
3. When an output is wrong or unexpected, how is it caught, corrected, and reported to overseers?

AI Nutrition Label or Applied Model Card

Name: **Developer:**	**Inquires or to report an issue:** abc@abc.com or +1 (999) 999- 9999

Release Stage:	**Release Date:**	**Version:**
Global Availability:	**Regulatory Approval, If applicable:**	

Summary: **Keywords:**	**Uses and Directions:** • **Intended use and workflow:** • **Primary intended users:** • **How to use:** • **Targeted patient population:** • **Cautioned out-of-scope settings and use cases:**

Warnings

- **Known risks and limitations:**
- **Known biases or ethical considerations:**
- **Clinical risk level:**

Trust Ingredients

AI System Facts:

- **Outcome(s) and output(s):**
- **Model type:**
- **Foundation models used in application, if applicable:**
- **Input data source:**
- **Output/Input data type:**
- **Development data characterization:**
- **Bias mitigation approaches:**
- **Ongoing Maintenance:**
- **Security and compliance environment practices or accreditations, if applicable:**
- **Transparency, Intelligibility, and Accountability mechanisms, if applicable:**

Transparency Information:

- **Funding source of the technical implementation:**
- **3rd Party Information, If Applicable:**
- **Stakeholders consulted during design of intervention (e.g. patients, providers):**

Key Metrics

Usefulness, Usability, and Efficacy		Fairness and Equity		Safety and Reliability	
Goal of metric(s):		**Goal of metric(s):**		**Goal of metric(s):**	
Result:	**Interpretation:**	**Result:**	**Interpretation:**	**Result:**	**Interpretation:**
Test Type:		**Test Type:**		**Test Type:**	
Testing Data Description:		**Testing Data Description:**		**Testing Data Description:**	
Validation Process and Justification:		**Validation Process and Justification:**		**Validation Process and Justification:**	

Resources

- **Evaluation References, If Available:**
- **Clinical Trial, If Available**
- **Reimbursement status, if applicable:**

CHAI

ChatGPT vs. Claude vs. OpenEvidence Model Card

ChatGPT (OpenAI)

Sourced from 59 pages on GPT-5 card & HealthBench website

ChatGPT for Healthcare

- HealthBench rubric consists of 5,000 conversations graded against "physician-written rubric criteria specific to that conversation" by 262 physicians in 60 countries and 26 medical specialties
- Tests the solution on "48,562 unique rubric criteria" and "includes only factually correct information supported by evidence or consensus"
- **Earlier this year, only one of the examples is tied to clinician use (one was tied to triaging acute scenarios)… now only 3 examples listed**

Testing Process

- Testing centered on disallowed content & hallucinations (self-harm, extremism, illicit content, etc.)
- Better translated materials into layman terms

Claude (Anthropic)

Sourced from 147 pages on 4.5 Sonnet System Card & Claude healthcare page

Claude for Healthcare

- Marketed for HIPAA-compliant administrative-focused use cases: "authorization, insurance, triaging, & documentation" and life science discovery use

Testing Process

- Tested on disallowed content and in-depth testing on safety, honesty, harmlessness
- Significantly more testing disclosures and information available

OpenEvidence

No public model card or system card; sourced from openevidence.com user guide (Clinical Curbside) & announcements

OpenEvidence for Healthcare

- Point-of-care clinical Q&A ("Clinical Curbside"): "quick, evidence-based answers to your clinical questions during patient care," with inline citations to guidelines & journals
- "Community Insights" answers may cite non-peer-reviewed sources (e.g., Curbsiders podcast)

Testing Process

- No published testing methodology; public claims only — "100% on USMLE" (Aug 2025); Mayo Clinic study: comparable to physician CDM in common scenarios (Apr 2025)

Content Agreements

- NEJM Group (Feb 2025), JAMA Network (Jun 2025), NCCN (Nov 2025), Wiley (Mar 2026); site banner adds Cochrane "and more"
- 12+ specialty-society partnerships 2025–26: ACC, AAFP, ACEP, ADA, ACOG, AUA, ASCO, SSO, GINA, NORD, AAO, AAO-HNSF

"While our evaluation results show the potential of LLMs in this area, none of them are effective enough for clinical settings. Common issues include extracting the incorrect entities, not using the correct equation or rules for a calculation task, or incorrectly performing the arithmetic for the computation."

Nikhil Khandekar et al., *MedCalc-Bench: Evaluating Large Language Models for Medical Calculations* (2024), arXiv:2406.12036.

OpenEvidence sources: openevidence.com/user-guide/clinical-curbside; openevidence.com/announcements (accessed June 9, 2026)

The risk shifts by tool — not just hallucination

Most attention goes to hallucination, but general-purpose LLMs and clinical AI point solutions carry distinct risk profiles.

General-purpose LLMs

PRIMARY RISK: Fabrication & data exposure

- **Hallucination from gaps —** no guidelines or journals at hand, so it invents plausible answers
- **Missing or fake citations —** confident tone, but sources are unverifiable or absent
- **Outdated & generic —** bounded by training cutoff; not clinical-grade or patient-specific
- **PHI exposure —** consumer tools may lack a BAA and can retain what you enter

Clinical AI point solutions

PRIMARY RISK: Overreliance on cited output

- **Residual hallucination, harder to catch —** unfaithful synthesis can hide behind a real citation
- **Misattribution —** a citation being present ≠ it supporting that specific claim
- **Retrieval error in → error out —** wrong, outdated, or wrong-population guideline retrieved
- **Automation bias —** citations look authoritative, so verification gets skipped

How AI risks map to your professional responsibilities

Each AI risk implicates a duty the nurse already owes — Texas BON Position Statement 15.31, Standards of Nursing Practice (Board Rule 217.11).

AI RISK	PROFESSIONAL RESPONSIBILITY (Board Rule 217.11)	WHY THE DUTY IS IMPLICATED
Inaccurate or hallucinated output	**Promote a safe environment [217.11(1)(B)] · Clarify questionable orders [(1)(N)]**	Acting on a wrong AI answer can harm a patient; the nurse must recognize and challenge it.
Overreliance / automation bias	**Exercise independent nursing judgment · Continuing competence [(1)(R)]**	Treating AI as the decision-maker abdicates the judgment and accountability the NPA requires.
Misattributed or unverifiable citations	**Competency with new technology [217.11(1)(H),(R)]**	Confirming a source truly supports the claim takes tool understanding — output isn't self-verifying.
PHI exposure / no BAA in place	**Protect client privacy & confidentiality [217.11(1)(E)]**	Entering patient data into ungoverned tools risks an impermissible disclosure.
Algorithmic bias / wrong-population output	**Provide nondiscriminatory care [217.11(1)(L)]**	Models built on narrow data can produce inequitable or unsafe guidance for some patients.
Can't explain the AI's role to the patient	**Patient education & informed decisions [217.11(1)(F)]**	The nurse must understand the AI well enough to inform the patient and support consent.
Unknown facility or legal rules on AI	**Comply with NPA, Board rules & policies [217.11(1)(A)]**	AI use must still conform to all applicable laws and facility policy.

Recent LLM Litigation

Case	Complaint	Legal Theories Pleaded
Florida AG / Dept. of Legal Affairs v. OpenAI + Sam Altman	Florida alleges OpenAI knowingly marketed ChatGPT, including to children, while concealing serious risks such as self-harm, addiction, and data collection from minors.	• FDUTPA violations • Negligence & gross negligence • Strict liability – design defect; Strict liability – failure to warn • Fraudulent misrepresentation • Public nuisance
Turner-Scott / Scott v. OpenAI + Sam Altman	Parents allege ChatGPT-4o gave their son personalized drug and medical advice and failed to recognize danger signs before his fatal accidental overdose.	• Strict product liability • Defective design & failure to warn; Negligence • CA Bus. & Prof. Code §§ 2052 & 4999.9 • CA UCL § 17200 • Negligent undertaking; wrongful death; survival
Garcia v. Character Technologies / Character.AI / Google	Plaintiffs allege Character.AI's chatbot contributed to a 14-year-old's suicide through compulsive, grooming-like interactions and a failure to activate safeguards.	• Wrongful death • Strict product liability – design defect & failure to warn • Negligence; negligence per se • Unjust enrichment • Intentional infliction of emotional distress • FL Unfair & Deceptive Trade Practices Act
Saucedo v. Sharp HealthCare / Sharp medical groups	Proposed class action alleges Sharp's ambient AI secretly recorded doctor-patient conversations and shared them with a third-party vendor without consent.	• CA Invasion of Privacy Act §§ 632 & 637.2 • CA Confidentiality of Medical Information Act • CA UCL § 17200 • Invasion of privacy / intrusion upon seclusion; Negligent misrepresentation
Commonwealth of PA, State Board of Medicine v. Character Technologies	Pennsylvania alleges a Character.AI chatbot held itself out as a licensed psychiatrist and supplied an invalid PA license number.	• Equity petition under PA Medical Practice Act §§ 422.1–422.53 • Injunctive relief for unauthorized practice of medicine

Liability

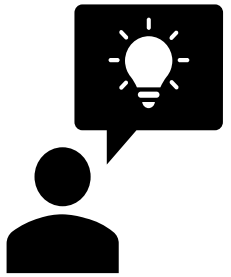
Case Review

- **Proving liability requires showing HOW an error occurred, not just that it did.** The AI "black box" makes this a high burden plaintiffs may struggle to meet.
- **AI performs differently across patient populations:** making it difficult to prove a clinician should have foreseen an error for a specific patient.
- **Courts do not CURRENTLY distinguish AI from traditional software:** creating unpredictable cross-case precedents that fail to account for AI's unique characteristics.

4-Factor Risk Assessment Framework

Not all AI tools carry equal risk. Calibrate adoption decisions and monitoring intensity by:

1. **Likelihood & nature of errors** — model quality, training data, task design, and clinical workflow integration
2. **Error detectability** — how much human visibility exists before an error reaches a patient
3. **Harm severity if uncaught** — clinical criticality of the function and patient acuity
4. **Litigation viability** — injury severity, ease of proving negligence, and causal clarity



Key Callouts

- **Developer liability will be shaped by private contracts and jurisdiction.**
- **Use contracts to demand training data transparency, indemnification splits, and post-deployment monitoring obligations**.
- **Standard of care morphing. Failing to adopt safe AI may itself become a basis for liability.**

Source: Mello & Guha, NEJM 390:271-278 (Jan 2024)

Risk Mitigation: Consolidated Recommendations

MITIGATION STRATEGIES

- **Audit logs** — capture what the AI decided, when, and where; set retention and weigh PSO / peer-review.
- **Post-deployment monitoring** — require vendors to notify of model drift and changes.
- **Insurance review** — confirm no AI-error exclusions or FDA-approval conditions.
- **Indemnification triggers** — for AI errors; carve AI errors out of liability caps.
- **Human-in-the-loop & override** — define when a clinician takes over and document the override.
- **Designated human review** — clinician finalizes output; test for hallucinations, accents, sensitive encounters.
- **Disclose AI identity up front** — label bots as AI; state scope and limits; never imply clinician authorship.
- **Consent for every voice** — patients, caregivers, family, interpreters, students, telehealth participants.
- **Feature-level disclosure** — flag recording on mute, after disconnect, or in background; voiceprint / speaker ID.
- **Revocation workflow** — real-time stop / pause with a clear retain-vs-delete rule for audio and transcripts.
- **Record-set decision** — decide whether recordings / transcripts are part of the designated record set.
- **Map interception risk** — vet third-party chat / analytics SDKs, session replay, and tracking pixels (wiretap exposure).

Practical Tips to Mitigate Your Risk

Everyday habits for safe, defensible AI use at the bedside

1

Get consent & disclose

Follow a standard consent / disclosure process before using AI live with a patient — e.g., “Is it okay if we use an AI assistant to help during your visit?”

2

Safeguard PHI under HIPAA

Confirm every tool has a signed Business Associate Agreement (BAA), share only the minimum necessary data, and chart accurately so the data feeding the tool is correct.

3

Monitor & audit with your institution

Partner with your organization on AI monitoring and auditing so you can use tools safely and raise appropriate challenges to inaccurate or unclear recommendations.

4

Keep a real human in the loop

Build genuine human review into your workflow. Verify what the AI produces — don’t be the “lazy human in the loop” who just rubber-stamps it.

5

Configure tools to fit you

Work with your technology and business leaders to configure tools to your preferences, comfort level, and clinical workflow.

References

1. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC medical education*. 2023;23(1):689.

2. Adams L, Fontaine E, Matheny M, Krishnan S. *An Artificial Intelligence Code of Conduct for Health and Medicine*. National Academies Press (US); 2025.

3. The White House. A National Policy Framework for Artificial Intelligence. Accessed March 21, 2026. https://www.whitehouse.gov/wp-content/uploads/2026/03/03.20.26-National-Policy-Framework-for-Artificial-Intelligence-Legislative-Recommendations.pdf

4. U.S. Department of Health and Human Services. HHS Artificial Intelligence Strategy & Implementation. Accessed March 17, 2026. https://www.hhs.gov/sites/default/files/hhs-artificial-intelligence-strategy.pdf

5. Shah R, Bozic KJ, Jayakumar P. Artificial intelligence in value-based health care. *HSS Journal®*. 2025;21(3):307-313.

6. Kim JY, Hasan A, Kueper J, et al. Establishing organizational AI governance in healthcare: a case study in Canada. *NPJ Digital Medicine*. 2025;8(1):522.

7. World Health Organization. *Ethics and governance of artificial intelligence for health: WHO guidance*. World Health Organization; 2021.

8. World Health Organization. *Ethics and governance of artificial intelligence for health: large multi-modal models. WHO guidance*. World Health Organization; 2024.

9. Dotan R, Blili-Hamelin B, Madhavan R, Matthews J, Scarpino J. Evolving AI risk management: A maturity model based on the NIST AI risk management framework. *arXiv preprint arXiv:240115229*. 2024;

10. Adepoju DA, Adepoju AG. Establishing ethical frameworks for scalable data engineering and governance in AI-driven healthcare systems. *Int J Res Publ Rev*. 2025;6(4):8710-8726.

11. Cary Jr MP, Russell RG, Silcox C, et al. Bridging the governance gap in health artificial intelligence: integrating nursing perspectives into the National Academy of Medicine Artificial Intelligence Code of Conduct. *Journal of the American Medical Informatics Association*. 2026:ocaf224.

12. El Arab RA, Al Moosa OA, Sagbakken M, Ghannam A, Abuadas FH, Somerville J, Al Mutair A. Integrative review of artificial intelligence applications in nursing: education, clinical practice, workload management, and professional perceptions. *Frontiers in Public Health*. 2025;13:1619378.

13. CHAI. The Responsible Use of AI in Healthcare (RUAIH). Accessed April 16, 2026. https://digitalassets.jointcommission.org/api/public/content/dcfcf4f1a0cc45cdb526b3cb034c68c2

14. Salwei ME, Morse K, Saria S, et al. Bridging the Gap: Consensus-Based Considerations for AI Usefulness in Healthcare. *The American Journal of Bioethics*. 2026/02/01 2026;26(2):1-6. doi:10.1080/15265161.2026.2617850

15. Keyes, T., Callahan, A., Pandya, A. S., Ambers, N., Banda, J. M., Fuentes, M., ... & Jain, S. S. (2026). Why and How to Monitor Deployed AI Systems in Health Care. NEJM Catalyst Innovations in Care Delivery, 7(6), CAT-25.

NURSIFYING AI

NAIIC Membership:

NURSING.FSU.EDU/NAIIC

Responsible AI in Nursing Micro-Credential:

NURSING.FSU.EDU/RESPONSILBEAI