IIElevenLabs

INDUSTRY REPORT · 2026

The Trust Framework for Healthcare Voice AI

Building Voice Agents That Clear Compliance Review, Earn Clinical Trust, and Scale Access

Ready to transform your operations?

Talk to a Healthcare Agent Deployment Expert.

Talk to a Healthcare Expert



John Durovsik

Agents Growth · ElevenLabs

INTRODUCTION

Governance at conversation speed

A hospital's after-hours line now routes through a voice agent instead of an answering service.
A Medicare enrollment center screens thousands of calls a day before a licensed agent picks up. A health plan's provider line resolves claim-status calls without a human on the line.

Healthcare is putting AI on the phone, and much of it has moved past the pilot stage: in NVIDIA's 2026 survey of more than 600 healthcare and life sciences professionals, 22% said AI agents are already deployed in their organizations, with another 19% expecting deployment within a year.[21]

EXECUTIVE SNAPSHOT

What you'll learn

1. Why the governance frameworks healthcare already uses, from NIST to the AMA, stop short of what a voice agent does on a live call.
2. Three phases that close the gap: design the boundary, enforce the boundary, prove the boundary held.
3. The responsibility ladder: which workflows an agent earns first, and what evidence moves it up a rung.
4. Three deployments with three different trust problems, including Banner Health, where go-live was conditioned on the buyer's own governance framework.
5. How five healthcare buyers read the framework differently, and why certifications set the floor while runtime behavior sets the ceiling.

AI AGENTS IN HEALTHCARE AND LIFE SCIENCES

NVIDIA 2026 SURVEY, 600+ PROFESSIONALS [21]

22%
say AI agents are already deployed in their organizations

19%
more expect deployment within a year

The question deciding each deployment has shifted from whether to use voice AI to how much responsibility to give it. Trust arrives in increments: an agent earns scheduling calls first, then insurance and intake questions, then a place in clinical support workflows, and each increment is granted on evidence.
This piece is about where that evidence comes from.

For conversational AI, the decisive layer of governance operates at conversation speed. A policy can be flawless on paper and fail in the four seconds after a caller says something nobody scripted. The framework below translates the field's governance consensus into that moment, and into the evidence a voice agent must produce before an organization hands it more responsibility.

THE TRANSLATION GAP

What the existing frameworks leave unanswered

A healthcare leader evaluating voice AI has no shortage of governance guidance. Deloitte, Gartner, NIST, KPMG, JAMA, and the AMA have each published a framework, and they converge on the same three-part shape: design controls before deployment, enforce them during operation, monitor them after. [1–7] The convergence is useful. It means the structure of trustworthy AI is close to settled, and nobody needs to invent it.

These existing frameworks stop short of the questions leaders face today. These frameworks are written at governance altitude, for AI systems in the abstract or for clinical prediction models reviewed by committees. A voice agent acts in real time, mid-conversation, with a patient on the line and no committee in the loop. NIST says manage risk continuously; it does not say what the agent should do in the next four seconds when a caller on a scheduling line says her chest hurts. The AMA toolkit says establish escalation policies; it does not say what a functioning escalation sounds like, how disclosure works when state law requires it verbally at the start and end of a call, or what data minimization means when the data is a live audio stream carrying PHI.

SIX FRAMEWORKS, ONE SHAPE1–7

Deloitte · Gartner · NIST · KPMG · JAMA · AMA

Design controls	Before deployment
Enforce them	During operation
Monitor them	After

WHAT A VOICE AGENT ADDS

Real time, mid-conversation, with a patient on the line and no committee in the loop.

The framework below closes that translation gap. It keeps the shape the field agrees on and adds what a voice agent specifically requires: the boundary it needs before launch, the enforcement that boundary demands on a live call, and the proof afterward that it held.

THE PRESSURE IS OPERATIONAL

The pressure to close the gap is operational. Eligibility and prior authorization work consumes the largest share of practice staff phone time, 45% in MGMA's March 2026 poll,8 and roughly 28.5 million Americans, about 9% of those age five and older, have limited English proficiency.10 McKinsey's workforce analysis found clinical-care labor productivity declined roughly 1% over two decades despite more than $150 billion a year in health IT, because organizations layer technology onto broken workflows.11 Governance is the gating variable the industry itself names: asked what shapes their approach to agentic AI, 40% of surveyed healthcare and life sciences organizations cited compliance with healthcare regulations, and performance and reliability topped the challenge list.21 Patients set conditions of their own. They are three times more likely to trust an agent embedded in a clinical system than a public chatbot, and roughly nine in ten say a clear path to a real person is essential.12

THE PRESSURE TO CLOSE THE GAP · KEY INDICATORS

Phone workload, language access and patient expectations all raise the bar for governance

45% — Phone

of practice staff phone time goes to eligibility and prior authorization work[8]

28.5M — Language

Americans, about 9% of those age five and older, have limited English proficiency[10]

~1% — Productivity

decline in clinical-care labor productivity over two decades, despite $150B+ a year in health IT[11]

40% — Compliance

of healthcare and life sciences organizations say regulatory compliance shapes their agentic AI approach[21]

3x — Trust

more likely to trust an agent embedded in a clinical system than a public chatbot[12]

~9 in 10 — Escalation

patients say a clear path to a real person is essential[12]

The framework

The three phases below keep the lifecycle the field converged on.
The names describe the job each phase performs for a voice agent.

1

Design the boundary

CONVERSATION

Scope limits, AI disclosure, retention, human-only moments

SYSTEMS OF RECORD

Source of truth, read / write rights, approvals, identity verification

BEFORE LAUNCH

Adversarial testing

→

2

Enforce the boundary

CONVERSATION

Disclosure where it matters, live escalation, data minimization, language and literacy

SYSTEMS OF RECORD

Grounded retrieval, permitted actions, safe failure

→

3

Prove the boundary held

CONVERSATION

Auditable calls, continuous evaluation

SYSTEMS OF RECORD

Actions verified

ONGOING

Certification upkeep

Post-deployment findings feed back into design

Design the boundary

Before an agent takes a live call, its perimeter has to exist in full. The conversational side covers what the agent will and will not do (no diagnosing, no treatment recommendations, no interpreting results, no approving or denying coverage), when and how it identifies itself as AI, what data it retains and for how long, and which moments belong to humans, including emergent symptoms, self-harm disclosures, and anything that triggers mandatory reporting.

The boundary is wider than conversation. Before launch, the organization also defines the agent's relationship to its systems of record: which system is authoritative for each piece of information, what the agent may read, what it may write or change, which actions require confirmation or human approval, and what the agent does when an integration is unavailable or returns conflicting information. The perimeter includes the actions and data access the organization delegates to the agent, and delegation is a design decision. When the workflow exposes protected information or permits sensitive actions, the boundary also defines what level of identity verification is required before the agent can proceed.

WHAT THE AGENT WILL NOT DO

- Diagnose
- Recommend treatment
- Interpret results
- Approve or deny coverage

MOMENTS THAT BELONG TO HUMANS

- Emergent symptoms
- Self-harm disclosures
- Mandatory-reporting triggers

ATTACK IT BEFORE PATIENTS DO

Designing the boundary also means attacking it: adversarial testing that simulates the difficult conversations, and the failed integrations, before any patient encounters either.

DISCLOSURE IS PART OF THE PERIMETER

Disclosure design is part of the perimeter, and it shows up in the first seconds of a call. MyPlanAdvocate's enrollment pre-qualification agent, which handles roughly 210,000 calls a month at peak with twice the conversion of the historical baseline, opens like this:

210,000
calls a month at peak

2x
Conversion vs the historical baseline

EXAMPLE: MYPLANADVOCATE CALL OPENING

AGENT "Hi, this is Ava calling on behalf of MyPlanAdvocate about your Medicare options. This call may be handled by AI. Is now a good time to talk?"

THE FRAMEWORK · PHASE 2

Enforce the boundary

Once the agent is live, design becomes behavior. Disclosure has to happen where it matters, and law is catching up to that view: California's AB 3030 requires health facilities to disclose AI-generated communications about a patient's clinical status verbally, at both the start and end of a call.[13]

Disclosure survives contact with patients. In EliseAI's deployments across health systems, 95% of callers who realize they are speaking with an agent complete the call anyway.

Escalation has to trigger and route in the moment, and data minimization has to happen at the point of use. Everlywell's engineering team made zero retention the deciding factor in its platform choice, so screening-outreach conversations with members were never stored longer than the call required. Enforcement also means adapting to the person: a guardrail that works only in English, or only in dense clinical language, is not functioning for a meaningful share of callers.

1. Recognizes the trigger
2. Declines the clinical question
3. Routes to the right human
4. Preserves the administrative task

The agent recognizes the trigger, declines the clinical question, routes to the right human, and preserves the administrative task. Every design decision in phase one exists to make those thirty seconds reliable.

CALIFORNIA AB 303013

Verbal disclosure of AI-generated communications about clinical status, at the start and end of the call.

ELISEAI · DISCLOSURE SURVIVES

95%

of callers who realize they are speaking with an agent complete the call anyway

EVERLYWELL · DATA MINIMIZATION

Zero retention

the deciding factor in its platform choice for screening outreach

WHAT CONVERSATIONAL ENFORCEMENT SOUNDS LIKE ON AN ADMINISTRATIVE LINE

CALLER "Before we finish, I've been having some chest pain since this morning. Should I still come in Thursday?"

AGENT "I can't assess symptoms, and chest pain is something a clinician should hear about right away. If this feels like an emergency, please hang up and call 911. Otherwise I can connect you to the nurse line right now and they'll take it from here. Your Thursday appointment stays on the calendar either way."

ENFORCEMENT · THE OPERATIONAL HALF

Grounded in the systems that govern the workflow

Enforcement also has an operational half. When the workflow requires authoritative healthcare data, the agent retrieves it from the system of record, takes permitted actions there, and fails safely when the system cannot provide a reliable answer. A trustworthy agent has to do two things at once: behave correctly in the conversation and stay grounded in the systems that govern the workflow.
A scheduling agent can follow every conversational guardrail and still fail by booking against stale availability. A benefits agent can communicate perfectly and still fail on outdated eligibility information. The difference between an agent that can conduct a convincing conversation and one that can reliably complete a healthcare workflow lives in this grounding.

TWO WAYS TO FAIL WITHOUT GROUNDING

Scheduling agent

✓ Follows every conversational guardrail
✕ Books against stale availability

Benefits agent

✓ Communicates perfectly
✕ Answers on outdated eligibility information

Grounded agent

Retrieves from the system of record, acts only where permitted, fails safely when the answer is not reliable

THE FRAMEWORK · PHASE 3

Prove the boundary held

Trust does not end at go-live. Every call needs to be auditable by a compliance team, at a depth a support-desk review does not reach. Performance needs continuous evaluation against the criteria set during design, on both dimensions: did the conversation stay inside the boundary, and did the actions land correctly in the systems of record.
A one-time validation at launch carries no weight six months later. Certifications need the same upkeep. Earning one happens once; keeping it current is ongoing work.

EVALUATED CONTINUOUSLY,
ON BOTH DIMENSIONS

1. Did the conversation stay inside the boundary?
2. Did the actions land correctly in the systems of record?

Every call auditable by a compliance team, at a depth a support-desk review does not reach.

A LINE THAT HAS TO HOLD

Where screening ends and judgment begins

A skeptical reader will notice an apparent contradiction: the boundary forbids clinical judgment, yet deployments described below run symptom assessments. The distinction matters enough to state plainly. Structured screening means the agent asks questions a clinical team authored, in the order the protocol sets, records the answers, and routes on thresholds clinicians defined in advance.

Judgment means interpreting what the answers mean, deciding what happens next, or advising the patient, and judgment never belongs to the agent. Clinicians write the protocol, own the thresholds, and act on the output. The agent administers the questionnaire and knows when to hand off. Held cleanly, that line lets voice agents extend clinical teams into high-volume screening work without ever making a clinical decision.

THE AGENT

Structured screening

Administers the questionnaire, knows when to hand off

1 Asks questions a clinical team authored

2 In the order the protocol sets

3 Records the answers

4 Routes on thresholds clinicians defined in advance

NEVER THE AGENT

Judgment

Clinicians write the protocol, own the thresholds, act on the output

× Interpreting what the answers mean

× Deciding what happens next

× Advising the patient

Clinicians write the protocol, own the thresholds, and act on the output.

The agent administers the questionnaire and knows when to hand off.

THE RESPONSIBILITY LADDER

Responsibility is earned in increments

The framework's phases describe one deployment. The more consequential pattern plays out across deployments, because organizations do not grant trust all at once. They extend it rung by rung, and each rung has an evidence requirement.

RUNG 3	Protocol-driven clinical support	Screening outreach · monitoring-alert response · post-discharge follow-up
GATE	Human-in-the-loop pilot, guardrail adherence scored on criteria agreed before launch. Never granted on a demo.	
RUNG 2	Regulated transactions	Benefits questions · enrollment · intake
GATE	Auditability, proven escalation, and the identity verification the workflow's data access demands.	
RUNG 1	Administrative access	Scheduling · eligibility checks · FAQs and routing
ENTRY	Answers live in systems of record; containment and accuracy measurable within weeks.	

The first rung is administrative access: scheduling, eligibility checks, routing. The answers live in systems of record, the risk surface is small, and containment and accuracy are measurable within weeks. The second rung covers regulated transactions such as benefits questions, enrollment, and intake, where the evidence requirement adds auditability, proven escalation, and the identity verification the workflow's data access demands. The third rung is protocol-driven clinical support: screening outreach, monitoring-alert response, post-discharge follow-up. Nobody should grant this rung on a demo. Cadence earned it through a human-in-the-loop pilot of roughly 2,000 calls, with the agent's guardrail adherence scored on evaluation criteria agreed before launch, and only then scaled to a quarter-million calls.

The industry's budgets are making the same turn from adopting to proving: 47% of organizations in NVIDIA's survey now prioritize optimizing AI workflows already in production, up from 34% a year earlier, while the share prioritizing new use-case identification fell from 47% to 37%.[21]

The climb runs on evidence, and it can run fast. Angle Health's first rung took six weeks from initial contact to live provider calls, and Cadence went from development to production in a month, because the controls each gate demands, simulation testing, audit trails, escalation routing, retention settings, governed integrations, come built into the platform.

34% → 47%
prioritize optimizing AI already in production[21]

47% → 37%
prioritize identifying new use cases[21]

6 weeks
Angle Health, first contact to live provider calls

1 month
Cadence, development to production

IN PRODUCTION

Three deployments, three trust problems

Each of the following faced a different hard question, and together they cover the framework end to end.

ANGLE HEALTH

Enforcement at volume

on a payer's provider line

CADENCE

The full lifecycle

at clinical stakes

BANNER HEALTH

Governance as a condition of go-live

at enterprise scale

90% of provider calls resolved end to end

Angle Health, a digital-first health insurer, fields thousands of calls a day, about 80% of them from doctors' offices asking about claim status, prior authorization, eligibility, and benefits. Its voice agents resolve 90% of provider calls end to end, verifying callers and pulling live claim and benefit data from the plan's core administration system. Deterministic workflows log required information and follow handoff rules during the call, and built-in call review lets the team confirm afterward that the rules held. Abraham Duplaa, Angle Health's Head of Product, puts the emphasis where the framework does.

90%

of provider calls resolved end to end

~80%

of calls come from doctors' offices

Claim status · prior authorization · eligibility · benefits, with live data from the plan's core administration system

> "ElevenAgents handled 90% of our call volume end to end, but what really impressed me was the 10% it was able to triage, understand why it couldn't handle it, and move it to the right team."
>
> Abraham Duplaa, Head of Product, Angle Health

CADENCE · THE FULL LIFECYCLE, AT CLINICAL STAKES

Screening at clinical stakes, decisions left to nurses

Cadence delivers chronic-disease management on behalf of more than 20 health systems, monitoring over 100,000 patients whose home devices generate more than 40,000 alerts a month. Its alert screening agent calls a patient when a vital goes out of range, runs a symptom assessment against Cadence's own clinical protocols, and warm transfers to a nurse on the 24/7 team when the conversation calls for it. The agent screens and escalates; nurses make the clinical decisions.

THE FRAMEWORK, END TO END

DESIGN	Evaluation Criteria Agreed Before Launch, Then A Supervised Pilot.
ENFORCE	Protocol-Driven Questions And Warm Transfer With Context.
PROVE	Calls Scored For Scope Adherence And False Positives, With The Agent Holding Its Clinical Guardrails Throughout Validation.

The operational results followed the trust work:

1h 48m → 3.5 min
median time from abnormal reading to patient contact

60%
of screened alerts resolved without clinician intervention

+47%
patients answered the agent more often than human outreach

BANNER HEALTH · GOVERNANCE AS A CONDITION OF GO-LIVE, AT ENTERPRISE SCALE

Inside the buyer's own AI governance framework

Banner Health, one of the nation's largest fully integrated nonprofit health systems, deployed ElevenAgents this year for patient-facing call handling, with the initial deployment live in primary care scheduling. The published shape of the deployment is the framework's enforcement phase in production: calls answered around the clock with no hold queue, appointments scheduled, rescheduled, and canceled directly in the EHR, warm transfer to a Banner team member with call context whenever a conversation needs one, and every agent operating inside Banner's own AI governance framework with human escalation built into every conversation. John Rares Almasan, Banner Health's senior vice president and chief AI, data and infrastructure officer, stated the bar in terms this framework recognizes.

> "Every AI system we deploy starts with the same question: does this make care safer, more accessible, and more trustworthy for our patients? This deployment clears that bar, with human escalation built into every conversation, full HIPAA compliance, and real-time oversight that lets us continuously improve the experience."
>
> John Rares Almasan, SVP and Chief AI, Data and Infrastructure Officer, Banner Health

The wider record runs through the same phases in narrower slices: MyPlanAdvocate's auditable enrollment guardrails, Everlywell's zero-retention outreach, and EliseAI's disclosure surviving 95% of the callers who notice it.

CERTIFICATION AND ACCURACY

Certifications set the floor

A certification tells you a system was designed and tested to meet a standard. It cannot tell you what happened on yesterday's calls. Sophisticated buyers hold onto that distinction: certifications set the floor of trust, and runtime evidence determines how much responsibility an agent should actually receive. Two kinds of certification have emerged in healthcare, and they answer different questions.

THE FIRST KIND · CERTIFIES THE HEALTH SYSTEM

The joint commission, responsible use of AI in healthcare

The joint commission's responsible use of ai in healthcare certification, launched in 2026 on guidance co-developed with the coalition for health ai, evaluates how an organization governs its own ai use; hackensack meridian health earned it first.[14]
For a system pursuing it, the vendor question is practical: does the platform produce the evidence the certification asks for, auditable call records, versioned agent changes, access controls, documented escalation. Those are elevenagents platform defaults.

THE SECOND KIND · CERTIFIES THE AI

AIUC-1, Built for AI agents around six risk categories

ElevenLabs became the first voice ai company to earn it, our agents subjected to 5,835 technical tests including voice-specific attacks such as background noise injection, pitch modulation, and multilingual prompts.[15, 16] The certification backs the first insurance coverage written specifically for ai agents,[17] and it sits alongside the security and privacy portfolio buyers should treat as table stakes, from SOC 2 through HIPAA-aligned operation with zero retention mode; the current list lives at Compliance.Elevenlabs.Io.[17]

5,835 Technical tests, including voice-specific attacks

CHAI's assurance labs and public model card registry signal where the market is heading: independent, reproducible verification. [18]

ACCURACY DESERVES THE SAME PROOF

Accuracy deserves the same standard of proof as safety, because clinical speech is unforgiving; hydroxyzine and hydralazine sit one misheard syllable apart. As of publication, our Scribe v2 Medical leads every model published on the public Eka Medical ASR benchmark at a 7.0% word error rate, on an evaluation set any buyer can re-run.[19] Cadence's production validation pointed the same way, with transcription accuracy holding on medication names across live patient calls.

SCRIBE V2 MEDICAL · EKA MEDICAL ASR BENCHMARK [19]

7.0%

word error rate, leading every published model

hydroxyzine — one misheard syllable — hydralazine

FIVE READERS

How each buyer reads this framework

A framework like this one gets read five ways, because five different leaders own five different pieces of the risk, and the weights shift with the setting: a hospital weighs disclosure and clinical escalation most heavily, a payer weighs auditability at enrollment-season volume. Knowing which piece is yours tells you where to start.

1	CIO STARTS AT · SYSTEM-OF-RECORD BOUNDARY	The CIO owns the systems the agent touches and the security review it must survive. Their reading starts at the boundary's system-of-record half: delegated access, tool gating, failure behavior when an integration degrades, and the certification floor their governance committee will ask about. Banner Health's deployment is their reference case, an agent operating inside the buyer's governance framework and writing directly to the EHR.
2	Innovation leader READS · THE WHOLE LIFECYCLE	The Innovation leader, whether the title reads chief innovation, digital, or AI officer, increasingly owns AI governance itself, and reads the whole lifecycle because their accountability spans it. The responsibility ladder is their planning tool and the evidence gates are their board slide. Almasan's question, does this make care safer, more accessible, and more trustworthy, is this reader's version of the framework in one sentence.
3	Operations leader READS · THE SEQUENCE	The Operations leader owns throughput and cost across every queue the agent will eventually touch. Their reading is the sequence: which workflow earns automation first, what containment and accuracy look like against baselines the organization already tracks, and when the evidence justifies the next rung. The ladder is their rollout plan, and grounding is the failure mode they watch for, because an agent that books against stale availability creates operational work instead of removing it.
4	Patient Access leader CONCENTRATES ON · ENFORCEMENT	The Patient Access leader owns the phones and the front-door queues the agent takes over first. Their reading concentrates on enforcement: what escalation sounds like, what resolution means on the calls their teams are measured by, and what patients experience at disclosure. They are usually the deployment's sponsor and its first beneficiary, and the evidence they need lives in the checklist below.
5	Clinical leader HOLDS · THE GATE	The Clinical leader holds the gate, and the framework earns their sign-off in one place above all: the line between screening and judgment. The Cadence model, where the agent administers clinician-authored protocols and nurses make every clinical decision, is the demonstration that the line can hold at scale.

THE 11-POINT CHECKLIST

What to check at each phase

Eleven questions recur across the deployments and certifications above, organized by the phase they belong to.

PHASE 1

Design The Boundary

1 Guardrail design

Are healthcare-specific boundaries built in: no diagnosis or treatment recommendation, no coverage determination, recognition of emergency symptoms, self-harm disclosures, and mandatory-reporting triggers?

2 Delegated access

Is it defined which systems are authoritative, what the agent may read, what it may write, which actions require confirmation or human approval, and what identity verification the workflow's data access requires?

3 Adversarial testing

Is There A Dedicated Environment For Testing Edge Cases Before Go-Live, Given That Only 44% Of Health Systems Report Having One Today?[20]

4 Certification floor

Does the vendor carry system-level security and ai-specific credentials, and does the deploying organization have its own governance in place?

PHASE 2

Enforce The Boundary

5 Disclosure

Does the system tell the patient or member they are speaking with ai, at the point where it matters?

6 Escalation

Is there a tested, reliable path to a human, and does the system actually use it when a conversation calls for it?

7 System-of-record grounding

Is the agent reading from and writing to the authoritative systems for the workflow, with defined behavior when those systems are unavailable, stale, or contradictory?

8 Data minimization

What gets retained, for how long, and does that match what the use case requires?

9 Plain-language delivery

Does the system adapt to the caller's language and health literacy, beyond the vendor's default script?

PHASE 3

Prove The Boundary Held

10 Auditability and monitoring

Can every call be reviewed by a compliance team, and is performance checked continuously against the design criteria, on both conversation and actions taken in systems of record?

11 Certification maintenance

Is the certification current, or was it a one-time badge?

CLOSING

The questions worth asking

1. What responsibility does the agent hold today?
2. What evidence did it produce to earn that responsibility?
3. What evidence will you require before it holds more

Organizations that can answer these three questions are already climbing the ladder safely, at whatever speed their evidence supports. Agents you can safely put in front of patients are built this way, one increment of proof at a time.

Ready to transform your patient access?

Talk to a Healthcare Agent Deployment Expert.

Talk to a Healthcare Expert

© 2025 ElevenLabs. All rights reserved.

Sources

1. Deloitte, "Trustworthy AI Governance in Practice"
2. Gartner, "AI Governance Needs More Than Policies: The AI TRiSM Framework"
3. NIST, "AI Risk Management Framework"
4. KPMG, "Global Tech Report 2026: Healthcare"
5. Chartis Group, "2026 Health System Outlook: Healthcare AI Strategy and Implementation"
6. Mello, Hernandez-Boussard, and Shah, JAMA, "AI, Health, and Health Care Today and Tomorrow"
7. AMA STEPS Forward, "Governance for Augmented Intelligence"
8. MGMA Stat, "Phones Are Still a Backlog Costing Medical Practices Time"
9. MGMA Stat, "Patient Access Priorities for 2026"
10. KFF, "Overview of Health Coverage and Care for Individuals with Limited English Proficiency"
11. McKinsey & Company, "The Real Future of Work in Healthcare"
12. Fierce Healthcare, "Patients Prefer Healthcare Providers' AI Agents to Public Chatbots, with Human Oversight Non-Negotiable"
13. California Legislature, "Assembly Bill 3030"
14. The Joint Commission, "Responsible Use of AI in Healthcare Certification"; Joint Commission and CHAI, "Guidance on Responsible Use of AI in Healthcare"
15. AIUC-1, "The Trustmark for Agents"
16. AIUC, "ElevenLabs Achieves AIUC-1 Certification"
17. ElevenLabs, "ElevenLabs Secures First-of-Its-Kind AI Agent Insurance"; ElevenLabs Trust Center
18. Coalition for Health AI, "CHAI Advances Assurance Lab Certification and Nutrition Label for Health AI"
19. Eka Care, "Eka Medical ASR Evaluation Dataset"
20. Becker's Hospital Review, "Most Health Systems Lack Dedicated AI Testing Platforms"
21. NVIDIA, "State of AI in Healthcare and Life Sciences: 2026 Trends"

© 2026 ElevenLabs. All rights reserved.