TRUSTLOGIX

EBOOK

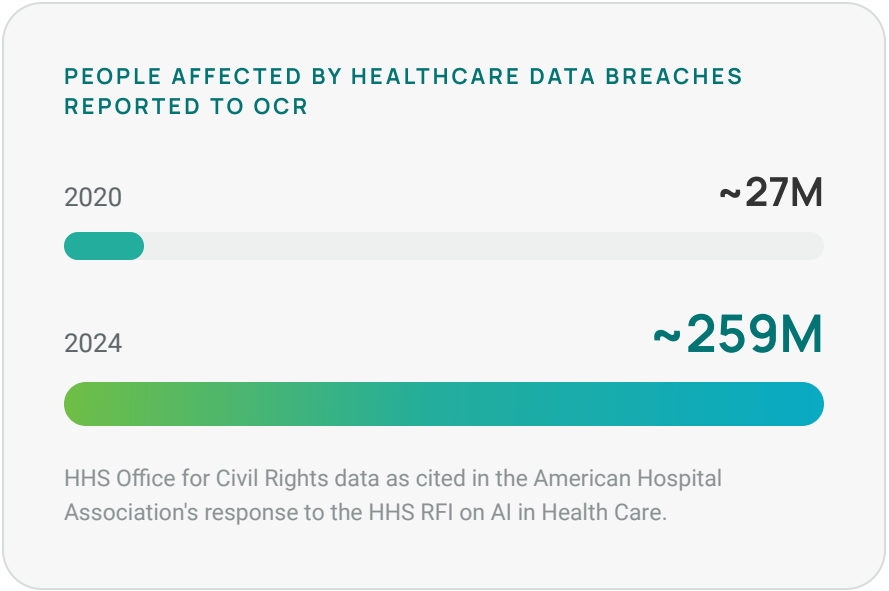
Four Ways Health Plans Can Secure AI Agent Data Access

Prior authorization, cross-plan data sharing, vendor agents, and audit evidence — where agentic AI meets PHI, and how access control has to change.

AI agents are moving into prior authorization, claims adjudication, care management, and member services faster than most health plans' access control programs were built to handle. These agents don't log in the way employees do. They don't request access through a ticketing queue. They connect directly to claims systems, clinical data stores, and member records, and they do it at a pace no manual review process can keep up with.

That speed is the point. It's also the risk. A health plan that gets AI agent access right can move prior authorization decisions, care coordination, and member outreach faster than ever. A health plan that gets it wrong exposes protected health information at a scale and speed no compliance team has dealt with before.

Regulators are already responding. In January 2025, the HHS Office for Civil Rights proposed the first major update to the HIPAA Security Rule in two decades, a change that would extend HIPAA protection to electronic PHI used in AI training data, prediction models, and algorithm outputs. At the same time, the population of people affected by healthcare data breaches reported to OCR grew from roughly 27 million in 2020 to roughly 259 million in 2024. The infrastructure connecting AI agents to sensitive data is expanding well ahead of the controls meant to govern it.

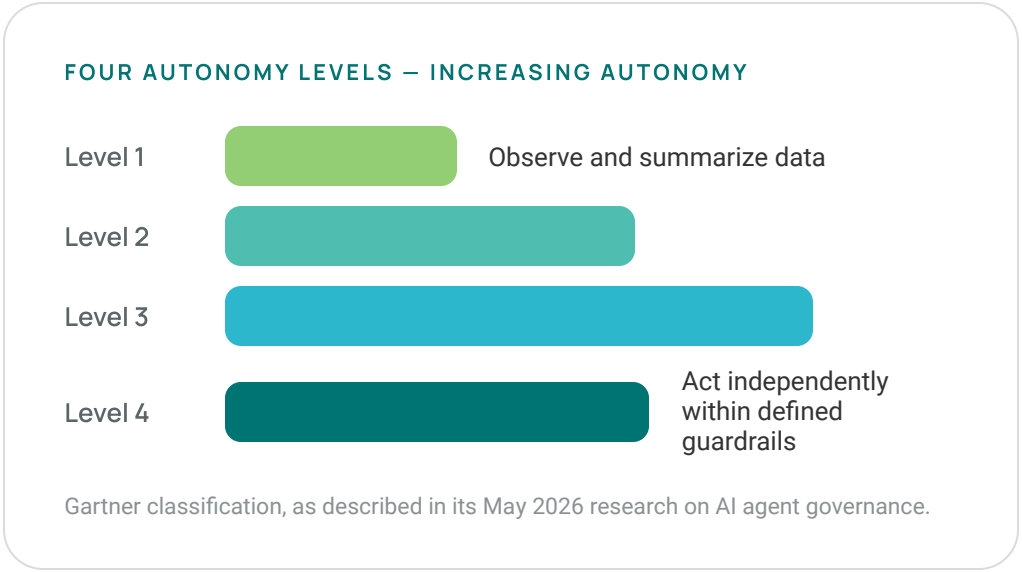


JANUARY 2025

The first major update to the HIPAA Security Rule proposed in two decades would extend HIPAA protection to electronic PHI used in AI training data, prediction models, and algorithm outputs.

Gartner's research on AI agent governance offers a useful way to think about the problem. Rather than treating governance as a single on-or-off switch, Gartner classifies agents into four autonomy levels, from agents that only observe and summarize data, up through agents that act independently within defined guardrails. Gartner's May 2026 research warns that applying the same governance controls to every agent regardless of its autonomy level is a leading cause of AI agent failure: overly rigid controls slow down low-risk agents and push teams toward shadow deployments, while overly loose controls let high-autonomy agents operate with far more access than their task requires.

For health plans, that autonomy-level lens maps directly onto four places where AI agents create new data access risk. This ebook walks through each one.



Gartner classification, as described in its May 2026 research on AI agent governance.

Overly rigid controls

Slow down low-risk agents and push teams toward shadow deployments.

Overly loose controls

Let high-autonomy agents operate with far more access than their task requires.

Four places where AI agents create new data access risk

1 **Prior authorization and care coordination agents**

Access scoped to the decision the agent is making, not to a case manager's role.

2 **Cross-plan and network data sharing, scoped by policy**

Business purpose, not platform credentials, decides what an agent can query.

3 **Third-party and vendor agent access**

A visibility problem before it becomes an access problem.

4 **Proving it: audit evidence when agents touch PHI**

Evidence captured at the moment of the decision, not reconstructed later.

1 Prior authorization and care coordination agents

Prior authorization and care coordination are two of the fastest-growing use cases for agentic AI in the payer space, and also two of the most sensitive. An agent reviewing a prior authorization request needs access to clinical documentation, diagnosis codes, and treatment history. A care coordination agent may need to see claims history, provider notes, and social determinants of health data to flag a member for outreach.

The risk isn't that these agents need PHI. It's that most access models grant them PHI at the same broad, role-based level a human case manager would get, rather than scoping access to the specific decision the agent is making in that moment. An agent evaluating a single prior authorization request doesn't need standing access to a member's entire clinical history; it needs the specific data elements relevant to that request, for the duration of that request.

Applying Gartner's autonomy framework here helps draw the line. A Level 1 or Level 2 agent that only summarizes clinical documentation or drafts a recommendation for human review can operate with narrower, read-only access and heavier logging. A Level 3 or Level 4 agent that can actually approve or deny a request, or trigger outreach, needs access scoped to task intent, evaluated at the moment of the request, not provisioned in advance as a standing entitlement.

ROLE-BASED, STANDING

Case manager entitlement

- Clinical documentation
- Diagnosis codes
- Treatment history
- Claims history
- Provider notes
- SDOH data
- Entire clinical history

SCOPED TO TASK INTENT

One prior authorization request

- Clinical documentation
- Diagnosis codes
- Treatment history
- ~~Entire clinical history~~

For the duration of that request.

DRAWING THE LINE BY AUTONOMY LEVEL

Level 1–2

Summarizes documentation or drafts a recommendation for human review — narrower, read-only access and heavier logging.

Level 3–4

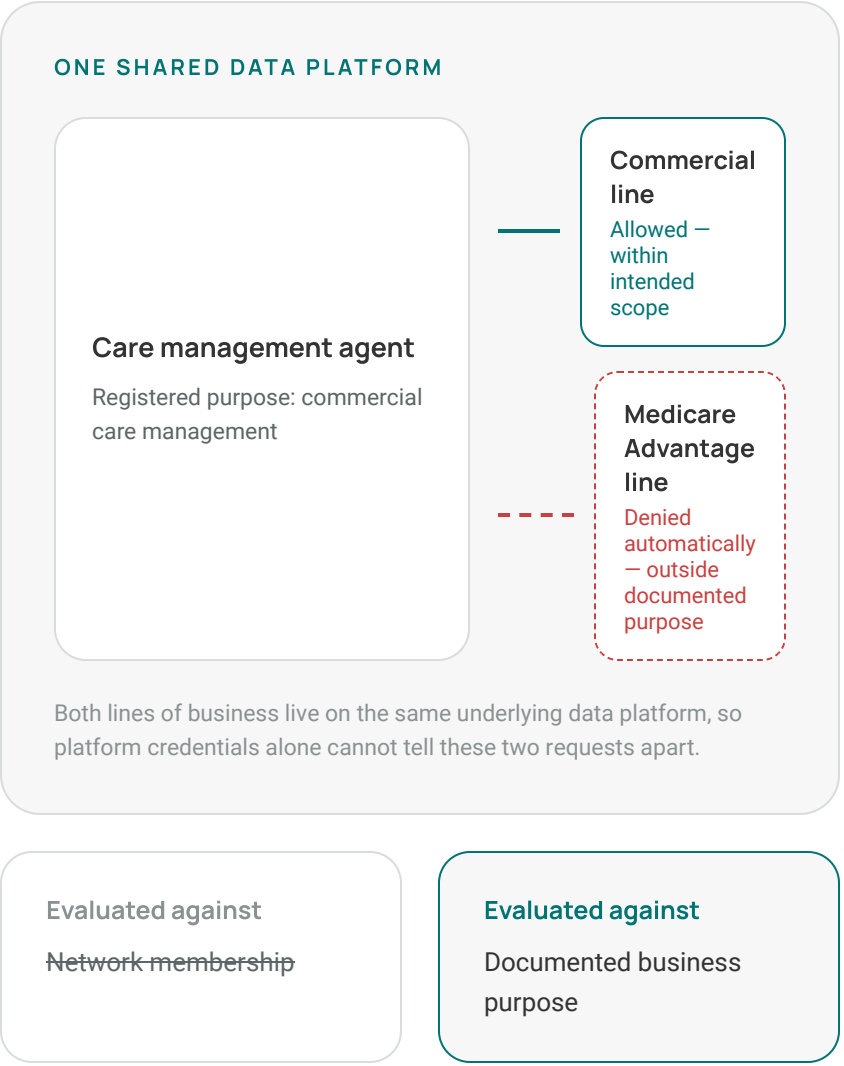
Approves or denies a request, or triggers outreach — access scoped to task intent, evaluated at the moment of the request.

2 Cross-plan and network data sharing, scoped by policy

Health plans increasingly operate across shared data infrastructure: multi-plan claims platforms, shared provider networks, care management systems that span multiple lines of business. AI agents built for one plan or one product line often have technical access to that broader shared environment, even when their intended purpose is much narrower.

This is where access control based on network membership or platform credentials breaks down. An agent built to support a commercial plan's care management program may technically be able to query data belonging to a Medicare Advantage line, simply because both live on the same underlying data platform. Network-level or platform-level access controls can't tell the difference between an agent operating within its intended scope and one reaching outside of it.

The fix is binding every agent to a documented business purpose, and evaluating each data request against that purpose rather than against network membership alone. An agent registered for commercial care management should be denied a request touching Medicare Advantage data automatically, regardless of what the underlying platform's access controls would technically allow.



3 Third-party and vendor agent access

Health plans work with a wide network of external parties: pharmacy benefit managers, care management vendors, analytics partners, and third-party administrators. Many of these vendors are now embedding AI agents into the services they deliver, agents that connect directly into the health plan's data environment to pull claims, eligibility, or clinical data.

This creates a visibility problem before it creates an access problem. Health plans frequently cannot answer basic questions about vendor-deployed agents: which ones have access to PHI, what specific data elements they can reach, and whether that access matches the scope of the contracted service. An analytics vendor's agent that was scoped to aggregate, de-identified reporting can end up with access to identifiable claims data simply because no one re-evaluated its entitlements after the initial integration.

Closing this gap requires the same registration and intent-binding approach used for internally built agents, extended to every third-party agent touching plan data. Each vendor agent should be registered with a documented purpose, evaluated against that purpose at runtime, and subject to the same audit trail as an internally built system. If a vendor's agent needs to be cut off, whether because of a security incident or a contract dispute, that cutoff needs to happen in seconds, not through a support ticket to the vendor.

THREE QUESTIONS HEALTH PLANS OFTEN CANNOT ANSWER

1. Which vendor agents have access to PHI?
2. What specific data elements can they reach?
3. Does that access match the scope of the contracted service?

ENTITLEMENT DRIFT — ILLUSTRATIVE

AT INTEGRATION	→	TODAY
Analytics vendor agent scoped to aggregate, de-identified reporting		Access to identifiable claims data — entitlements never re-evaluated

Cutting off a vendor agent should take seconds, not a support ticket to the vendor.

4 Proving it: audit evidence when agents touch PHI

Health plans face frequent audits from state insurance regulators, CMS, and OCR, and those audits increasingly ask questions that weren't designed with AI agents in mind: who accessed this member's data, under what authority, and why. When the "who" is an autonomous agent rather than a named employee, most existing audit processes have no good answer.

The HHS Office for Civil Rights has already made clear that using an AI system does not relieve a covered entity of its HIPAA compliance obligations, and that covered entities remain responsible for breach notification even when the breach originates with an AI vendor. Forrester's 2026 healthcare predictions point to a similar theme: health organizations will need to rethink oversight of both AI governance and vendor partnerships as AI moves deeper into core operations. Regulators are not going to accept "the agent did it" as an answer to an audit question.

The evidence health plans need looks different from a standard user access log. It has to show, for every agent decision touching PHI: which agent made the request, what policy authorized it, what data was returned or withheld, and whether the request matched the agent's documented purpose. That level of detail has to be captured automatically, at the moment of the decision, because reconstructing it after the fact from application logs is rarely possible.

Evidence for one agent decision	Captured at the moment of the decision
Which agent made the request	Registered agent identity
What policy authorized it	Policy in force at request time
What data was returned or withheld	Fields released, fields masked
Whether it matched the documented purpose	Purpose match, recorded

State insurance regulators

CMS

OCR

> Regulators are not going to accept "the agent did it" as an answer to an audit question.

The four risks above share a common root cause: access models built for human users and static roles don't hold up against AI agents that operate at machine speed and touch data in ways no static role was designed to anticipate. Gartner's guidance on proportional, autonomy-based governance offers the right mental model. Applying it well requires binding every agent, internal or vendor-deployed, to a documented purpose, evaluating that purpose at the moment of each data request, and capturing audit-ready evidence automatically.

TrustLogix, the AI agent data security platform, helps health plans apply exactly this model: registering AI agents alongside human users, enforcing intent-based access control at the point sensitive data is actually accessed, and producing continuous audit evidence for SOC 2, HIPAA, and state insurance regulatory review. To see how this works against your own environment, schedule time with the TrustLogix team.

See how this works against your own environment

SCHEDULE TIME WITH TRUSTLOGIX

SOC 2	HIPAA	State insurance review

SOURCES

Sources: Gartner press release, "Gartner Says Applying Uniform Governance Across AI Agents Will Lead to Enterprise AI Agent Failure," May 26, 2026. HHS Office for Civil Rights data as cited in the American Hospital Association's response to the HHS RFI on AI in Health Care. HIPAA Journal, "When AI Technology and HIPAA Collide." Forrester, "Predictions 2026: The Year AI Tests The Heart Of Healthcare."