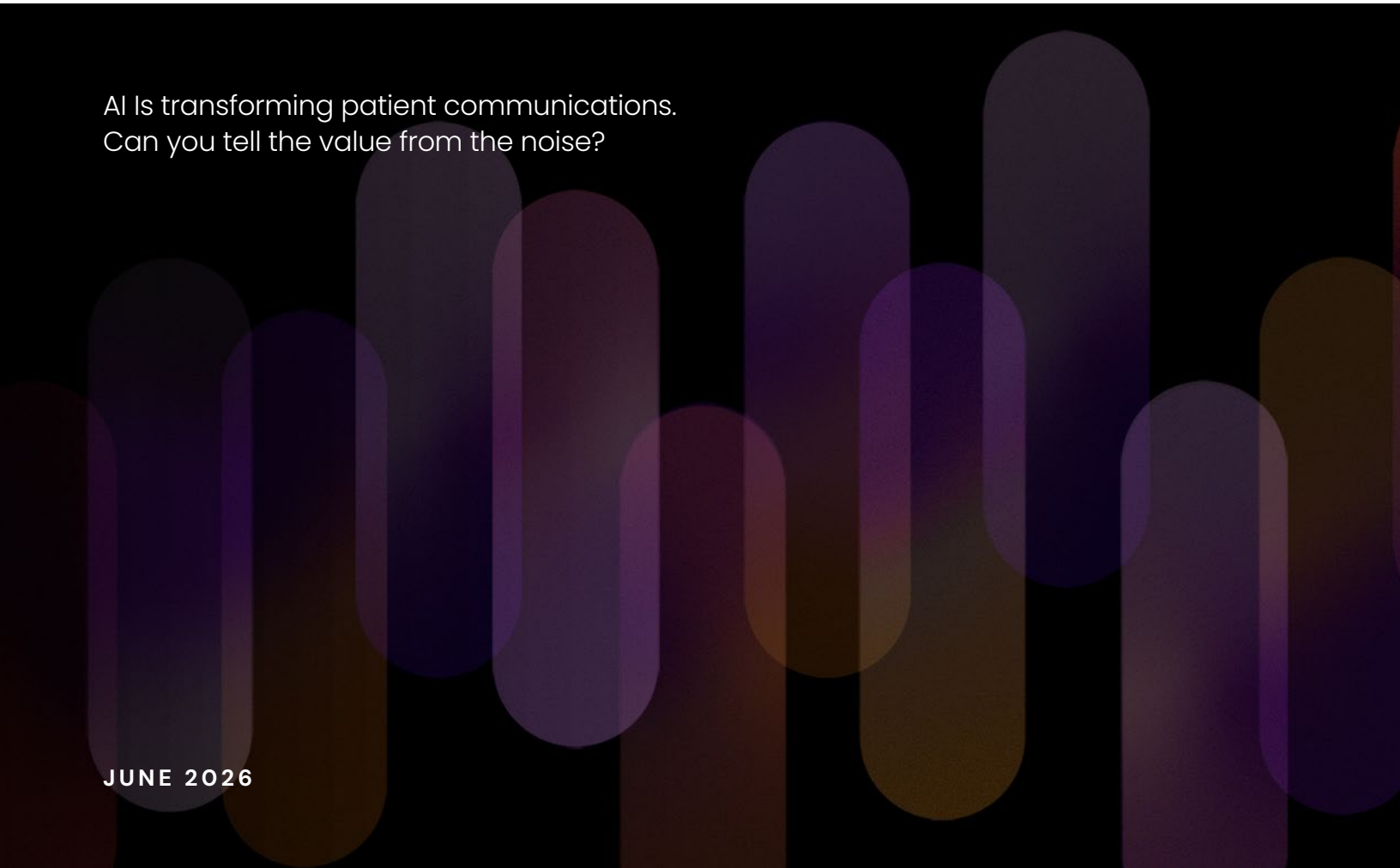
artera
Experience by Your Side

EDUCATIONAL SERIES

Decoding AI Agents for Healthcare

Expert Guidance to Help You Master the Shift to Agentic AI



JUNE 2026

Contents

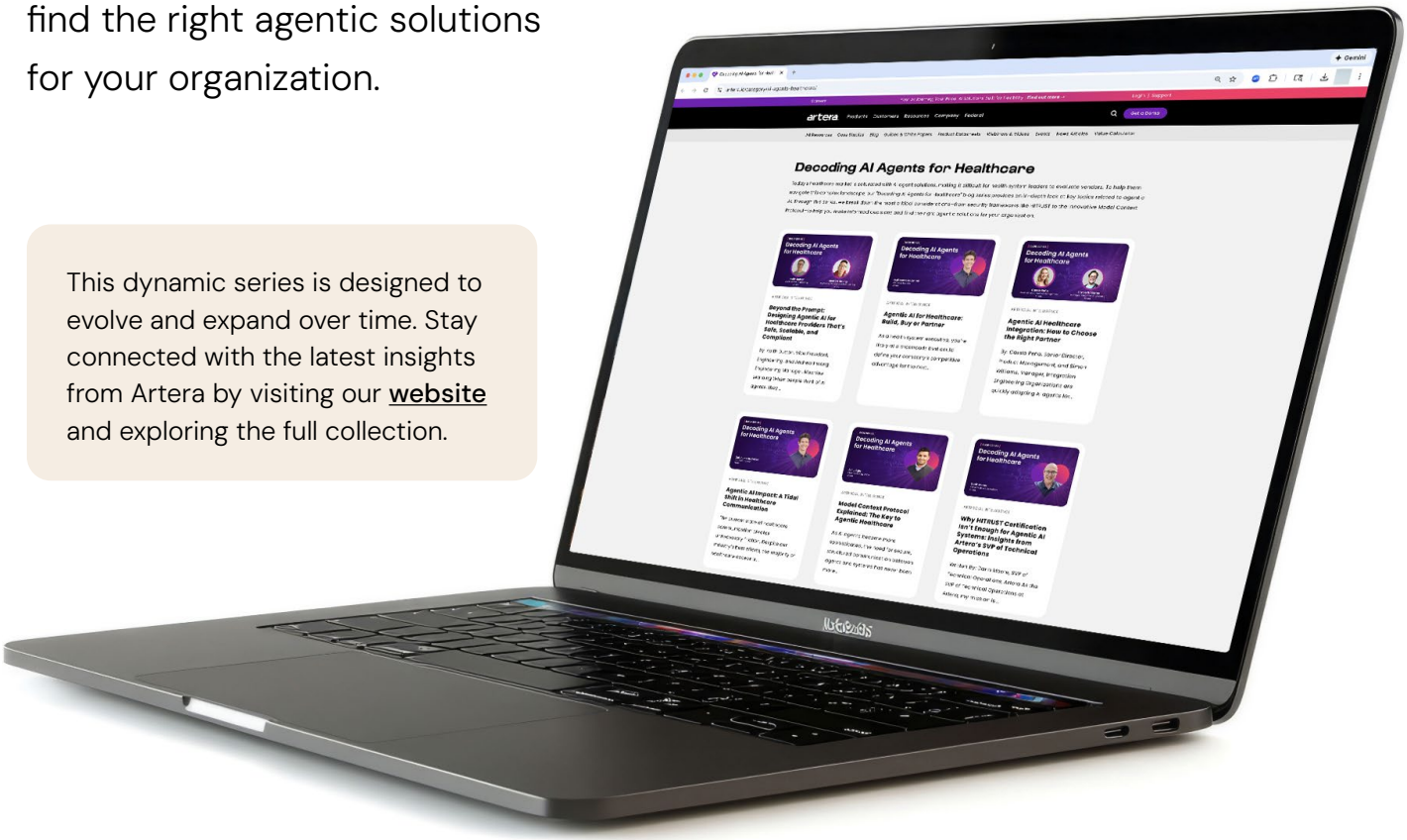
Introduction

Today's healthcare market is saturated with AI agent solutions, making it difficult for health system leaders to evaluate vendors. To help them navigate this complex landscape, our "Decoding AI Agents for Healthcare" perspective series offers an in-depth look at key topics related to agentic AI. Through this series, Artera experts break down the most critical considerations — from security frameworks like HITRUST to the innovative Model Context Protocol — to help you make informed decisions and find the right agentic solutions for your organization.

>80%

of health system respondents said they didn't have the resources to identify, select and implement AI solutions at their organization — according to a recent HFMA survey of 233 health systems.[1]

This dynamic series is designed to evolve and expand over time. Stay connected with the latest insights from Artera by visiting our **website** and exploring the full collection.



[1] Healthcare Financial Management Association, "**Health System Readiness For Artificial Intelligence**," August 2025

Agentic AI Impact: A Tidal Shift in Healthcare Communication

AUTHOR:
Guillaume de Zwirek
CEO and Co-Founder, Artera

The current state of healthcare communication creates unnecessary friction. Despite our industry's best efforts, the majority of healthcare access is still coordinated over the phone. Across our customer base, we consistently find that more than 70% of call volume relates to basic administrative tasks: confirming appointments, canceling, rescheduling, scheduling new visits, and department transfers.

These are fundamentally simple tasks that create bottlenecks in the system. Patients face long hold times, high abandonment rates, and limited access to care coordination outside business hours. Meanwhile, healthcare staff spend valuable time on routine tasks instead of focusing on higher-acuity patient needs.

Patient communication isn't just a part of patient access; it's the foundation of it. When we remove communication barriers, we dramatically improve access to care. I believe agentic AI is poised to accelerate the breakdown of common patient communications barriers — faster and more effectively than ever before.

Agentic AI: A Transformative Force for Patient Communications

Agentic AI represents the next major tidal wave hitting healthcare. Hundreds of companies are entering this space, with hundreds of millions of dollars being raised. The pace of change is extraordinary — faster than any technology revolution we've experienced in our lifetimes, including mobile, web, and social media.

We're seeing new infrastructure upgrades that meaningfully advance the technology's potential on what feels like a daily basis. This rapid innovation is driving improvements across the board — from reducing latency to improving background noise suppression.

At Artera, we're deeply committed to driving innovation in agentic AI. Over the past few months alone, we've updated our underlying infrastructure more than 20 times, achieving exponential improvements to product quality. Our latency now sits below 500 milliseconds, we've deployed dozens of MCP tools and servers, our continuous learning framework is in production and ingesting live transcripts, and we're supporting multiple languages. These aren't just incremental improvements — they're transformational leaps that directly benefit patient care.

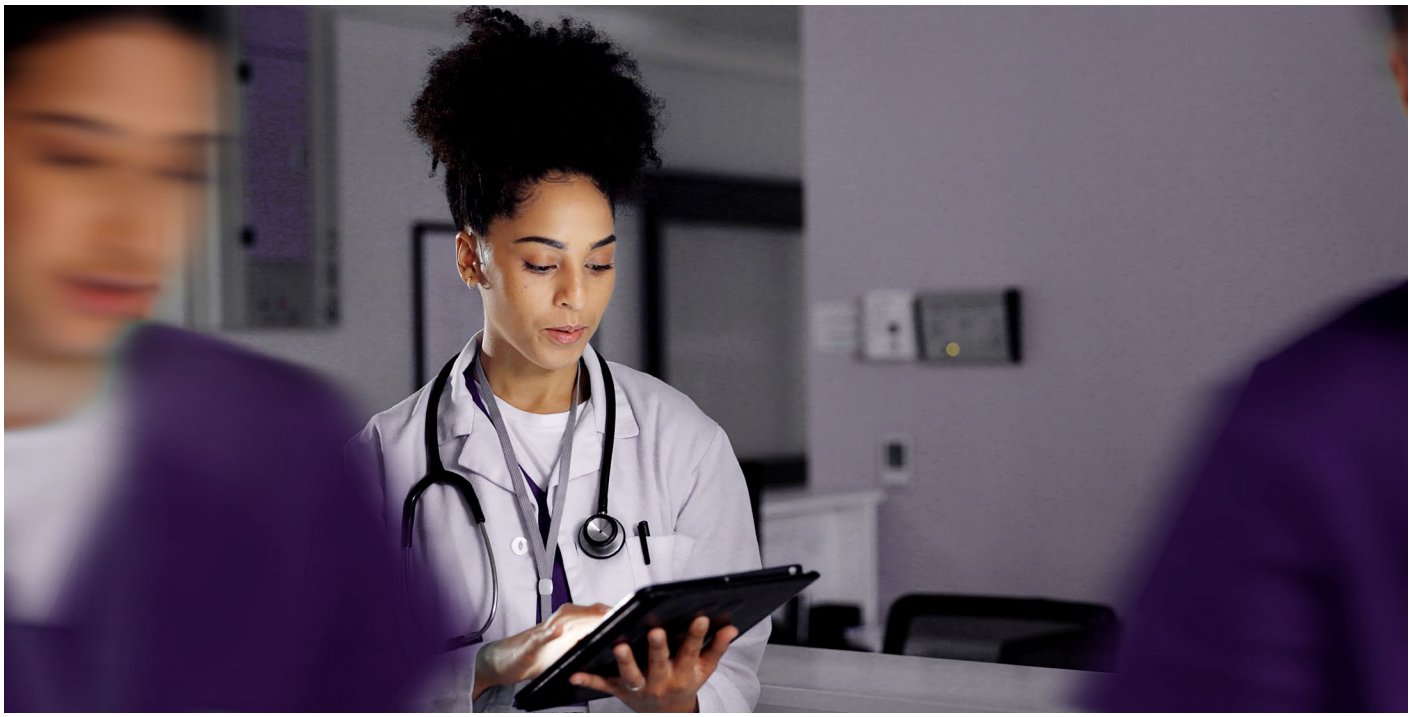
Deterministic AI Agents: A Stepping Stone to Autonomous AI Agents

Despite significant technological advancements, our core mission has remained unchanged for a decade: making healthcare number one in customer service. To us, "customer service" refers to how patients experience care outside the four walls of a hospital or clinic.

Our focus has been specifically on communications — enabling seamless, asynchronous interactions between healthcare providers and patients, no matter where they are. The rise of artificial intelligence has fundamentally changed how this experience will unfold in healthcare moving forward.

Our AI-powered, deterministic Flows Agents deliver two strategic benefits that align with our vision:

First, it serves as a definitive knowledge base of proven pathways that have been hardened over many years across hundreds of institutions and for millions of



patients. This knowledge repository is an invaluable tool for training fully autonomous agents.

Second, Flows Agents act as a stepping stone into fully autonomous AI. Healthcare operates under strict regulations, where technology — especially autonomous AI — can have life-and-death implications. Flows uses deterministic logic combined with natural language understanding (NLU) to guide patients through specific automated journeys. These pathways eliminate risks associated with hallucinations and jailbreaking, helping build initial comfort with AI among healthcare leadership, while creating a clear path toward full autonomy through our AI Agents.

Where to Start: Automating Routine Interactions That Create Barriers to Care

Agentic AI offers the opportunity to automate the routine administrative phone tasks I initially mentioned, while making them available 24/7 at a quality bar that approaches that of real humans. By reducing the burden of simple interactions, we elevate staff to focus on higher-acuity patient needs, reduce hold times and abandonment rates, and ultimately improve access to care.

As system interoperability continues to advance, we'll be able to streamline more routine patient tasks, creating a truly personalized, concierge-like experience for every patient while simultaneously reducing healthcare operating costs.

This enhanced accessibility will span all communication channels — voice, messaging, and web — ensuring patients can connect anytime, anywhere. At Artera, we're excited to shape the future of patient communication, where every individual benefits from 24/7 concierge-like care.

The More Things Change, the More They Stay the Same

While technology is changing rapidly, the fundamental challenges of healthcare communication remain the same. Patients need frictionless, asynchronous communication channels that allow them to engage with their care teams on their own terms and schedules. Providers, on the other hand, need efficient, automated workflows with the flexibility to involve human intervention when necessary.

I believe our decade of experience in tackling these core challenges gives us a distinct edge as we innovate with agentic AI. By building on a foundation of proven solutions and deep market knowledge, we not only leverage the potential of technology but also address the real, pressing needs of the industry.

Why HITRUST Certification Isn't Enough for Agentic AI Systems

AUTHOR:
Darin Moore
SVP of Technical Operations, Artera

Given the dynamic nature and rapid change of the agentic AI landscape, we have a unique opportunity today to ensure that our security protocols remain agile and resilient in the face of new challenges. If this past year has taught us anything, it's that as AI agents become more advanced and independent, the risks of data breaches, hallucinations and leaks can escalate quickly.

In this new era of agentic AI, data security requires a fundamental shift in strategy, and can no longer rely on static, point-in-time assessments. Instead, it demands continuous monitoring, multi-layered security frameworks and the integration of human oversight with AI-powered validation.

Healthcare providers seeking agentic AI solutions need partners who truly understand this and have built robust security systems designed specifically with agentic AI in mind.

Why HITRUST Alone Falls Short in the AI Era of Healthcare

Traditional frameworks like HITRUST are a solid starting point for protecting healthcare data, but they just can't keep up with how fast agentic AI systems evolve. While HITRUST shows a commitment to safeguarding PHI, securing agentic AI requires a whole new approach.

Here's the thing: agentic AI doesn't play by the same rules. These systems are constantly learning, adapting and making decisions on their own. What worked yesterday might not work today, and something secure this morning could have vulnerabilities by the afternoon. A one-time security assessment just doesn't cut it anymore — we have to be vigilantly guarding the way that AI is using our data.

It gets trickier when you factor in how AI models get updated, retrained or tweaked between security reviews. Every change can bring new risks or behaviors that weren't there before. Traditional frameworks simply don't have the flexibility to keep up with these rapid changes, leaving organizations open to threats that didn't even exist during their last compliance check.

Beyond Compliance: A Comprehensive Multi-Pillar Approach to Security

Just relying on HITRUST isn't enough anymore in this new era. Working with vendors with multiple certifications gives you stronger, layered protection. That's why leading health tech companies are choosing a mix of certifications to handle the dynamic nature of AI security. At Artera, we view security as a puzzle where each certification plays a specific role:

- **HITRUST:** the foundational layer for healthcare; demonstrates a commitment to safeguarding PHI
- **SOC 2 Type 2:** third-party audit that highlights strong internal controls around data and systems — it's a key signal of operational maturity for the business as a whole
- **ISO 27001:** general framework that provides the foundation for information security management systems in place
- **ISO 27017:** certification that specifically addresses cloud service security
- **ISO 27018:** certification that focuses on personally identifiable information (PII) protection in an organization's environment
- **ISO 27701:** certification that covers privacy management and an organization's commitment to keeping any privacy-related information confidential

As you can see, each certification plays a different role. When these pieces come together, they create a multi-pillar approach to security.

At Artera, we're not just meeting these standards — we're also pursuing FedRAMP High authorization, which is the Federal Risk and Authorization Management Program's most rigorous security baseline for cloud services handling highly sensitive government data

(in fact, Artera recently achieved "in process" FedRAMP High designation).

So why does this matter? Pursuing FedRAMP High status reflects our commitment to the highest level of security protocols, elevating our approach to data protection and enhancing our understanding of the evolving security landscape.

Security Considerations for Evaluating Agentic AI Partners

Beyond those certifications listed above, health system leaders should focus on three fundamental areas when assessing potential agentic AI vendors: data containment, spillage prevention and hallucination mitigation. These represent the most significant risks unique to AI systems, and require specialized approaches that traditional security frameworks don't address.

- **Data Containment**
  Involves ensuring that PHI and PII remain within secure, controlled environments, rather than being exposed to publicly accessible large language models (LLMs).
- **Spillage Prevention**
  Addresses the risk of information crossing between different patient sessions or unauthorized data access.
- **Hallucination Mitigation**
  Reduces or eliminates the generation of false, misleading, or nonsensical information by artificial intelligence models, particularly large language models (LLMs).

In addition to the preventive measures mentioned, continuous monitoring and real-time alerts are essential while agents are active.

Building a Culture of Security, Not Just Compliance

While no system is ever 100% secure, we can do a lot to protect ourselves by using every available safeguard and holding ourselves accountable. The goal is to keep both internal and external threats from compromising our systems. Just as important is having a clear audit trail so we can handle any issues that come up. Above all, we need to protect the healthcare data with all we've got. This includes fostering a culture of security and continuous improvement.

At Artera, I'm proud to say that security isn't just a checkbox or a compliance exercise. It's a core business principle and vital investment. Over the past few years, I've witnessed a remarkable cultural shift within our organization. Security has become a collective effort embedded in everything we do.

I've observed a growing interest in security across teams, functions, and employees. Colleagues are asking insightful questions, actively expanding their knowledge, and sharing valuable security insights throughout the company. What stands out most is the heightened curiosity and engagement. It's both inspiring and encouraging to witness this level of commitment.

Preparing for the Future of Agentic AI Security

As AI continues to play a bigger role in healthcare, keeping systems secure is only going to get more complicated and more important. The organizations that prioritize strong security partnerships now will be better positioned to take full advantage of AI's benefits while keeping patients' trust intact.

When choosing an agentic AI partner, it's a good idea to focus on vendors who not only have solid security measures in place today but are also committed to staying ahead of future challenges. I encourage providers to look for vendors who stay on top of AI security trends, invest in research and innovation, and can quickly adapt to new threats with effective solutions.

Model Context Protocol Explained: The Key to Agentic Healthcare

AUTHOR:
Ashu Agte
Technical Advisor, Artera

As AI agents become more sophisticated, the need for secure, structured communication between agents and systems has never been more important. Enter Model Context Protocol (MCP) — a new approach that's redefining how AI agents interact with external systems while maintaining strict security boundaries.

While traditional APIs have served machine-to-machine communication well, they now fall short when it comes to agentic AI interactions. MCP fills this gap by providing a specialized protocol designed specifically for AI agents, complete with built-in security features that help prevent data spillage and reduce hallucinations.

What is Model Context Protocol (MCP)?

Model Context Protocol is a new standard for connecting AI models to external tools, data sources, and services, so they can more effectively communicate. Essentially, it functions as an API designed specifically for AI agents. Developed by Anthropic as an open source protocol in late 2024, MCP has quickly gained traction across the industry, despite being less than a year old.

The protocol operates on a simple but powerful premise: instead of giving agents direct database access or unlimited system permissions, MCP creates a controlled interface that defines exactly what an agent can and cannot do. This approach fundamentally changes how we think about agent-system integration.

What are the Three Pillars of MCP?

MCP architecture consists of three core components that work together to create a comprehensive communication framework that addresses the key challenges of agent deployment: capability definition, information access, and response consistency.

PILLAR 1 Tools: the Agent's Capabilities

Tools represent the specific actions an agent can perform within a system. These are discrete functions that agents can call to interact with external services. Each tool has a defined scope and purpose. An agent cannot perform actions beyond its available toolset, creating natural boundaries around what's possible during any interaction.

PILLAR 2 Resources: Static Information Repository

Resources encompass all the static information an agent needs to function effectively. This includes structured data like databases, documents, and reference materials that don't change frequently. Resources provide agents with the contextual knowledge they need, without requiring real-time database queries for every piece of static information.

PILLAR 3 Prompts: Contextual Communication Guidelines

Tied to the available resources and tools, prompts define how agents should respond in specific situations. They're pre-written response templates that ensure consistent, appropriate communication based on the context of the interaction.

Security Through Structure: How MCP Protects Data

One of MCP's most significant advantages is its approach to security, which operates on multiple levels to protect sensitive information and prevent unauthorized access.

- **Hallucination Mitigation:** Traditional agent implementations often gave AI systems direct database access, creating opportunities for hallucinations when agents generated plausible-sounding but incorrect information. MCP addresses this by normalizing data exchange and reducing ambiguity.

  For example, when an MCP server receives a specific date from an agent, like "September 26, 2025," (rather than sharing the numbers in a different order, such as 26-09-25), there's little to no room for misinterpretation. The MCP can translate the data into its own specification for the agent, providing structured, verified responses, rather than constructing replies from raw database queries. This structured approach significantly reduces the likelihood of agents fabricating information (hallucinations).

  The protocol also limits agents to only the information explicitly provided by the tools they call. If a tool is designed to verify patient appointments, it returns only verification status — nothing more. This prevents agents from accessing or inferring additional data beyond their designated scope.

- **Data Containment and Access Control:** MCP creates strict boundaries around data access through its tool-based architecture. Agents can only access information through predefined tools, and each tool has specific parameters and return values. This approach prevents data spillage in several ways: limited scope, no direct database access and structured responses.

  If someone attempts to trick an agent into providing unauthorized information — like requesting a patient's social security number — the agent simply has no tool capable of retrieving that data. The response would be: "I don't have the capability to access that information. Would you like me to forward you to a human to answer that?"

- **Preventing Jailbreaking Attempts:** Jailbreaking occurs when users try to manipulate agents into providing information or performing actions they shouldn't. Classic examples include convincing an AI that harmful requests are actually for fictional purposes or creative projects. MCP's architecture makes jailbreaking significantly more difficult because agents physically cannot access information beyond their tool capabilities.

  Even if an agent hallucinates and generates a fake social security number or medical record number, that information isn't sourced from actual patient data — it's purely fabricated and can be detected and flagged by monitoring systems, like Judge LLMs.

The Future of Agent-System Communication

Model Context Protocol represents a fundamental shift in how we architect AI agent interactions. By providing structured, secure communication channels, MCP enables more sophisticated agent capabilities while maintaining strict security boundaries.

Many tech companies, including Artera, are already implementing MCP servers to integrate agent interactions with their platforms. This growing adoption suggests that MCP is on track to become a standard protocol across the tech industry, similar to how REST APIs became ubiquitous for web services.

While MCP shows great promise, we're prioritizing security as the protocol continues to mature. For example, our MCP server operates within a controlled environment, accessible only to authorized agents, rather than being publicly available on the internet.

As the protocol matures, we anticipate enhanced security standards, broader industry adoption, and more sophisticated toolsets that enable agents to handle increasingly complex workflows. I believe that organizations — such as Artera — which adopt MCP early are well-positioned to leverage these advances in agentic AI while maintaining robust security practices.

Agentic AI Healthcare Integration: How to Choose the Right Partner

AUTHORS:

Cassie Pena
Senior Director,
Product Management,
Artera

Simon Williams
Manager,
Integration Engineering,
Artera

Organizations are quickly adopting AI agents for healthcare to streamline operations, reduce staff workload and enhance the patient experience. But here's the thing: the success of these tools comes down to one big factor: how well they integrate with your existing systems. Even the smartest AI is only as good as its ability to connect to your data and workflows. Without seamless integration, you risk expensive technology that doesn't fully deliver.

The tricky part? Every vendor claims to offer "deep EHR integration" or "seamless connectivity." For healthcare IT leaders, the challenge is cutting through the buzzwords to figure out which partner can deliver real results. Let's break down what to look for when assessing an AI agent partner based on actual integration depth and scope.

Integration Capabilities Impact AI Agent Success

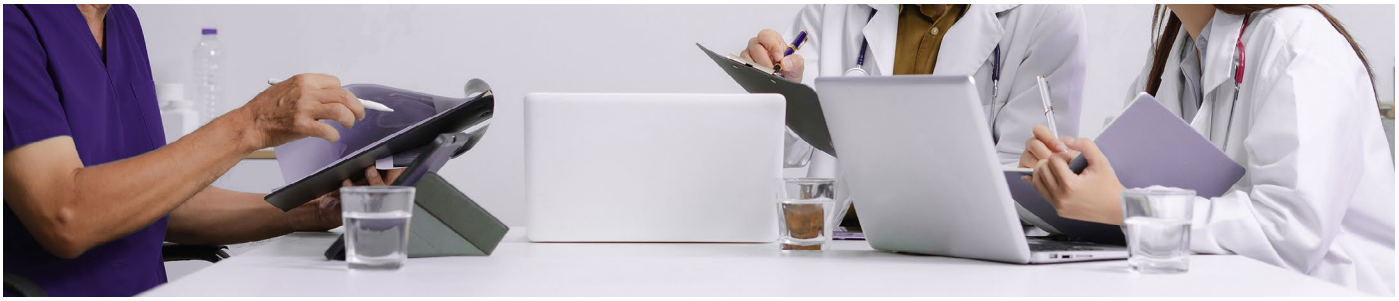
AI agents need access to the right data to truly be effective. Think EHRs, scheduling tools, and billing platforms: these are the core systems running patient care. Real-time data access allows AI agents to make smart decisions based on current patient information, schedules, and protocols. They also need to communicate back, updating records, booking appointments, and triggering workflows.

If the integration isn't solid, you're left with fragmented data and manual workarounds, which defeats the purpose of automation. To get the most out of AI, you need a partner who can connect all the dots seamlessly.

Evaluating Integration Scope and Diversity

Health system leaders must thoroughly assess and understand the integration capabilities available from a potential agentic AI partner. When evaluating vendors, healthcare organizations should look beyond broad integration claims and marketing to truly grasp the depth of their offerings.

Effective integration demands careful consideration of several critical dimensions, including experience and distribution, the breadth and diversity of EHR integration types and overarching ecosystem strategies.



Below are several topics and questions to consider when evaluating an agentic AI vendor's integration capabilities across certain dimensions:

DIMENSION	SIGNIFICANCE	POTENTIAL QUESTIONS TO ASK	THE ARTERA DIFFERENCE
Customer Volume & EHR Integration Diversity	Many vendors can showcase one or two basic integrations, but scaling AI agents across healthcare teams and environments demands proven expertise with multiple data sources. It's important to seek out vendors who can demonstrate experience with multiple customers using your EHR platform, the ability to handle diverse data feed types beyond basic scheduling, and production workflows that actively process patient interactions.	"How many customers have you successfully integrated with our specific EHR system, and across how many different data feeds / sources?"	At Artera, we maintain a comprehensive internal repository of EHR data feeds and integration types, which enables efficient knowledge transfer across teams and allows us to quickly reference EHR integration setups for customer support and shared learning.
Diverse Integration Methods	EHR integration isn't a one-size-fits-all solution. It's essential to partner with a vendor who has a thorough understanding of various integration methods and can guide your team toward the approach that best aligns with your needs. Choosing a vendor limited to a single integration method could restrict your options and hinder your success.	"Do you support multiple integration pathways, and can you give examples of when each was used successfully? How do you determine the right approach for different data types?"	At Artera, our integrations extend across multiple data feeds and API endpoints. Our in-house integration engineering team supports the full spectrum of healthcare data exchange methods, including: • HL7 messaging for real-time clinical data exchange • Flat file for structured files like CSV, TXT or TSV • FHIR APIs for modern, standards-based connectivity • SFTP file transfers for batch data processing • Custom development for unique organizational requirements
Workflow Support	Beyond a strong ecosystem, assess the complexity of workflows the vendor can handle, and their ability to support diverse teams and functions across an entire health system. Does the vendor's solution only assist with scheduling (i.e., just in the call center), or can it manage a broader range of healthcare workflows across different teams?	"What use cases do you support along the patient journey? Do you solve for one workflow, or can you expand to more complex use cases?"	At Artera, we work with customers across various areas and departments, including patient access and intake, value-based care, and chronic condition management teams. Our flexible use cases and workflow solutions allow us to scale and adapt to meet the unique needs of your organization. While some AI agent vendors focus on single-use cases like reducing call volume, Artera's integration engine supports comprehensive patient journey workflows like post-discharge follow-up, referral management and more.
Ecosystems & Partnerships	An AI agent's value multiplies when it can interact with your broader tech ecosystem. Vendors with established marketplace partnerships can unlock new workflows faster than those building relationships from scratch.	"What existing partnerships/contracts do you have in place? How quickly can you enable integrations with our other technology vendors?	Beyond EHR integration, Artera's Marketplace provides pre-built connections to 50+ digital health vendors across the care continuum. These established partnerships enable AI agents to access data and trigger actions across pop health, rev cycle management and more.

What Questions Should You Ask When Evaluating Agentic AI Partners for Healthcare?

Here's a quick list of questions to guide your conversations with potential agentic AI partners:

Technical Capabilities	• How many EHR systems have you integrated with in production? • What's the average number of data feeds you connect to per EHR implementation? • Do you have in-house integration engineering, or do you rely on third-party bridge vendors? • Which integration methods do you support (HL7, FHIR, API, SFTP)?.
Operational Experience	• How many healthcare customers are you currently supporting? • What workflows can you support across the entire patient journey? • How do you handle ongoing integration maintenance and updates? • What level of IT resource commitment do you require from our team?
Partnership Network	• Which third-party healthcare vendors do you currently integrate with? • Do you have existing partnership agreements in place? If not, how long would it take to get contracts in place? • Do you have the infrastructure and platform set up to support a true marketplace?

What Are the Hidden Costs of Poor Integration?

Choosing an AI agent vendor with insufficient integration depth can result in serious downsides:

First, custom development expenses quickly accumulate when vendors lack pre-built connectors for your existing systems. Second, operational overhead can increase significantly, as staff must manually transfer data between disconnected platforms, wasting valuable time and resources. Third, scalability limitations emerge, as vendors may struggle to adapt their solutions to your organization's evolving growth and changing needs.

Perhaps most importantly, weak integration prevents AI agents from accessing the full context required for smart decision-making. This directly reduces their effectiveness and hinders user adoption, undermining the very purpose of implementing AI.

Artera's Integration Advantage

With over 10 years of healthcare-specific integration experience, Artera brings unmatched expertise to the AI agent space. We've supported 1,000+ healthcare organizations with EHR integrations across a variety of workflows, giving us the knowledge and tools to deliver robust, scalable solutions.

Making the Right Choice for Long-Term Success

As you evaluate potential partners, ask yourself one critical question: "Can this vendor truly support my organization's long-term success?" If the answer is yes, you're on the right track.

Agentic AI for Healthcare: Build, Buy or Partner

AUTHOR:
Guillaume de Zwirek
CEO and Co-Founder, Artera

As a health system executive, you're likely at a crossroads that could define your company's competitive advantage for the next decade. Agentic AI is emerging as the next breakthrough, making the real question not if you should embrace it, but how to integrate it effectively.

While tools like ambient AI scribes have made AI a common term in healthcare, the next frontier is automating high-volume operational tasks. To get there, you have three options: build your own solutions, buy off-the-shelf products, or partner with a specialized vendor.

Understanding Agentic AI

Agentic AI represents a sophisticated convergence of multiple technologies working together to create human-like interactions that can complete complex tasks autonomously. Unlike simple chatbots or single-function AI tools, these systems integrate multiple technologies, including large language models, AI-generated voice capabilities, speech transcription and new standards like MCP for connecting AI systems to software.

Here are some real-world healthcare applications of agentic AI:

- Appointment scheduling
- Prescription management
- Password resets
- Insurance collection
- Bill processing

Artera data reveals that a substantial 45 percent of call center inquiries revolve around appointment verification, cancellations, and rescheduling. These are prime candidates for AI automation.

These AI agents work around the clock, 24/7, delivering consistent service and taking care of routine tasks that used to require human staff. The technology has matured to the point where it can deliver human-like experiences that often outperform traditional call center performance in speed, accuracy and experience.

Build vs. Buy vs. Partner
AI Tech Stack: "The Metal"

Before we dive into the three approaches, let's take a moment to break down the key layers needed to build an AI Agent. Think of these layers as building blocks, often provided by different vendors and serving various needs. I like to refer to this AI tech stack as "the metal."

- **Large Language Models (LLMs):** Understands & generates human-like text; there are many model providers, and the best depends on the use-case and modality (e.g. Gemini Flash 2.5 or OpenAI Realtime are the lowest latency as of this writing).
- **Orchestration:** Foundation layer that simplifies the process of integrating large language models with tools, while providing base observability and scalability for your agents
- **Real-Time Voice:** Includes two core services, often provided by distinct vendors depending on the use case. Speech-to-text (STT) transcribes audio into text, text-to-speech (TTS) turns LLM text output into spoken word, most typically using an "AI-generated voice" that is lifelike and infinitely customizable.
- **Telephony:** The core gateway connection into all the telecommunications and email (SMTP, POP3, IMAP) providers. For the most versatile agent, you will want to consider voice (PBX or SIP/V -OIP), SMS, RCS, and email.
- **Tools:** In healthcare, the "skills" you want your AI agent to fulfill (e.g. scheduling, prescription refill, insurance update) typically exist in the EHR/PM and are accessed via FHIR, HL7, SFTP or custom web services. As of this writing, best practice for leveraging tools involves creating a "translation" layer using an open standard like Model Context Protocol (MCP).
- **Evaluation:** A high-performing AI agent requires significant prompt engineering, wherein the instructions for the LLM are tweaked to reliably achieve the desired outcome. Evaluation

Agentic AI for Healthcare: Build, Buy or Partner? *continued*

Building In-House AI Agents

The first approach, building an AI Agent in-house, demands substantial technical expertise and financial investment. This involves engaging with the lowest-level infrastructure and meticulously integrating the various components listed above. By carefully stitching these together, one could construct a bespoke solution.

In this case, however, constructing a bespoke solution means you must work and maintain direct relationships with core AI Infrastructure providers to find solutions tailored to your needs, navigating a complex landscape and purchasing from each of them individually.

ADVANTAGES	DISADVANTAGES
• Complete control over tech stack and implementation • Customization aligned with specific organizational needs • Highest quality product (provided the right supporting team) • Lowest transaction costs (provided large enough volume)	• $3M minimum ongoing annualized investment (minimum of 2 Devops Engineers, 2 Machine Learning Engineers, 2 Senior Software Engineers, 1 Product Manager + hosting and software spend) • Need for specialized in-house talent across multiple languages, e.g. Python for ML, YAML, JSON and JavaScript for SWE, and being current on various AI-specific standards like A2A and MCP • Ongoing maintenance, all while staying up to speed with the latest technology advancements and agility to replace as new benchmarks emerge

Purchasing an Off-the-Shelf AI Agent (OEM Vendor)

Alternatively, health systems can work with a vendor that essentially rebrands a horizontal provider's technology as their own: a practice known as OEMing (Original Equipment Manufacturing). OEM vendors provide a wrapper around existing middleware platforms (e.g. Vapi.ai or Bland.ai) that facilitate the creation of AI Agents, but are not specifically designed for healthcare.

This approach can mean paying double the actual technology cost while limiting direct access to the latest innovations and updates of the core provider within the AI Tech Stack or "the metal."

Another more preferable option (assuming in-house expertise around AI integration standards and baseline development capacity) would be a healthcare organization acting as its own OEM and integrating directly with the middleware platforms mentioned above. This will reduce vendor costs, while increasing maintenance spend and providing more direct control over the end product.

ADVANTAGES	DISADVANTAGES
• Single point of contact/vendor to manage • Faster implementation timeline compared to building in-house • Lower upfront costs compared to building in-house	• Higher long-term costs due to markup on underlying technology • Limited customization capabilities • Dependence on third-party middleware providers • Less secure (additional entry point into your systems) • Less agility and control to replace as new benchmarks emerge

Partnering with AI Healthcare-Specialized Vendors

To steer clear of the above issues, I suggest health systems skip third-party vendors and work directly with those plugged directly into the metal.

Which leads us to the third option: contracting with a vendor focused exclusively on healthcare (healthcare-vertical providers), who have direct access to those in the AI tech stack. Specialized healthcare AI vendors typically invest millions or more in developing healthcare-specific solutions, understanding the unique requirements of healthcare interoperability, security, and compliance, while maintaining direct relationships with core AI infrastructure providers.

Essentially, these vendors orchestrate and bundle all of the AI tech stack components technology into a simple solution for healthcare providers, making it the fastest and most cost-effective way to deploy AI Agents.

ADVANTAGES	DISADVANTAGES
• Healthcare-specific expertise and compliance knowledge • Direct integration with core AI infrastructure • Proven real-world deployment experience • Shared investment in healthcare-relevant innovations • Faster time to value with lower risk • Many customization capabilities	• Dependence on partner's technology roadmap • Limited control over underlying technology decisions • Agentic ai for healthcare

The Core Tradeoff: Control Versus Speed

At its core, the build versus buy versus partner decision comes down to a tradeoff between control and speed.

Building in-house gives you full ownership of your tech stack, deeper integration into internal systems, and full control of sensitive data. However, it often requires longer development timelines, and significant upfront and ongoing investment

Buying off-the-shelf may offer faster deployment and lower initial costs (versus building), but can limit your ability to customize, restrict access to core data, and create long-term scalability issues.

The more strategic approach for many healthcare organizations is to partner with a vendor that understands the unique complexity of healthcare systems. By working with Artera, health systems accelerate time to value, leveraging years of specialized expertise, a compliant AI infrastructure, and deep integrations already built for real-world deployment. This approach reduces risk and cost while giving you the flexibility to scale with confidence.

Choosing between these paths is not always a binary decision. In many cases, the optimal strategy is a hybrid model that combines the customization benefits of in-house development with the speed, scalability, and stability of a proven AI partner like Artera.

Lessons Learned

We've spent years investing in AI, and with the rapid rise of Agentic AI, we've quickly embraced, configured and deployed it for many of our customers. Here's what we've learned along the way.

Technology Evolution Speed: The pace of change in Agentic AI is unprecedented. Technologies that required significant custom development just months ago are now available as standard features from major providers. This rapid evolution makes long-term technology investments particularly risky for organizations building in-house solutions.

Security and Compliance Complexity: Agentic AI systems are advancing at a rapid pace, requiring a fundamental shift in how we approach data security. Traditional static, point-in-time assessments are no longer sufficient. Instead, safeguarding these dynamic systems calls for continuous monitoring, robust multi-layered security frameworks, and a seamless integration of human oversight with AI-driven validation.

Real-World Deployment Value: The competitive advantage lies not in the underlying AI technology itself, but in healthcare-specific implementation experience. Understanding how to handle edge cases, manage patient interactions, and integrate with healthcare workflows requires extensive domain expertise.

Cost Structure Reality: The current pricing for AI solutions is unsustainable at scale. Similar to the dot-com boom, the current market is subsidized, with artificially low AI prices. When the inevitable market correction happens, many vendors won't survive. Therefore, organizations must partner with vendors that are not only financially stable enough to withstand the crash but also have sustainable business models for long-term success.

Beyond the Prompt: Designing Agentic AI for Healthcare Providers That's Safe, Scalable, and Compliant

AUTHORS:

Keith Dutton
Vice President, Engineering, Artera

Andrew Hwang
Engineering Manager, Machine Learning, Artera

When people think of AI agents, they often picture a powerful Large Language Model (LLM) that can handle tasks with just a simple "prompt." But building effective AI agents for healthcare is a whole different ballgame. These agents manage critical, multi-step workflows where the margin for error is virtually nonexistent. With incredibly high stakes, safety, accuracy, and stringent compliance are non-negotiable.

Consequently, developing production-ready, reliable and HIPAA-compliant AI agents for the healthcare industry not only demands advanced prompt engineering but a full ecosystem of solid backend tools, smart data pipelines, advanced analytics, and strict compliance frameworks. In this context, the prompt is really the foundation of a much bigger, highly connected system built to work seamlessly together.

What Specialized Prompt Engineering Is & Why It Matters

Language models are essentially rich repositories of information. Our goal with prompting them is to provide clear, precise instructions and guidance, ensuring they produce responses that align with our desired outcomes. It involves a full process of writing, refining and optimizing outputs.

Given the complexity of healthcare-related workflows, AI agents require explicit, highly structured instructions to successfully conduct natural conversations, all while adhering to strict safety and compliance guardrails. This is particularly critical in an MCP (Model Context Protocol) context, where we craft prompts to support and leverage these complex instructions.

Effective Prompt Engineering Techniques

Designing agentic AI for healthcare providers that's safe and compliant involves a disciplined, multi-layered approach that integrates both technical expertise and strategic design. Below are some core techniques essential to the prompt engineering process:

1. **Narrow Scope and Consistency to Create Reliable Healthcare Agents**

   For an agent to perform reliably in healthcare, it needs a clear job. For example, a scheduling agent should only focus on things like scheduling, rescheduling, or canceling appointments. This can include tasks such as verifying patient identity, checking provider availability, navigating location preferences, selecting appointment types, sending confirmations, managing waitlists, handling appointment reminders, and following up on missed or canceled visits.

When designers lay out exactly what an agent can and can't do, it keeps the conversation on track. An overly broad prompt often yields unhelpful results from the agent.

2. **Safety Guardrails to Prevent Hallucinations**

Prompts must include explicit "do/don't" instructions to enforce safety. For example, an agent might be told, "You are not a doctor; do not provide medical advice." These constraints prevent

agents from making clinical judgments, offering diagnoses, or answering questions that should be handled by licensed professionals. Additional guardrails may include restrictions around accessing or referencing sensitive data, such as insurance information, prescription history, or protected health details unless verified through appropriate tools.

Agentic prompts within the healthcare space are also designed to ensure agents handle ambiguous responses appropriately. If a patient answers a yes-or-no question with "maybe," the agent knows to re-ask the question until it receives a valid answer, rather than making assumptions. In high-stakes workflows, such as confirming surgical prep or managing medication instructions, these safety protocols ensure the agent stays within approved parameters, escalating to human staff when needed.

3. Modular and Scalable Design

Writing a new, complex prompt for every customer or use case is inefficient. Instead, adopting a modular template system streamlines the process. A foundational "healthcare agent" template can include universal safety guardrails and ethical protocols, while a secondary "use case" template customizes the agent for specific workflows, such as scheduling or prescription refills. This approach ensures consistency while allowing for easy specialization.

Resist the urge to over-engineer agent prompts for a quick fix, as some vendors may throw everything into an agent prompt in service of quick implementation. While this might seem efficient for a fast go-live, it's brittle and introduces risk. Whereas thoughtfully designed, intent-based MCP (Model Context Protocol) tools can increase performance, reduce the risk of hallucination and improve scalability.

4. Iterative and Flexible Prompts

Prompts must be designed for continuous refinement. A rigid or overly detailed prompt can lead to conflicts or unpredictable behavior. Modular, flexible prompts allow teams to quickly test and modify specific sections as needed without a complete rewrite. This iterative approach enables rapid improvements based on real-world feedback.

Measuring, Testing, and Improving AI Workflows

Testing and evaluation are critical to building reliable prompts. The process often begins by breaking down workflows from top to bottom into individual components and testing them in isolation. In simple terms: we have a goal of what the agent should be able to do from point A to point B, and so, how do we get it to point B?

Once these components are refined, end-to-end tests ensure they work together seamlessly.

For example, for a scheduling agent, you would break down the process into multiple pieces, or checkpoints: verifying patient information, identifying why the patient is calling in, recognizing which providers the patient can see, confirming eligibility, etc. In order to create a unified experience, we would need to make sure each step works in an isolated fashion before stitching them together to successfully book the appointment.

How Tools Enhance the Effectiveness of AI Agents

When we initially started building agents, we relied heavily on prompts to guide them. But we quickly learned that, instead, giving them the right tools is what really levels up the agent's capabilities, helping it explicitly understand when and how to perform actions within the given context.

Tools enable agents to understand what actions to take in a given context without overloading prompts with excessive instructions. By abstracting actions into the form of tools, the process becomes less error-prone, as the agent can choose from a predefined set of tools based on the situation.

We're moving towards tools like MCP handling more of the information, acting as communication nodes for the agent to complete workflows. This shift will continue as language models get better and faster, allowing us to integrate improved solutions.

Prompts and Tools Must Work Together

As AI keeps evolving, prompts are getting shorter as models get smarter and tools become more powerful. We're already seeing improvements in how AI "thinks through" complex responses. Multi-agent systems are also on the rise, with specialized agents handling tasks like patient verification or scheduling appointments. This modular setup makes them faster, safer, and easier to build.

In the future, better security and compliance will let agents take on bigger jobs, like processing payments or other high-trust tasks. But the key to success stays the same: combining a specialized, compliant prompt foundation with a solid system of tools, metrics, and constant improvement. Prompt engineering is important, but it's just one piece of the puzzle for building safe, reliable AI agents for healthcare.

Measuring What Matters: Performance Metrics for Voice AI Agents in Healthcare

AUTHOR:
Zach O'Bea
Principal Product Manager, Artera

The rise of agentic AI is super exciting, but let's be honest — it can also feel pretty overwhelming. This is uncharted territory for all of us. Large Language Models (LLMs) often seem like a "black box," and the array of techniques available to monitor them only makes things more complicated.

The rise of agentic AI is super exciting, but let's be honest — it can also feel pretty overwhelming. This is uncharted territory for all of us. Large Language Models (LLMs) often seem like a "black box," and the array of techniques available to monitor them only makes things more complicated.

As Principal Product Manager at Artera, my goal is to make this whole process less intimidating and more transparent. This is why I want to illustrate how Artera evaluates the performance of these non-deterministic, agentic systems, and what metrics we look for when measuring success.

By sharing what's happening behind the scenes, we hope to inspire curiosity and encourage our customers to ask the tough questions. After all, the real "winners" in the AI space will be the ones who learn to be savvy, informed users of this technology.

Being a savvy user starts with understanding performance. Unlike traditional software, where success is black-and-white, Agentic AI lives in the gray areas of human conversation. This requires a framework that moves beyond technical uptime to measure the quality and empathy of the patient experience.

Let's break down the metrics we focus on below.

Speed Matters: Measuring Latency in Voice AI Healthcare Agents

In voice-based AI, silence creates friction. Just a few seconds of dead air can cause patients to hang up or lose trust. That's why speed is such a big deal for us.

1. Time to First Token (TTFT)

This measures how quickly the AI responds after a patient says "Hello." Think of it as the AI's "hello back" moment. Similar to typing into a blank ChatGPT window, the first response tends to be the slowest due to a cold start— when the model loads its instructions and processes the initial request before replying. At Artera, we strive to keep initial response time latency to no more than 500 milliseconds (ms), as anything longer can feel awkward or even lead to patient frustration.

2. Average Turn Latency

Once the conversation is rolling, we look at "turn latency." This measures how long it takes for the AI to reply after the patient finishes speaking. If this suddenly gets slower, it usually means there's a system issue, like a sluggish API or a tricky query. We track these hiccups closely and get real-time alerts so we can jump in and potentially fix things quickly, keeping conversations smooth and frustration-free.

Quality of Life (QoL) Metrics: AI Workflow Reliability & Task Success Metrics

Speed is important, but it means little if the AI fails to get the job done. This is where "quality of life" metrics become crucial, as they're specifically designed to measure success within the context of a particular workflow. By asking questions like, "Is the AI completing tasks accurately, or is it creating errors?" we can gather the data needed to fine-tune our systems.

3. Patient Identification Success Rate

Let's look at a scheduling workflow, which is one of our most in-demand

use cases. To get a patient on the books, the AI must know exactly who it's talking to. It'll ask for a name and date of birth, then double-check those details against the EMR. We track how often the AI nails this step — basically its "win rate." When this goes smoothly, the rest of the call ideally falls into place. But if that success rate starts to dip? That's a red flag that something in the initial verification step needs to be investigated.

4. Tool Success Rate

We also keep a close eye on the "health" of the dozens of various tools baked into each workflow. By monitoring every success and failure, we can jump on issues the moment they happen, like if an agent suddenly can't book an appointment. If things go sideways, we know immediately that an investigation is required to understand if it's an EHR outage, a weird formatting glitch, or some other technical hiccup.

Patient Experience Metrics

Even if the AI is fast and our tools are working correctly, we want to have more qualitative analysis of the experience of a patient while they're on a call with our agent. Measuring this is tricky but essential … that's why we include a metric in our "LLM as a judge" that's all about patient experience.

5. Patient Experience Score

To gauge the patient experience score, we use an internal agent to review patient conversations and give them a score from 1 to 3:

- **3 (Excellent):** The interaction with a patient was smooth, easy, and completed without issues.
- **2 (Good):** The task was done, but there were minor hiccups, like asking the patient to repeat themselves.
- **1 (Poor):** The call went off the rails — confusion, frustration, or the need to escalate to a human.

Our goal? Maximize "3" scores and drive "1" scores to zero. This gives us actionable insights into how patients feel about their interactions and helps us fine-tune the AI.

Other Metrics We're Tracking (and Iterating On)

- **Average Length of Calls** (in minutes): Sometimes a long call is legitimate, sometimes it's a sign of an underlying issue
- **Count of Agent Conversations** (number of inbound / outbound conversations): important for customers to know how many calls their agent is handling each day
- **Call Outcomes** Determines whether a session was "successful" — this is currently defined as those where the workflow was completed autonomously without human intervention. We're continuing to refine this, as there are nuances to certain handoffs.
- **Workflow adherence** This is how closely an agent follows a given set of instructions. Using an LLM as judge, this metric determines whether the agent deviated from the specified workflow and to what extent.
- **Handoff Reason Analysis** We provide a detailed examination of why agents escalate interactions to human representatives, identifying the specific factors that necessitate a handoff.

In the future, we plan to enhance our analytics capabilities by introducing session sentiment analysis and allowing for different agent use cases to have their own workflow-specific rubrics and metrics. Soon, our customers will be able to review both audio and text files from conversations to identify tone of voice and patient inflection, unlocking a new level of conversational intelligence.

Turning Metrics Into Action: How AI Performance Data Drives Better Healthcare Outcomes

Evaluating these metrics is the key to making things better. When something goes wrong, they help us spot and fix it fast.

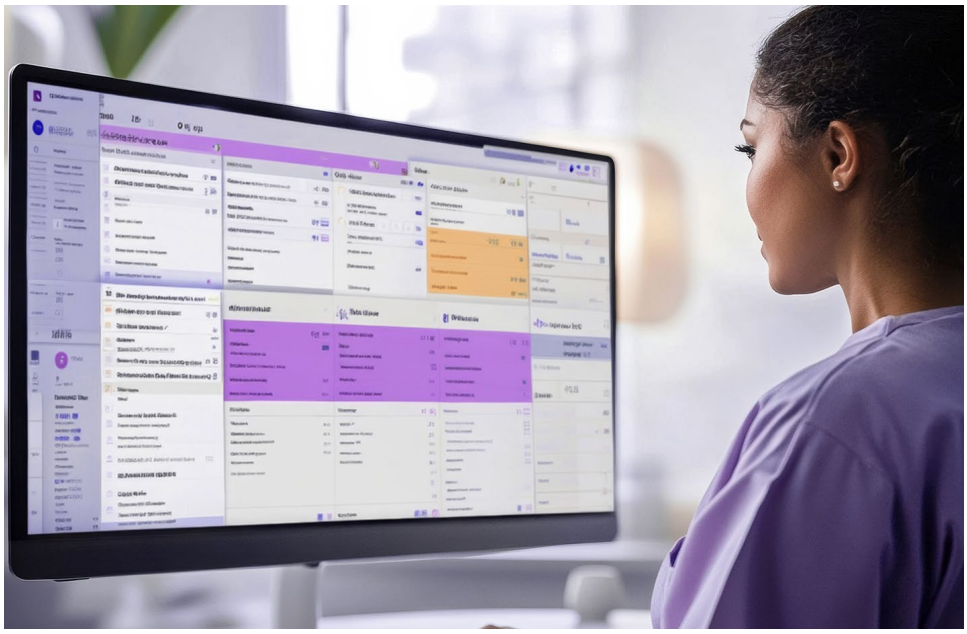
By analyzing agent-patient conversations with our LLM as judge, we gain both quantitative and qualitative insights into our agents' performance. This allows us to pinpoint specific areas for improvement. Manually reviewing every transcript is impractical at scale, so these techniques are essential for efficiently detecting agent hallucinations and identifying trends in technical failures or inconsistencies. Most importantly, this process provides our customers with the confidence that our agents are performing as promised.

If you're a healthcare leader thinking about Agentic AI, here's my advice: don't stop at the demo. Ask for the data. Look into how performance is tracked and monitored. With the right metrics, you can ensure that your AI agents truly deliver and transform the patient experience.

3 Critical Factors for Building a Scalable Digital Workforce

AUTHOR:
Zach Wood
Chief Product & Strategy Officer, Artera

The "digital workforce" is no longer a future concept; it's here. And, with the growing demand to adopt AI solutions that can ease administrative burdens and improve patient access, the challenge isn't whether to act — it's knowing where to start and how to build something that will deliver measurable, lasting impact — not just today, but three, five, even ten years from now.



While it's tempting to choose a solution that addresses your most immediate pain point, a scalable digital workforce can't be assembled from disconnected tools. It requires a strategic approach that governs how digital agents interact with patients, staff, and one another.

Too many vendors offer band-aid solutions that fix one problem while creating new ones: fragmented patient experiences, siloed data, and systems that don't talk to each other. The smarter path is investing in a partner who will help you craft a unified digital workforce designed to scale with your organization.

Here are three critical factors to consider to get started:

1. Ensure Humans and AI Agents Can Work Together

Even the most advanced digital workforce won't automate everything. Complex, sensitive situations will always need a human touch, and the success of your strategy depends on how well digital agents can collaborate with staff.

The goal is the "warm handoff" — a digital agent briefs staff before a call, or summarizes context to the patient before bringing a human in. The result: patients don't repeat themselves, and staff step in as specialists with context already loaded, not triaging from zero.

At Artera, we believe the future isn't just automation — it's the interplay between humans and agents. Our platform is purpose-built for a human-in-the-loop approach, ensuring staff are intelligently pulled in exactly when they're needed. If that resonates, let it shape how you evaluate vendors: ask how a platform handles escalations from AI to live staff, and what context carries over in that handoff.

2. Prioritize Connected Intelligence

For a digital workforce to be truly effective, it must possess "connected intelligence" — the ability to understand the entire history of a relationship, not just a single instance.

This requires a "switchboard" architecture that manages session context across the enterprise; it ensures that when a patient texts in "Wednesday at 2 p.m. works," the agent doesn't have to guess the intent. Instead, the system pins that new message to the broader history — whether it's a referral notification from yesterday or a follow-up from an hour ago, it acts with the same institutional memory as a dedicated care team.

Memory is what turns a fragmented chat into a seamless, concierge experience where the patient feels known — moving the needle from 'automation' to 'recognition.' When you evaluate solutions, look for this ability to hold context over time and across channels. A vendor that only sees the message in front of them is solving for a single task. A partner that sees the entire conversation history is a strategic asset dedicated to improving the patient experience.

We believe healthcare should feel human, even when it's powered by tech. Artera provides the connective tissue between backend agents and frontline staff, ensuring every touchpoint is transparent, personalized, and coherent, reducing friction across the journey.

3. Choose a Trusted Partner, Not Just a Product

Where you start matters: the foundation you build on, the vendor you choose, and the architecture you commit to will define everything that follows. And in a market full of vendors who will say yes to whatever you ask, that distinction is easy to miss.

A true long-term partner brings two things a band-aid solution can't: a track record of delivering on their roadmap, and a strong opinion on how to get you there. Rather than promising everything at once, they'll help you phase your approach, starting with high-value, lower-complexity use cases like managing appointments, then expanding to scheduling new patients, and eventually handling full registration workflows. Each phase builds on the last, with the people, processes, and technology to support it.

Point solutions are appealing because they're quick to deploy. But when your needs grow (which they will), you'll find yourself stitching together solutions that don't talk to each other, with no clear path forward. A long-term partner, on the other hand, grows with you.

Building for the Long Term

Healthcare is full of AI promises right now. Isolated tools might solve immediate problems, but they create fragmented experiences that are challenging to build on down the line.

A true digital workforce is an extension of your organization, built to collaborate effectively with humans, automate complex workflows with personalization, and scale with every new challenge you face. Prioritize these factors now, and your AI investment will deliver value today and in the future.

Bringing Administrative AI Agents into Your Healthcare Organization: The Foundations (Part 1)

AUTHOR:
Jessica Oveys
VP of Product Management, Artera

The healthcare industry has operated under one non-negotiable principle for decades: it has to be perfect. Lives are on the line. Regulations are strict. Mistakes aren't just costly — they can be catastrophic.

But something is shifting. AI is making healthcare faster — not just at answering patient calls or scheduling appointments, but at identifying problems, testing solutions, and releasing improvements. The failure-to-fix cycle that once took months now takes days, sometimes hours. That changes everything.

For providers considering administrative AI agents: don't be afraid. You don't need perfection on day one. You need a problem, a small team, a willingness to iterate — and a partner who sees around corners with you.

Start Here: Trade Traditional SaaS for a Dynamic AI Services Model

The biggest shift in your operations isn't technical — it's letting go of the rigid healthcare IT implementation mindset. Historically, deploying new technology meant a grueling sprint toward a single "Day One" launch, with absolute perfection expected upfront, because changing a workflow later meant waiting until an annual software enhancement cycle, complex updates to printed guides and extended user trainings.

AI agents break that paradigm. To move with real agility, healthcare organizations have to adapt past hands-off SaaS contracts and toward vendors that offer an AI Services Model, like Artera.

Instead of looking for a single solution to a single problem, organizations have to start focusing on finding a holistic AI partner: human builders with combined healthcare and AI expertise who work directly with your organization to solve its unique challenges — at the speed of software, with custom solutions.

That partnership is the engine that lets you trade "upfront perfection" for continuous, rapid progress. Rather than delaying a launch for six months to anticipate every patient scenario, you deploy a secure, compliant baseline workflow. From there, your AI services partner monitors live patient interactions, sees what's actually happening, and iterates on the fly. Optimization work that used to dictate a 12-month roadmap gets executed in days or weeks, not quarters.

This is how you leverage AI in this new agentic world. Everything below is the foundation that makes it work.

Foundation 1: Identify the Operational Problems Worth Solving

Before diving into complex workflows and documentation, ask two foundational questions:

1. What's the biggest challenge our patients face when engaging with us?
2. What am I spending my time doing that's beneath my skill set?

These often lead to the same place: better patient engagement and freeing providers from low-value administrative tasks. Patients want to connect with their provider, not a burnt-out provider who spent the night documenting referrals.

Once you've identified a focus area, map the pain points. Where are patients getting stuck? Where are staff hitting bottlenecks? Is poor outbound communication creating new inbound problems?

If you aren't sure where to look first,

revisit your digital transformation roadmap from the last 2 - 3 years. Did you hit those goals? If not, start there. Those unmet objectives are often ideal entry points for AI — core patient needs like scheduling, paying, and accessing care haven't changed; they've just gotten more complicated.

But here's what I want you to hold in the back of your mind as you do this exercise: the problems you can name today are just the beginning. The most valuable problems AI will solve in your organization are ones you haven't identified yet — because you've never had a system capable of seeing them (more on that in a follow-up piece).

Foundation 2: Determine What's Prime for Agents

Not every organization is ready to dive 100% into AI Agents, so not every workflow is a good fit for AI agents — right away. Here's the quick litmus test for finding what is "agent-ready" for your org:

- **Repeatable:** The process happens frequently across the organization.
- **High touch:** Multiple people perform it regularly.
- **Low clinical risk:** Minimal chance of adverse patient outcomes.

If you're new to AI agents, start with workflows that check all three boxes. As you build confidence, you can tackle more complex, higher-stakes use cases, especially with a partner who can evolve with you.

Today, you no longer need a roadmap with multiple use cases mapped out for the next 3 years. You start with a core problem and let the agents bring you the unforeseen problems you don't know about. Your "roadmap" builds over time programmatically, quickly, and continuously.

Foundation 3: Assemble a Small, Thoughtful Team

This is the hardest habit to break. Large, cross-functional teams that meet weekly slow you down. Deploying AI agents calls for a small, agile group that deeply understands the business need and can communicate updates to the rest of the organization.

Don't exclude the skeptics. Include the doctor who doesn't love AI and the front-office staffer who swears they'll never use it. Their hesitation is valuable; you don't have to act on every concern, but those concerns reveal blind spots you'd otherwise miss.

Keep one perspective front and center: patient engagement AI is very different from clinical decision-making AI. Today's patients interact with AI daily, from booking travel to managing finances. They can handle occasional friction, and they don't expect absolute perfection, so your team shouldn't let the fear of it paralyze progress.

Foundation 4: Build and Continuously Refine Your Operational Documentation

Deploying an AI agent starts with a foundation of truth. AI agents don't guess your clinic's rules or protocols from the open internet. To stay safe and compliant, you train the agent on your internal documentation: operational guidelines, prep instructions, and standard operating procedures. Think of it as a secure, closed knowledge base — every answer the AI gives a patient is pulled directly from this internal playbook, which keeps it from hallucinating.

Start with your most experienced people: the ones who've been there 10, 15 years. Interview them. Shadow them. Capture what they do that they never wrote down: the personal touches, the reminders they give patients, the shortcuts they've developed.

But this isn't a one-time exercise. Staff will deviate from the original process and make adjustments, and you'll want the AI to know about them. That means committing to continuously refining your documentation.

Final Thoughts

If you take one thing from this: you don't have to be 100% ready — you just have to start.

Find the right partner to iterate alongside you. Start with patient pain points and administrative burdens, pick repeatable, high-touch, low-risk workflows, build a small team, and commit to maintaining your knowledge base. In the world of AI, the goal was never perfection before launch: it's progress.

But deploying well is just the beginning. The organizations that pull ahead aren't only the ones that launch AI agents successfully — they're the ones that understand what those agents are about to reveal, and stay ready to act on it. That's where this is really going, and it's the subject of Part 2: The Frontier.

Bringing Administrative AI Agents into Your Healthcare Organization: The Frontier (Part 2)

AUTHOR:
Jessica Oveys
VP of Product Management, Artera

In Part 1, we covered how to prepare for and deploy administrative AI agents: the foundations. But deploying well is just the starting line. The real shift is what comes next.

When we relaunched as artera.io, it wasn't a cosmetic change — it was a declaration. We're no longer a patient communications company that happens to use AI. We're an AI company that understands patient communication is the connective tissue of healthcare operations.

That distinction matters because the next wave of value in healthcare isn't going to come from making existing workflows faster. It's going to come from AI surfacing workflows that should exist but don't — problems no one has staffed for because no one knew they were problems.

If you've already started deploying administrative AI agents — identified your first use cases, built a small team, trained the agent on your documentation — you've done the hard, practical work. (If you haven't, start there first). Those foundations are table stakes.

This is about where they lead.

Where This Is Really Going

The organizations that win aren't just the ones that deploy AI agents well. They're the ones that understand what AI is about to reveal — and are open to making quick adjustments when it does.

AI will show you the calls that never came.

Today, if a patient doesn't call to reschedule, you assume they're coming. If they no-show, you note it and move on. But an AI system that understands your patient population — their communication patterns, their barriers, their history — can read the absence of engagement as a signal. The 68-year-old diabetic who always confirms but went silent this cycle. The post-surgical patient whose engagement pattern predicts a complication call in 72 hours.

These aren't clinical decisions. They're administrative intelligence — the operational equivalent of early warning systems that never existed, because no human team could monitor at that resolution. AI doesn't just answer the phone better. It notices what the phone never rang about.

AI will expose the cost of your workarounds.

Every healthcare organization has staff who've built elaborate manual processes to compensate for broken integrations, unclear scheduling rules, or EHR limitations. Those workarounds are invisible to leadership because they work — until the person who invented them leaves.

AI agents, by attempting to follow your documented processes exactly, will immediately fail wherever a human was silently papering over a gap. That failure isn't a bug — it's a diagnostic. Within weeks of deploying an agent, you'll have a map of every undocumented exception, every tribal-knowledge dependency, every process that only works because someone named Linda has been there since 2008. That map alone is worth the investment, before the agent handles a single patient call.

AI will redefine what "staffing" means.

We're not far from a fundamental rethinking of what "staffing" means for healthcare administration. Not replacement, but reconfiguration. Today, you staff a call center based

on call volume. You hire schedulers based on appointment complexity. You add billers based on denial rates.

In the near future, you won't staff based on volume. You'll staff based on the exception rate. AI handles the majority that follows the rules; your human team handles the outliers that require judgment, empathy, or creativity. And that outlier isn't static — every exception the AI escalates is a learning opportunity, so over time the exception rate compresses. Not to zero, but enough that the role of administrative staff shifts from processing to judgment. From clerk to analyst. From phone operator to patient advocate.

That's not a cost story. It's a dignity story. The best people in healthcare administration are overqualified for what they spend their days doing. AI doesn't take their jobs — it gives them back the jobs they were hired to do. And it doesn't depersonalize the patient experience; it means that when a patient is scared, lost, or struggling and needs a hands-on touch, someone is free to give it.

What We Haven't Solved Yet

Let me be transparent about what we're still building toward — because these open problems define the next era.

Multi-system orchestration.

Today's AI agents are strong within a single domain: scheduling, intake, FAQ. But patients don't experience their care in domains. They experience it as one continuous relationship that happens to cross your EHR, your billing system, your referral network, and three different phone trees. The agent that can hold context across those systems — that understands a scheduling question is actually a transportation barrier masquerading as a no-show pattern — doesn't fully exist yet. We're building toward it.

Cultural fluency at scale.

We can localize language. We can translate. But cultural fluency — understanding that a patient's silence means deference to a family decision-maker, not disengagement; that "I'll think about it" means no in one community and yes in another — is a layer of intelligence we're still learning to encode. AI trained on operational data reflects operational assumptions. The real question is whether AI can help us see those assumptions for the first time.

The trust gap between "it works" and "I trust it."

We've proven containment rates, reduced hold times, improved show rates. But there's a gap between a healthcare organization seeing results and trusting the system enough to expand its scope. Bridging that gap isn't a technology problem — it's a relationship problem. And it's why the services model matters more than the software.

The Window Is Closing

Here's what keeps me up at night: the healthcare organizations that start now won't just be ahead in 18 months. They'll be unreachable. Because AI compounds. Every patient interaction trains the next one. Every exception documented makes the system smarter. Every week of live data creates a moat that no amount of catch-up spending can bridge.

The gap between "started early" and "started late" isn't linear — it's exponential. The organizations deploying today aren't just solving today's problems. They're generating the institutional intelligence that will let them solve tomorrow's problems before their competitors have even named them.

You don't have to be perfect. But you do have to start. And the window where "starting" still means "early" is closing faster than this industry realizes.

About Artera

Artera is an agentic company strengthening how healthcare providers communicate and care for patients. As an agentic partner, we bring over a decade of healthcare experience to address urgent workflows from day one and build custom solutions as healthcare providers' needs evolve. Trusted by 1,000+ specialties, FQHCs, health systems, and federal agencies, Artera strengthens and enhances patient relationships across every interaction — from intake and scheduling to referral management, post-visit care, and more.

2B+
annual communications

200M+
patients engaged annually

11+
years experience

1,000
provider customers

class D certification

Artera's blog posts and press releases are for informational purposes only and are not legal advice. Artera assumes no responsibility for the accuracy, completeness, or timeliness of blogs and non-legally required press releases. Claims for damages arising from decisions based on this release are expressly disclaimed, to the extent permitted by law.

This dynamic series is designed to evolve and expand over time. Stay connected with the latest perspectives from Artera by visiting our website and **exploring the full collection.**

JUNE 2026

artera.io | sales@artera.io | 1-833-234-9355