MASTECH DIGITAL

Whitepaper

Agentic AI for Prior Authorization: From administrative burden to an Interoperable, Auditable operations fabric

Author - Shravanti Mitra
Industry Principal, Health Sciences

Executive Summary:

Prior Authorization (PA) has become one of the most visible operational bottlenecks in U.S. healthcare—driving care delays, avoidable denials, clinician abrasion, and increased admin cost. The near-term opportunity is not just automation, but a shift to agentic workflows: multiple specialized AI agents that retrieve evidence, interpret policy, reason clinically, and coordinate bi-directional exchange while maintaining auditability. This blueprint outlines how to move from rigid rules to adaptive intelligence.

Source: AMA 2024 Survey

93%	39	13+ hrs	31%	3–7 Days
Care Delays	PA Requests week/Physician	Weekly Admin Burden	PA Denial Rate	Decision Timeline

The Problem: A System at a Breaking Point

Current benchmarking reveals a fragmented, high-friction workflow that punishes every stakeholder. From the AMA's 2024 physician survey, the scale of operational friction is unsustainable. Physicians and staff are buried in chart mining, while payers struggle with incomplete submissions that trigger manual pends.

Why Traditional Automation hasn't Solved it

Many organizations have attempted to solve this with portals, forms, and RPA, yet the inefficiency persists. The failure stems from inherent limitations in legacy approaches:

Rules-Based Brittleness:

Rigid if-then logic breaks whenever policies change or edge cases arise, requiring constant maintenance.

Limited Clinical Reasoning:

RPA cannot understand medical necessity or justify why a treatment aligns with complex guidelines.

Siloed Systems:

Payer and provider workflows remain isolated with no real-time data sharing.

Fragmented Data:

Critical clinical evidence is scattered across structured EHR fields, PDF notes, labs, and imaging systems.

Opaque Decisions:

Black box denials provide no clear clinical rationale, leading to unnecessary appeals.

Prior Authorization Friction Flywheel (Today)

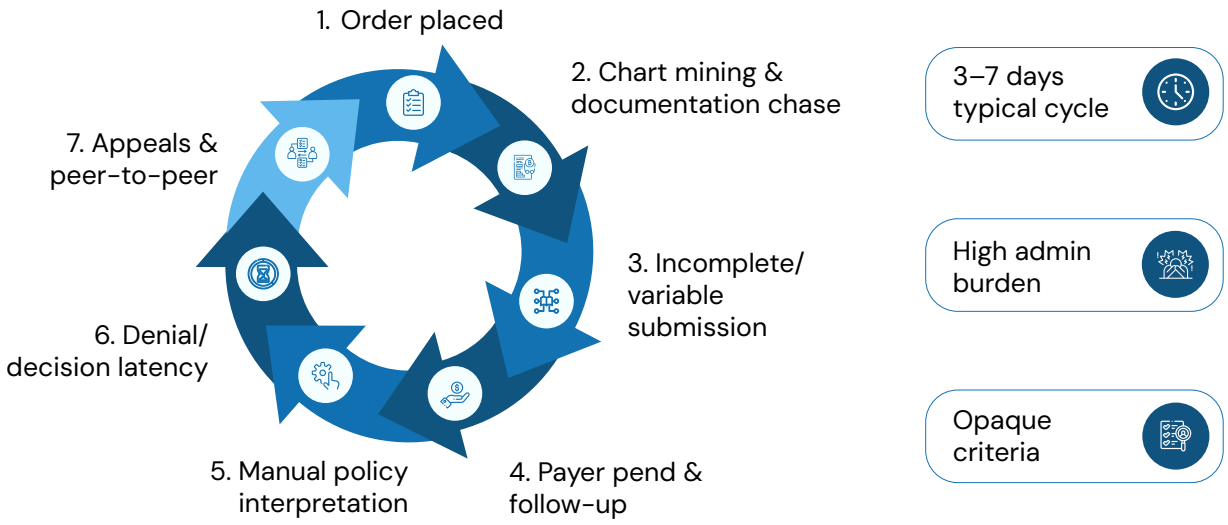

Figure 1 : Prior Authorization Friction Flywheel (Today)

The Shift: From Linear Processing to Orchestrated Multi-Agents

What Agentic Prior Auth Means in Operational Terms

Agentic PA is not merely a chatbot or generative text tool. It is an orchestrated ecosystem of specialized agents that decompose the complex PA process into autonomous but coordinated tasks. **Mastech Digital** defines this architecture through eight core roles:

01 **Intake Agent:** Validates request completeness and triggers workflows.

02 **Context Synthesis Agent:** Builds patient profile from EHR history/notes/workflows.

03 **Evidence Collection Agent:** Retrieves guidelines, labs, and imaging data.

04 **Policy Intelligence Agent:** Interprets PDF/text policies into logic.

05 **Reasoning Agent:** Applies clinical logic to determine necessity.

06 **Narrative Generation Agent:** Drafts evidence-backed clinical letters.

07 **Payer Comms Agent:** Manages API /EDI submissions & status tracking.

08 **Appeal/Optimization Agent:** Learns from outcomes to improve yield.

The New Workflow: Minutes for Routine, Hours for Complex

This approach transforms a serial, manual process into parallelized, intelligent operations. Routine cases are processed in minutes with high auto-approval rates, while complex cases are prepared with full dossiers for human review.

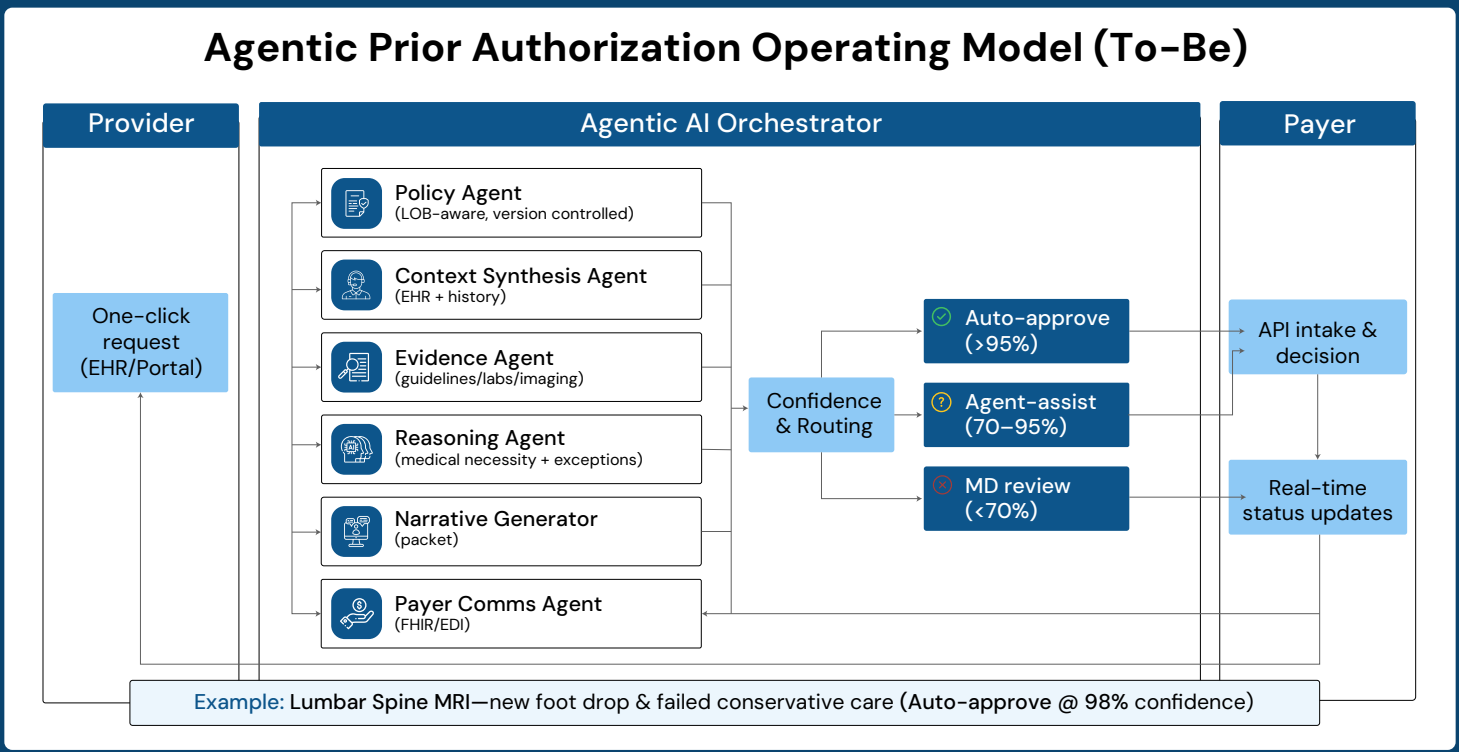

Figure 2: Agentic Prior Authorization Operating Model (To-Be) - Orchestrated intelligence with multi-agent collaboration

Clinical Vignette: Why Agentic AI succeeds where rules fail

Scenario: A 45-year-old patient requires a Lumbar Spine MRI . They have a new onset of foot drop (neurological deficit) but have completed only 4 weeks of physical therapy (PT).

Traditional Rules-Based Automation	Agentic AI Reasoning
Outcome: AUTO-DENY The rules engine checks the Conservative Therapy field. It sees 4 weeks is less than the required 6 weeks. It lacks the context to understand exceptions. **Result:** Denial, delay, physician appeal.	**Outcome: APPROVE** The Reasoning Agent identifies **foot drop** in the clinical notes. It cross-references the Policy Agent, which highlights an exception: Neurological deficit exempts PT requirement. **Result:** Approval with audit trail citing exception.

Why Now: The Interoperability Tailwind

The shift to Agentic Prior Authorization is not merely technological, it is increasingly regulatory and market-driven.

CMS Interoperability and Prior Authorization Final Rule (CMS-0057-F)

Published in January 2024, this landmark rule requires impacted payers (Medicare Advantage, Medicaid managed care, CHIP, and QHP issuers on FFEs) to implement several critical capabilities:

Prior Authorization API:

By January 1, 2027, payers must provide a FHIR-based API that communicates approval/denial decisions and specific denial reasons in real-time.

Decision Timeframes:

Starting January 1, 2026, payers must respond within 72 hours for expedited requests and 7 calendar days for standard requests (excluding drugs).

Denial Transparency:

All denials must include specific reasons, enabling providers to address gaps and resubmit efficiently.

Public Metrics Reporting:

Annual publication of PA volume, approval rates, and turnaround times creates accountability and competitive pressure.

Source: Centers for Medicare & Medicaid Services. (2024). CMS Interoperability and Prior Authorization Final Rule (CMS-0057-F).
https://www.cms.gov/newsroom/fact-sheets/cms-interoperability-and-prior-authorization-final-rule-cms-0057-f

Industry-led reform: Gold Carding and Selective Application

Beyond regulatory mandates, market forces are driving innovation. **Gold Carding** programs which exempt high-performing providers from PA requirements, are expanding rapidly. UnitedHealthcare announced in 2025 that it had added 1,800 providers to its Gold Card program, recognizing that trust-based models can co-exist with oversight for outliers.

Agentic AI enables dynamic Gold Carding: rather than static provider lists, systems can evaluate real-time performance (approval rates, outcomes data, guideline concordance) and automatically adjust PA requirements accordingly. This creates a virtuous cycle where high-quality care is rewarded with reduced administrative burden.

The Pharma Perspective: Hub Services and Access Programs

For pharmaceutical manufacturers and specialty pharmacies, Prior Authorization friction directly impacts patient access and medication adherence. Agentic system offers:

Hub service optimization:

Automating PA support for high-cost therapies (Oncology, Rare disease) to accelerate Time-to-Therapy.

Real-world evidence integration:

Pulling outcomes data from registries and EHRs to support off-label use justifications.

Patient assistance program coordination:

Seamlessly routing denied PAs to co-pay assistance or free drug programs based on eligibility.

Navigating the Medicare and Medicaid PA Labyrinth: How Agentic AI Addresses Program-Specific Complexity

While Commercial Payer Prior Authorization is challenging, Medicare and Medicaid present exponentially greater complexity due to program structure, state variation, and regulatory constraints. Organizations serving these populations must navigate a fragmented landscape where a single universal PA solution is insufficient.

The Medicare Advantage (MA) Challenge: Plan Diversity and CMS Oversight

Medicare Advantage plans operate under **dual accountability**: they must comply with CMS regulations while maintaining competitive plan designs. This creates unique PA friction points:

- **Supplemental benefit variation:** While traditional Medicare Part A/B services have standardized coverage rules, MA plans offer supplemental benefits (dental, vision, transportation) with plan-specific PA requirements. Agentic systems must maintain Benefit-Tier-Aware policy engines that distinguish between core Medicare coverage determinations and plan-specific supplemental authorizations.
- **Star Ratings pressure:** CMS Star Ratings penalize plans for appeals and member complaints related to PA. According to the 2024 AMA survey, 75% of physicians report that PA delays in MA plans have increased over the past five years. Agentic AI can reduce this friction by providing proactive coverage determination alerts at the point of care—warning providers before they order services that will require PA.

- **Part D (pharmacy) complexity:** Medicare Part D formularies require Step Therapy and Prior Authorization for high-cost medications, but tier placement and step-therapy sequences vary dramatically by plan. The Policy Intelligence Agent must query plan-specific formulary files and interpret exception pathways for situations like drug intolerance or contraindications.

Real-world Example: A patient with heart failure prescribed Entresto (sacubitril/valsartan) may face different Step-Therapy requirements across MA plans - some require ACE inhibitor failure first, others allow direct access for NYHA Class II-III patients. An Agentic system trained on multi-plan formularies can instantly determine which pathway applies and assemble the required evidence.

The Medicaid Maze: State-by-State Variation and Managed Care Fragmentation

Medicaid presents the most complex PA landscape in U.S. healthcare. Each state operates its own program with distinct coverage policies, and most states use Managed Care Organizations (MCOs) that add another layer of plan-specific rules. Key challenges include:

- **50+ state policy sets:** Medicaid fee-for-service programs in states like California, New York, and Texas have entirely different PA requirements for the same procedure. For example, Durable Medical Equipment (DME) like wheelchairs may require face-to-face physician assessments in some states but not others. Agentic systems must incorporate state-specific policy versioning and automatically route requests through the correct regulatory framework.
- **MCO variability within states:** In states with multiple Medicaid MCOs (e.g., Florida with 11 statewide plans), PA requirements for the same service can differ dramatically. The Evidence Collection Agent must not only retrieve clinical data but also map the member to the correct MCO and apply that plan's specific medical policy.
- **EPSDT mandates:** The Early and Periodic Screening, Diagnostic, and Treatment (EPSDT) benefit requires Medicaid to cover medically necessary services for children, even if the service isn't typically covered for adults. This exception logic is rarely well-documented in MCO medical policies. Agentic AI systems must include member age-aware exception handlers that automatically flag EPSDT eligibility and adjust approval pathways accordingly.
- **Retroactive eligibility complexity:** Unlike commercial insurance, Medicaid members can have retroactive eligibility up to 3 months prior to application. This means PA requests may be submitted for services already rendered, requiring the Intake Agent to distinguish between prospective and retrospective authorization scenarios and route them appropriately.

How Leading Organizations Are Addressing These Nuances

Forward-thinking payers and technology vendors are deploying several strategies to manage Medicare/Medicaid PA complexity:

- **Multi-tenant policy repositories:** Rather than hard-coding rules, organizations are building composable policy libraries where federal baseline rules (CMS NCDs/LCDs) are layered with state-specific Medicaid policies and plan-specific supplemental criteria. The Policy Intelligence Agent queries this hierarchy dynamically based on member eligibility metadata.
- **Real-time eligibility integration:** Agentic systems integrate with state Medicaid Management Information Systems (MMIS) and Medicare eligibility APIs to validate coverage status, benefit package, and retroactive eligibility before initiating PA workflows. This prevents processing requests for ineligible members or incorrect coverage periods.
- **Geospatial policy mapping:** Some advanced implementations use member ZIP code and provider location to automatically determine applicable state Medicaid rules and local coverage determinations (LCDs). This is particularly critical for DME suppliers operating across multiple states.
- **Appeal tracking and policy feedback loops:** Because Medicare Advantage and Medicaid MCOs face regulatory scrutiny of appeal overturn rates, leading organizations use the Learning & Optimization Agent to track which PA denials are overturned on appeal. When overturn rates exceed thresholds (e.g., >15% for a specific service/diagnosis combination), this signals that the underlying policy criteria may be too restrictive or incorrectly applied. The system flags these patterns for clinical policy review.

The Agentic Solution: Program-Aware, Adaptive Intelligence

Traditional PA automation fails in Medicare/Medicaid environments because it relies on static rule sets. Agentic AI succeeds through contextual program awareness:

01

The Intake Agent identifies the member's program (MA, Medicaid FFS, Medicaid MCO), benefit tier and state jurisdiction.

02

The Policy Intelligence Agent retrieves the correct policy hierarchy: federal baseline → state Medicaid policy → MCO-specific criteria → plan supplemental rules.

03

The Reasoning Agent applies program-specific exception logic (e.g., EPSDT for children, Medicare coverage gap exceptions).

04

The Payer Comms Agent formats the submission according to program-specific requirements—X12 278 for Medicare, state Medicaid portal formats, or MCO-specific FHIR APIs.

This architecture transforms an impossibly complex manual process into a manageable, auditable workflow. Rather than requiring staff to memorize 50 state Medicaid policies and hundreds of MCO medical policy documents, the agentic system becomes the institutional memory that ensures every PA request follows the correct pathway.

Source: American Medical Association. (2024). 2024 AMA prior authorization physician survey. https://www.ama-assn.org/system/files/prior-authorization-survey.pdf

Multi-Stakeholder Benefits: Quantifying the Value Proposition

Agentic prior authorization creates measurable value across the entire healthcare ecosystem. Unlike traditional automation that merely shifts bottlenecks, intelligent agent orchestration eliminates friction at its source.

Provider Impact: Reclaiming Clinical Time

According to the 2024 AMA survey, physicians report spending an average of 13 hours per week on prior authorization tasks, with practices completing approximately 39 PA requests per physician weekly. Agentic systems can reduce this burden through:

Automated evidence assembly:

The Context Synthesis and Evidence Collection agents eliminate manual chart mining, pulling structured and unstructured data directly from EHR systems.

Real-time decision transparency:

Providers receive instant notifications on approval likelihood before submitting requests, reducing unnecessary work.

Reduced peer-to-peer reviews:

With 65% of physicians reporting P2P review requirements, agentic systems that generate comprehensive clinical narratives can reduce these time-consuming interactions significantly.

Source: American Medical Association. (2024). 2024 AMA prior authorization physician survey.
https://www.ama-assn.org/system/files/prior-authorization-survey.pdf

Payer Impact: Operational Efficiency at Scale

The 2024 CAQH Index estimates that manual prior authorization costs approximately $3.41 per transaction for payers. Electronic automation can reduce this per transaction, with agentic AI offering further optimization through:

Higher first-pass approval rates:

By ensuring complete, policy-compliant submissions upfront, agentic systems can increase initial approval rates.

Reduced appeals volume:

Clear, evidence-based denials with specific rationale decrease inappropriate appeals.

Clinical reviewer optimization:

Routing only complex cases (10-20% of volume) to medical directors allows expensive clinical resources to focus on true medical necessity determinations.

Source: CAQH. (2024). 2024 CAQH Index Report.
https://www.caqh.org/insights/caqh-index-report

Patient Impact: Access and Outcomes

The most critical metric - patient outcomes will see the most dramatic improvement. The AMA survey found that 93% of physicians report PA-related care delays, with 82% citing treatment abandonment. Agentic systems address this through:

Accelerated decision timelines:

Reducing 3-7 day waits to hours or minutes for routine cases.

Reduced serious adverse events:

The AMA reports that 29% of physicians have seen PA lead to serious adverse events; faster processing directly impacts clinical safety.

Financial transparency:

Real-time determination of out-of-pocket costs eliminates surprise bills and reduces treatment abandonment due to cost uncertainty

The Blueprint: Architecture, Governance, and KPIs that matter

Reference Architecture: Build an Agentic Fabric, Not a Monolith

Successful implementation requires a multi-layer ecosystem rather than a rigid software suite.
The architecture consists of:

Experience Layer:
EHR Embedded Apps (SMART on FHIR) and Provider/Payer Portals.

Intelligence Layer:
The swarm of specialized agents (Reasoning, Evidence, Policy).

Data Layer:
Unified access to EHRs, PACS, Labs, and Clinical Guidelines.

Governance Layer:
The critical human-in-the-loop controls, audit logging, and compliance checks.

Agent Specialization and Orchestration

Each agent operates as an autonomous unit with defined responsibilities, tools, and decision boundaries. Consider the Reasoning Agent workflow for a biologic drug request:

1. **Retrieval:** The agent queries the Policy Intelligence Agent for current step-therapy requirements (e.g., "Must fail methotrexate for 3 months").
2. **Evidence Synthesis:** It cross-references medication history from the Context Synthesis Agent, identifying documented MTX trial with duration and adverse effects.
3. **Guideline Alignment:** The Evidence Collection Agent provides NCCN or ACR guidelines supporting biologic use in this clinical context.
4. **Exception Handling:** If step-therapy is incomplete but contraindications exist (e.g., liver toxicity from MTX), the agent applies exception logic rather than rigid denial.
5. **Confidence Scoring:** The system calculates a confidence score (0–100%) based on data completeness, policy match strength, and guideline alignment.

This multi-agent collaboration mirrors clinical decision-making far more accurately than rules-based systems, which would simply check: "MTX trial ≥ 3 months? No → DENY."

Continuous Learning and Model Optimization

A critical advantage of agentic systems is their ability to improve through feedback loops.
The **Learning & Optimization Agent** tracks:

- **Decision outcomes:** Which auto-approvals were correct vs. which required human override?
- **Appeal patterns:** What denial reasons consistently result in successful appeals, indicating policy misinterpretation or outdated criteria?
- **False positive/negative rates:** Are we over-approving inappropriate requests, or over-denying medically necessary care?

This creates a "**digital clinical mentor**" that becomes more accurate over time, similar to how resident physicians improve through case experience and attending feedback.

Operational KPIs That Matter

Success must be measured across all stakeholders to ensure the system is working holistically:

Provider KPIs:

Reduction in time spent per request, increase in first-pass approval rates, and improvement in staff productivity.

Payer KPIs:

Lower cost per authorization, higher auto-approval (touchless) rates, reduced appeal volumes, and higher clinical accuracy.

Patient KPIs:

Faster time-to-treatment, improved continuity of care, and greater transparency on out-of-pocket costs.

Governance and Safety: The Non-Negotiables

Deploying agentic AI in clinical workflows demands rigorous governance frameworks. The system must maintain HIPAA compliance at every layer, with comprehensive audit trails that capture agent reasoning paths, data sources consulted, and confidence scores for every decision. Bias detection mechanisms must continuously monitor for disparities in approval rates across demographic groups or geographic regions.

The **Human-in-the-Loop protocol** is non-negotiable - cases with confidence scores below defined thresholds (typically 70-80%) must be routed to clinical reviewers. This creates a safety net while allowing the system to continuously learn from expert corrections.

Integration with Value-Based Care and Risk-Adjustment Models

As healthcare transitions from fee-for-service to value-based reimbursement, prior authorization must evolve from a cost control mechanism to a quality and outcomes optimization tool. Agentic systems enable this shift through:

- **HCC and risk-adjustment awareness:** Medicare Advantage plans operate under Hierarchical Condition Category (HCC) risk-adjustment models where comprehensive diagnosis capture drives revenue. The Context Synthesis Agent can flag when a PA request contains diagnoses that improve HCC scores (e.g., chronic kidney disease Stage 4 vs. unspecified CKD), prompting providers to ensure proper documentation that benefits both clinical accuracy and plan financial viability.
- **Outcomes-based PA adjustments:** For Accountable Care Organizations (ACOs) and bundled payment models, PA criteria can shift based on episode-of-care outcomes. For example, if a provider's total knee replacement patients consistently achieve above-benchmark outcomes (lower readmissions, faster functional recovery), the system can automatically grant expedited approval or waive PA for subsequent cases - dynamic gold carding based on outcomes, not just volume.
- **Care gap identification:** While processing a PA request for a diabetic patient, the Context Synthesis Agent might discover the patient hasn't had a diabetic eye exam in 18 months (a HEDIS measure gap). The system can flag this to care managers, turning PA from a gatekeeping function to a care coordination trigger.
- **Total cost of care predictions:** Advanced implementations integrate predictive analytics where the Reasoning Agent estimates not just "is this medically necessary?" but "will approving this reduce downstream costs?" For example, approving a Continuous Glucose Monitor (CGM) for a poorly controlled diabetic may prevent future ER visits and hospitalizations—a net savings even if the CGM itself is expensive.

Conclusion: Preparing for the future

Prior authorization will not disappear - utilization management remains a necessary function in a value-based care ecosystem. But its implementation can and must evolve.
Agentic AI transforms PA from a compliance burden into a **strategic asset:** a system that ensures appropriate care, protects against waste, and does so with transparency, speed, and clinical nuance.

Organizations that move early on this transformation will gain competitive advantages in provider satisfaction, member experience, and operational cost structure. Those that delay
will find themselves unable to meet regulatory timelines, competitive benchmarks, or the expectations of an increasingly digital-native patient population.

The era of 'computer says no' is ending. The era of **intelligent, auditable, collaborative decision-making** is here.

References

1. American Medical Association. (2024). 2024 AMA prior authorization physician survey. Retrieved from https://www.ama-assn.org/system/files/prior-authorization-survey.pdf

2. Centers for Medicare & Medicaid Services. (2024). CMS Interoperability and Prior Authorization Final Rule. Retrieved from https://www.cms.gov/newsroom/fact-sheets/cms-interoperability-and-prior-authorization-final-rule-cms-0057-f

3. HL7 International. (2024). Da Vinci Prior Authorization Support (PAS) Implementation Guide. Retrieved from https://build.fhir.org/ig/HL7/davinci-pas/

4. CAQH.2024).2024 CAQH Index Report. Retrieved from https://www.caqh.org/insights/caqh-index-report

Mastech Digital is a Data Science-led, Data Engineering-driven partner helping enterprises build AI-ready data infrastructures for real-time insights and measurable outcomes. We unite advanced Data Engineering, cutting-edge AI/ML, and top-tier talent to maximize data ROI and accelerate innovation. Our offerings span Trust-Driven Data Foundations, AI-powered operations, adaptive CX, generative AI, and autonomous decisioning—delivered with speed, agility, and governance. Partnering with leading platforms and hyperscalers, we design secure, scalable, cloud-native solutions that enable clients to modernize, innovate, and lead in the AI-powered future.
Learn more at **www.mastechdigital.com.**

Let's Get in Touch

+1.800.627.8323
+1 412.787.2100
+1 412.494.9272

1305 Cherrington Parkway
Building 210, Suite 400 Moon
Township, Pittsburg, PA - 15108

experience@mastechdigital.com