Why agentic AI is pharmacy's strategic imperative for growth

Composable architectures and intelligent orchestration empower pharmacy companies to scale clinical impact and drive competitive differentiation.

An 'agentic pharmacy' reimagines healthcare delivery. By deploying specialized AI agents to manage cognitive workflows, organizations amplify pharmacists rather than replace them, unlocking superior clinical outcomes, accelerating operations, and driving competitive market growth.

Intelligent pharmacy orchestration with AI: Here's what matters

- Competitive pressure has become existential. D2C telehealth and digital pharmacy models are eroding the traditional captive pharmacy relationship.
- AI capability has crossed the clinical threshold: LLMs can now autonomously reason across complex clinical workflows with sufficient accuracy.
- A human-in-the-loop governance model ensures licensed pharmacists authorize all AI recommendations, allowing for scalable predictive analytics without regulatory exposure.
- Composable architectures utilizing modern APIs enable rapid integration of modular AI components, accelerating ROI—within six months, without massive legacy overhauls.
- Maturing HL7 FHIR APIs and real-time pharmacy benefit networks have created the precise data substrate required for operational intelligence.

Amplifying the pharmacist. Not replacing them

At its best, pharmacy is the most accessible point of contact in the healthcare system: the provider patients visit more often than their physician, the professional who holds the complete medication picture. Operational constraints have chronically prevented pharmacies from realizing that potential. The agentic pharmacy changes this by deploying a composable layer of specialized AI agents across every pharmacy workflow, handling the cognitive burden of population management, surfacing patients who need intervention, and drafting communications that build durable relationships. The pharmacist remains the authorizing clinical leader. The agents handle the volume. Here are three crucial considerations:

- **Composability over monolith:** Independent, swappable agents assembled like components. Any model, data source, or integration is replaced without re-engineering the whole system.
- **Agents draft, pharmacists authorize:** Every clinical recommendation passes through a licensed pharmacist before reaching a patient. Human oversight is the operating model—not a safety net.
- **Intelligence at every layer:** AI is not bolted onto existing workflows. It is the orchestration layer itself—reasoning, personalizing, and learning from pharmacist feedback continuously.

The architecture imperative: Five reasons why pharmacy's next move cannot be incremental

- **Adherence collapse:** Medication non-adherence costs the US healthcare system an estimated $100–300 billion annually in avoidable hospitalizations and wasted medication. Yet the dominant intervention, the batch refill reminder, addresses only one of six distinct non-adherence root causes. A one-size-fits-all reminder is not a solution.
- **Cost shock at the counter:** Between 30–45% of patients who encounter an unexpected medication cost at the pharmacy counter abandon their prescription entirely. This price discovery failure is entirely preventable, but current architecture cannot deliver pricing intelligence upstream of the dispensing moment.
- **Medication Therapy Management (MTM) at artisan scale:** A pharmacist can realistically deliver 3–5 high-quality MTM sessions per day. The bottleneck is not clinical skill but documentation. Subjective objective assessment plan (SOAP) notes, medication action plans (MAPs), and prescriber letters consume 40–60% of session time.
- **Channel fragmentation:** The average pharmacy patient exists in a fragmented state: a different record in the IVR, the app, the POS, and the pharmacist's workstation. Each channel holds an incomplete picture. The result is impersonal experiences, duplicated effort, and clinical risk from medication blind spots.
- **Specialty pharmacy complexity:** A single specialty prescription can trigger dozens of interconnected workflows. Managing this manually is slow, error-prone, and expensive. Siloed automation creates brittleness without solving the coordination problem.

Market timing: Why 2025–2027 is the critical window

Three forces are converging to create a narrow window of structural advantage for organizations that act now.

AI capability has crossed the clinical threshold

LLMs now demonstrate the ability to reason across complex, multi-step clinical workflows—from drafting prior authorizations to delivering personalized adherence outreach—with sufficient accuracy to operate under human oversight. Agentic frameworks have matured, enabling seamless coordination across tools, APIs, and data sources without constant human intervention.

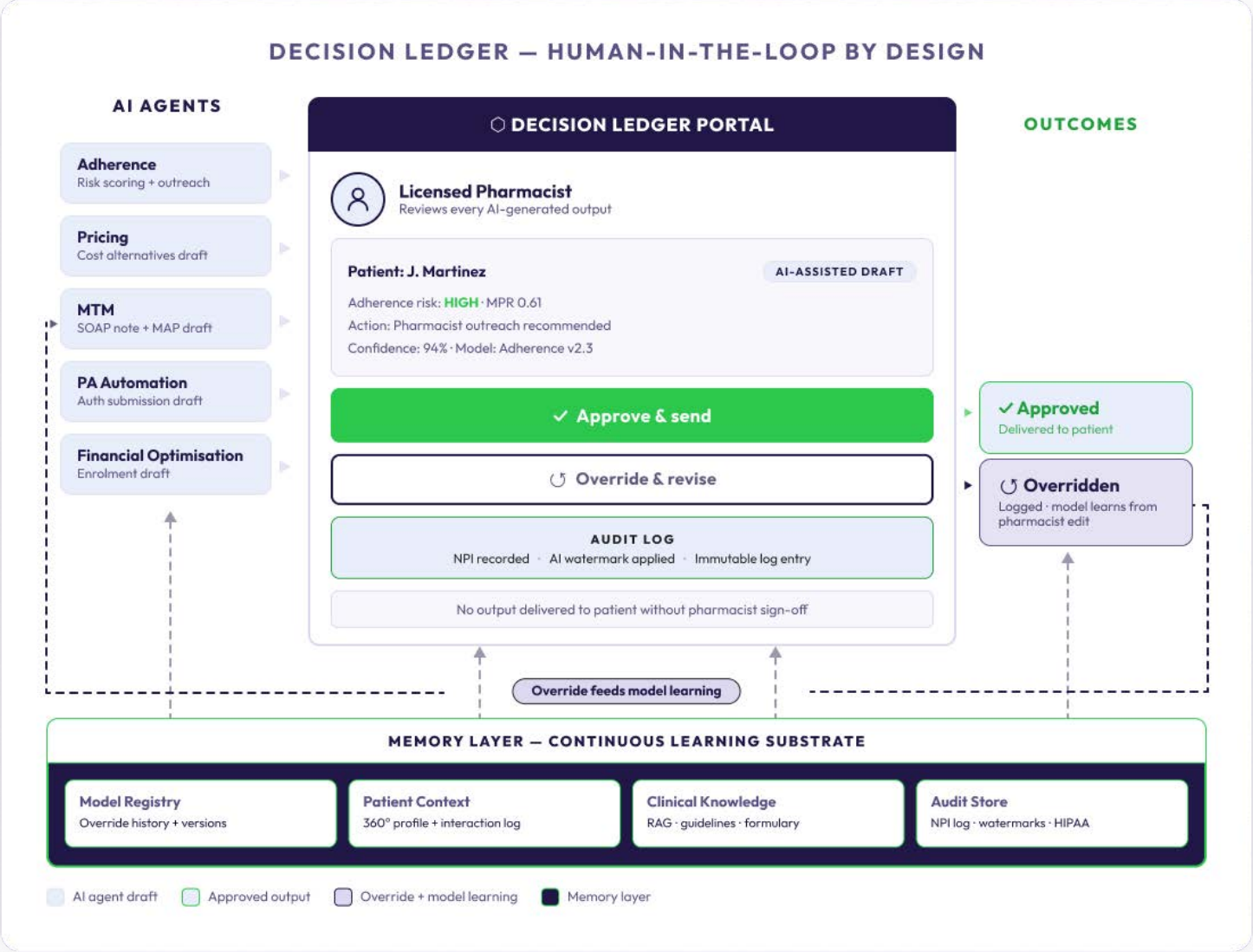
Data infrastructure is finally ready

The maturation of HL7 FHIR APIs, real-time pharmacy benefit networks, and EHR interoperability has established a robust data foundation for intelligent operations. For the first time, a unified API abstraction can map to every clinical model in the marketplace, offering unparalleled flexibility to integrate best-of-breed components as the AI landscape evolves.

Competitive pressure has become existential

D2C telehealth and digital pharmacy models are disrupting traditional pharmacy relationships. Differentiation through intelligent, personalized patient experiences is no longer optional but essential for survival. Organizations that invest in these capabilities now will secure structural advantages in patient retention, clinical outcomes, and payer contracting that will be nearly impossible to replicate three years from now.

Architecture: Five blocks that define the category shift



Composable agent marketplace: Instead of relying on a monolithic platform, the agentic pharmacy breaks operations into specialized, healthcare-trained agents that can be independently deployed, replaced, or upgraded without disrupting the broader system. Organizations can select and compose agents tailored to their specific workflows such as prior authorization, omnichannel patient engagement, claims processing, and financial optimization, enabling seamless expansion over time without starting from scratch.

Vendor-agnostic model abstraction: With new clinical LLMs and agentic frameworks emerging quarterly, organizations hardwired to a single model vendor face disruptive migrations whenever the landscape shifts. A common API abstraction maps to every clinical model in the marketplace, managing translation, modality adjustments, PHI gatekeeping, and response validation. This ensures the application layer remains unchanged, even as underlying models evolve.

The decision ledger: Human-in-the-loop by design: The decision ledger portal offers a unified interface where licensed pharmacists can review AI-generated summaries and recommendations with clear approval and override controls. Every agent's output is logged, every override contributes to model improvement, and every AI-generated communication is watermarked and attributed to the reviewing pharmacist's NPI. No clinical recommendation reaches a patient without explicit authorization.

Intelligent orchestration with emotional design: The orchestration layer optimizes for both clinical accuracy and the patient's emotional experience, treating the journey from 'worried' to 'satisfied' as a design objective. Clinical logic and governance are cleanly separated across six observable, auditable stages: input, context preparation, care orchestration, governance validation, response delivery, and feedback.

End-to-end specialty pharmacy automation: A single specialty prescription triggers an intelligent pipeline encompassing intake triage, eligibility and benefits verification, prior authorization drafting, clinical review, REMS compliance, financial optimization, fulfillment, and post-dispensing monitoring. For example, the financial optimization agent can reduce a $2,500 copay to $500 through automated copay card enrollment—a workflow that previously required manual coordination across multiple departments.

PATIENT JOURNEY

EMOTIONAL ARC

01

User input

Capture sentiment, language and raw intent from patient

"Me duele el brazo después de la cirugía"

Worried

02

Intent & context

Language detection, translation and intent routing to engines

2 discrete intents detected + routed

Uncertain

03

Care & task orchestration

Clinical LLM for care inquiry, EMR integration for scheduling

Clinical LLM: post-op care · EMR API: appointment

Processing

04

Validation & governance

Clinical accuracy check, PHI compliance and emotional tone gate

Accuracy ✓ PHI ✓ Tone: reassuring

Reassured

05

Response delivery

Bilingual response with confidence levels, citations and care context

Spanish + English · 94% confidence

Confident

06

User feedback

Capture satisfaction signal, close emotional arc, feed learning loop

Arc: resolved · Satisfaction: captured

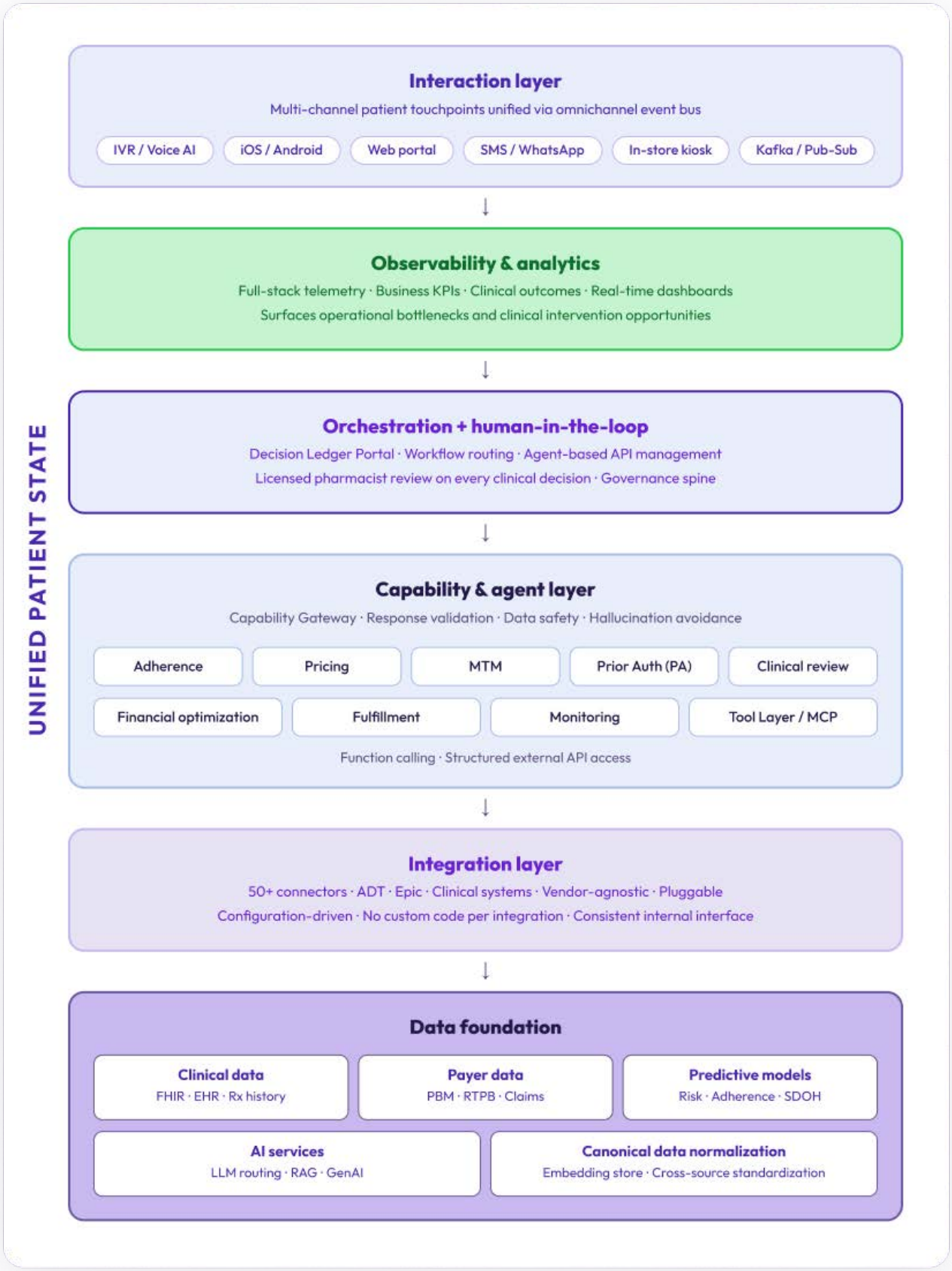
Satisfied

Core architecture principle

Clinical logic separated from governance — each stage independently optimized, validated, and auditable.

The agentic pharmacy stack

The agentic pharmacy is a fully composable system, with an orchestration layer that delivers complete value chains of functionality, reusable across experience channels. The architecture is structured into six distinct layers, each serving a specific function while maintaining seamless, clean interfaces with adjacent layers.



The interaction layer: One patient, one record. Collapsing the channel gap

The typical pharmacy patient exists in a fragmented state: separate records in the IVR system, the app, the POS, and the pharmacist's workstation. Each channel holds a partial view of the patient, but none provides the complete picture. This fragmentation results in an experience that feels impersonal at best and medically risky at worst.

The patient-centric architecture addresses this issue with a unified patient data platform: a real-time, event-sourced patient state that serves as the authoritative record across all channels. Every interaction, regardless of where it occurs, reads from and writes to the same patient state. This ensures that the pharmacist at the counter sees exactly what the app displays.

Event sourcing + CQRS for healthcare: Patient interactions are modeled as an immutable stream of events (e.g., RxFilled, AdherenceMessageSent, PharmacistConsultCompleted, PricingAlternativeAccepted). The current patient state is a projection of this event stream. This architecture not only provides a complete audit trail critical for compliance but also enables time-travel queries essential for quality reviews.

The in-store intelligence layer: When a patient approaches the pharmacy counter, the pharmacist's workstation surfaces an AI-generated patient brief. This brief includes current medications, recent adherence flags, outstanding MTM opportunities, and any pricing issues to resolve. The dynamically generated context card provides the pharmacist with the information that previously took 3–5 minutes to manually compile.

Integration layer: An intelligence substrate

The intelligence capabilities described above are powered by an integrated data layer designed for clinical-grade reliability, auditability, and continuous improvement.

Data input and ingestion: The ingestion layer normalizes data from multiple sources, including:

Knowledge stores: Workflow registries, business rules, clinical guidelines.
External and partner data: Prescription benefit networks, PA history, pharmacy networks, payer portals, lab reports.
Patient journey data: EHR records, interaction history, adherence data.

Data flows through ingestion and curation pipelines with canonical data normalization, producing curated datasets such as adherence history, patient 360 profiles, payer formulary rules, unified medication history, and historical PA records.

Context and intelligence layers: A context store built on Retrieval-Augmented Generation (RAG) provides the knowledge layer that grounds agent reasoning in verified clinical content. The ML Layer houses both classic

machine learning models (e.g., cost optimization, PA prediction, clinical review, rejection scoring) and deployed model endpoints managed through a model registry. An LLM Model Gateway routes requests to the appropriate language models, with tool-calling capabilities that allow agents to invoke structured external APIs.

Explainability and observability: Every AI-generated output is accompanied by contextual evidence via the RAG pipeline, confidence scores (both response-level and context-level), agent execution traces for auditability (via frameworks like Langsmith), and feature attribution through interpretability frameworks (e.g., SHAP/LIME). This observability layer ensures that pharmacists, compliance officers, and quality teams can interrogate any recommendation the system produces, understanding not just what the AI suggested, but why.

The agent layer: Specialized agents across every workflow

Each agent evaluates the patient's current state, selects the appropriate tool from a curated library, executes an action, and iterates as needed. When confidence falls below a defined threshold or clinical judgment is required, the agent escalates to a human pharmacist. This is not rule-based automation but clinical reasoning at population scale.

Agent	What it does	Escalates when?
Adherence	Scores non-adherence risk daily per patient. Identifies root cause: forgetting, cost, side effects, belief, or inertia - and selects a targeted intervention, not a generic reminder.	MPR < 0.70 in a high-risk condition patient.
Pricing	Queries insurance, GoodRx, copay cards, 340B, and LIS in parallel at prescription verification. Surfaces lowest verified cost before dispensing. Auto-drafts formulary alternatives when cost exceeds threshold.	Out-of-pocket above threshold; formulary exception needed.
MTM	Identifies CMR-eligible patients, performs DDI screening, surfaces clinical gaps, and drafts SOAP notes, MAPs, and prescriber letters from structured session data.	Complex polypharmacy; clinical discrepancy found.
PA Automation	Drafts prior authorization submissions matched to payer-specific policy, retrieves historical success patterns, and submits. Reduces manual PA prep from hours to minutes.	PA denial; clinical conflict requiring pharmacist review.
Financial Optimization	Runs full program eligibility matching - copay cards, patient assistance, LIS, 340B - and drafts enrollment submissions. Minimizes patient cost at every prescription event.	High affordability risk; copay above threshold.
Monitoring	Post-dispensing refill tracking, adherence monitoring, and rejection scoring. Triggers proactive outreach before patients disengage.	Adherence risk score crosses configured threshold.
Escalation	Routes complex cases to the pharmacist work queue with a structured case summary. Ensures nothing falls through gaps between agents.	Agent confidence below threshold on any output.

Agent capability deep dive

Root-cause-aware adherence: The science of not quitting

Non-adherence is not a single behavior but a family of behaviors with distinct causes: forgetting, cost shock, side effect mismanagement, health belief conflicts, social isolation, and inertia. A one-size-fits-all refill reminder addresses only one of these six root causes and risks irritating patients experiencing the other five.

Adherence risk scoring

The foundation model is trained on pharmacy dispensing history, claims data, social determinants of health (SDOH) features, and engagement signals. It generates a per-patient, per-medication adherence risk score, updated daily. Key feature categories include:

- **Behavioral signals:** Refill timing patterns, pick-up delays, channel engagement rates.
- **Clinical context:** Condition complexity, polypharmacy burden, recent hospitalization.
- **Financial signals:** Copay levels, past cost-related abandonment, insurance type.
- **SDOH features:** Transportation access, language preference, social support proxies.
- **Output:** A daily updated adherence risk score that drives the intervention priority queue.

The adherence agent evaluates each patient's non-adherence risk daily, reasons about the likely root cause using behavioral signals, clinical context, financial history, and social determinants of health, and selects the intervention most likely to address that specific patient's behavior. For example: A patient whose pattern correlates with cost issues receives proactive pricing alternatives. A patient showing signs of side effect burden is offered a pharmacist consultation. The result is meaningful engagement tailored to the patient's needs; not generic noise.

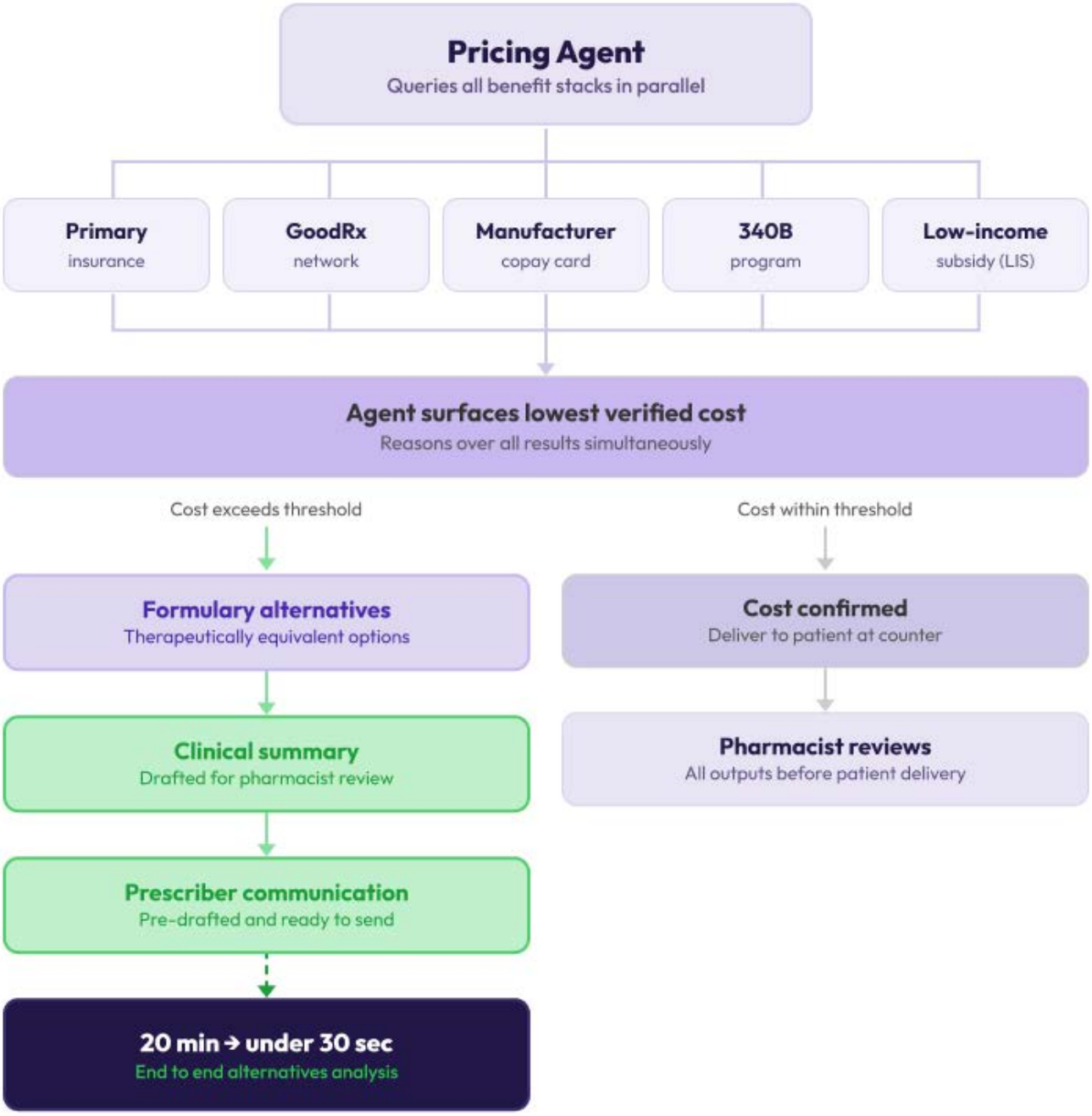
Upstream pricing intelligence: Eliminating cost shock

Real-time Pharmacy Benefit (RTPB) architecture

No patient should discover their medication's cost at the point of dispensing. The pricing agent pushes cost intelligence upstream into the prescribing encounter, digital confirmation flows, and automated outreach, addressing cost friction before it leads to abandonment. When costs exceed a configurable threshold, the agent automatically:

- **Generates a formulary alternatives analysis.**
- **Produces a clinical summary for pharmacist review.**
- **Pre-drafts the prescriber communication.**

What once required 20 minutes of pharmacy staff time is now completed in under 30 seconds.


Pricing Agent
Queries all benefit stacks in parallel
Primary
insurance
GoodRx
network
Manufacturer
copay card
340B
program
Low-income
subsidy (LIS)
Agent surfaces lowest verified cost
Reasons over all results simultaneously
Cost exceeds threshold
Cost within threshold
Formulary alternatives
Therapeutically equivalent options
Clinical summary
Drafted for pharmacist review
Prescriber communication
Pre-drafted and ready to send
20 min → under 30 sec
End to end alternatives analysis
Cost confirmed
Deliver to patient at counter
Pharmacist reviews
All outputs before patient delivery

THE MTM documentation flywheel

MTM is one of the highest-value services a pharmacist can provide, yet it remains chronically under-delivered. A comprehensive medication review (CMR) that identifies a drug-drug interaction or an untreated indication can prevent hospitalizations, reduce costs, and save lives. However, CMRs are time-intensive and documentation-heavy. In the current model, a pharmacist can realistically deliver only 3–5 high-quality MTM sessions per day.

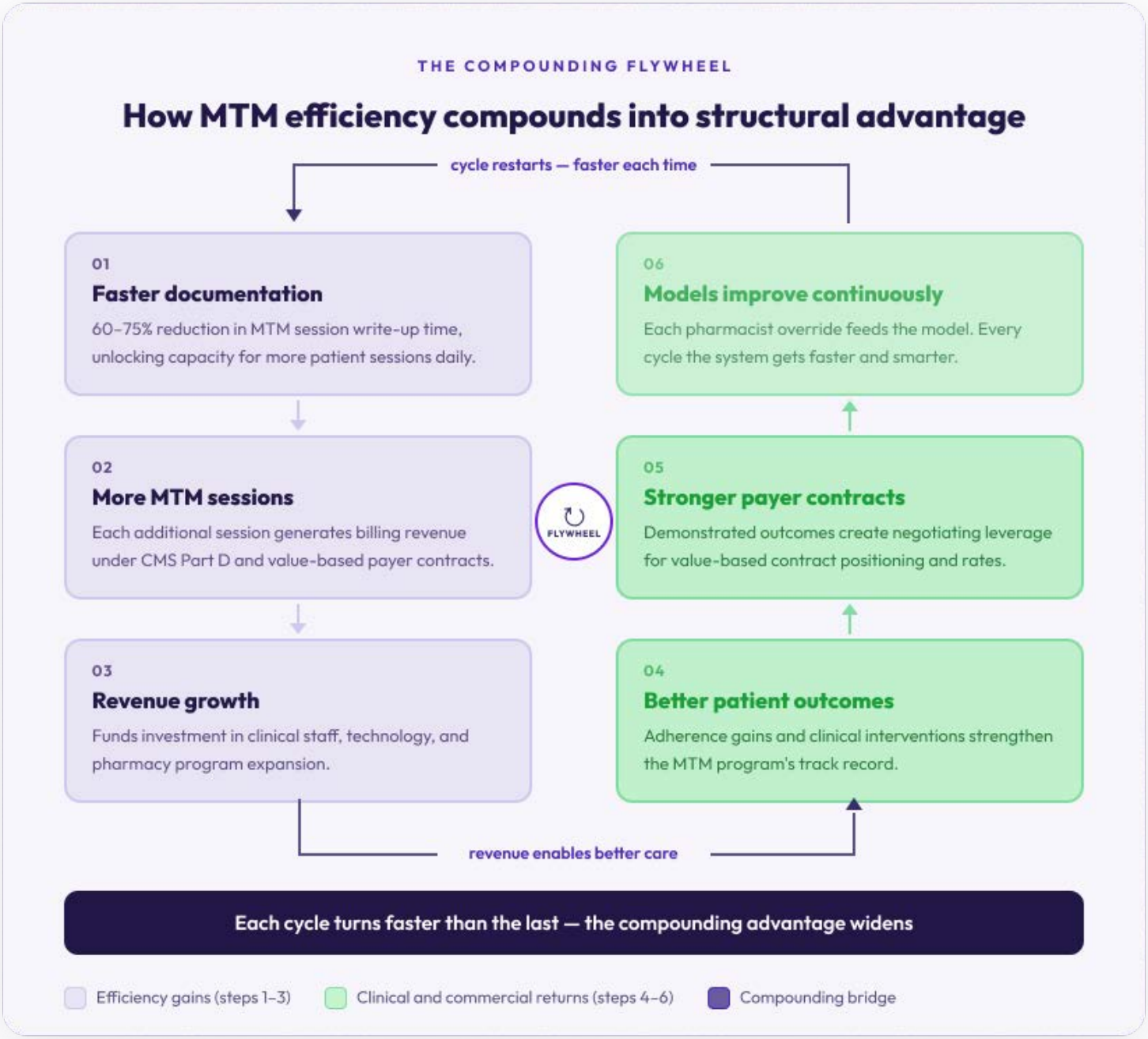
MTM agent capabilities

The MTM agent overcomes these constraints through six core capabilities:

- **Population Identification:** Identifies CMR-eligible patients using Rx history, claims, and payer enrollment data.
- **Medication Reconciliation:** Generates a FHIR-sourced comprehensive medication list with drug-drug interaction (DDI) screening via a clinical knowledge graph.
- **Clinical Gap Analysis:** Uses Retrieval-Augmented Generation (RAG) over clinical guidelines to surface untreated indications, monitoring gaps, and adherence issues.
- **SOAP Note Drafting:** Employs GenAI to draft pharmacist SOAP notes from structured session data, which are reviewed and edited before finalization.
- **Prescriber Communication:** Auto-drafts a MAP and personal medication list for both prescriber and patient.
- **Billing Automation:** Suggests CPT codes and assembles payer-specific documentation packages for submission.

The GenAI documentation flywheel

A pharmacist conducting a CMR currently spends an estimated 40–60% of their time on documentation tasks, such as writing SOAP notes, composing medication action plans, and drafting prescriber letters. The MTM Agent streamlines this process by generating first-draft documentation from a structured session transcript. The pharmacist then reviews, edits, and approves the drafts. This approach reduces documentation time by 60–75%, significantly increasing the number of patients a pharmacist can serve.



Trust and governance: Built for regulated clinical environments

Deploying AI in a clinical setting presents governance challenges as significant as the technical ones. The agentic pharmacy addresses these challenges with a governance framework spanning eight critical domains: authentication and role-based access, performance analytics, response validation, data safety, bias detection, hallucination avoidance, PHI gatekeeping, and comprehensive audit logging. Every production model operates under continuous monitoring, including performance drift detection (daily), fairness metric surveillance (weekly), and adversarial input testing (quarterly). If a model's performance on a protected demographic segment degrades by more than 5%, an automatic review gate is triggered.

The system is architected to comply with HIPAA, state pharmacy practice acts, and FTC guidelines on AI-generated health communications. Pharmacist autonomy is safeguarded by design: any agent recommendation can be overridden at any time, and every override contributes to improving future model behavior.

Pharmacist autonomy and professional responsibility
The agents are explicitly designed to support — not replace — pharmacist judgment. Pharmacists can override any agent recommendation at any time, with each override logged to enhance future model performance. The system transparently surfaces confidence levels and the evidence behind each recommendation, empowering pharmacists to interrogate and understand the reasoning.

No lock-in by design
Vendor-agnostic model abstraction means any AI model can be swapped without re-engineering the application layer. Your investment compounds regardless of how the AI landscape shifts.

Deployable in regulated environments today
The human-in-the-loop architecture is what makes agentic AI viable in clinical settings right now. No recommendation reaches a patient without licensed pharmacist review.

Complete audit trail
Every AI output is logged with the reviewing pharmacist's NPI. Every override feeds model improvement. Every AI-assisted communication is watermarked. Regulatory readiness is built in, not bolted on.

Talent and retention advantage
Pharmacists who practice at the top of their license - as clinical leaders supported by intelligent agents - are more engaged and easier to recruit. The system makes pharmacy the profession it was trained to be.

The rule-based automation trap

Here's something to think about. AI as a bolted-on tool or rigid, monolithic automation creates brittle operations. Intelligence must be the orchestration layer itself. Where licensed pharmacists authorize, and composable, specialized agents seamlessly execute.

Implementation: From assessment to value in 3–6 months

Implementation doesn't require building every component from scratch. A modular accelerator approach featuring prebuilt, healthcare-trained agents and workflow blueprints integrates seamlessly into existing technology stacks through a configuration-driven integration layer. This layer includes over 50 pre-built connectors across pharmacy, payer, and EHR systems. This approach significantly compresses deployment timelines: what would typically require 12–18 months of custom development is reduced to 3–6 months of configuration, testing, and deployment. Organizations can realize production value within the first quarter of go-live.

The architecture is explicitly designed for progressive adoption. Organizations can start by deploying a single high-impact agent such as MTM documentation, pricing intelligence, or PA automation to validate value before expanding. There's no need for a big-bang migration or wholesale system replacement. Each agent delivers value independently and compounds when combined with others.

Agentic pharmacy: A strategic pivot from incremental automation to intelligent advantage

- Agentic AI establishes a new operating model, scaling pharmacist judgment, clinical impact, and patient trust through governed human-in-the-loop orchestration.
- Composable, vendor-agnostic architectures future-proof pharmacy strategy, enabling rapid deployment, innovation, and measurable returns without legacy overhaul.
- Upstream pricing intelligence, root-cause adherence, and automated MTM documentation directly address cost abandonment, capacity constraints, and clinical variability.
- A unified patient record and intelligent orchestration layer eliminate channel fragmentation, reduce risk, and deliver consistent, emotionally intelligent experiences across touchpoints.
- Early movers can lock in defensible differentiation across retention, outcomes, payer leverage, and workforce engagement.

About Brillio

Brillio is a digital technology services company that drives AI-first engineering and design-led experiences for global enterprises. Born digital in 2014, its consulting-led services span Customer Experience, Data & AI, Product Engineering, and Digital Infrastructure. With an industry-leading NPS of 71, Brillio accelerates time to market through its proprietary BrillioOne.ai platform, powered by AI-ready talent with deep domain expertise.

Brillio is the official Digital Transformation Partner and the official Data and AI Services Provider of Atlassian Williams Racing. Brillio partners with leading technology providers including Microsoft, AWS, Google Cloud, Salesforce, Adobe, Databricks, and Snowflake and operates with 6,000+ “Brillians” across 15 global delivery centers. Consistently recognized as a Great Place to Work® since 2021, Brillio blends innovation, talent, and purpose to deliver measurable outcomes for clients and fulfilling careers for employees.

https://www.brillio.com/
Contact Us: info@brillio.com

brillio