

From Payment Integrity to **Payment Intelligence**:

Six Shifts That Will Reshape Health Plan
Payments in the Age of AI

Abstract

Health plan payment integrity was built for a slower world with fixed rules, periodic refreshes, and review teams working through unstructured claims data. That model is hitting its limits at the same moment providers are deploying AI to make billing, documentation, and appeals faster and more sophisticated than payer programs were designed to handle. The result is a widening gap between the velocity of payment error and the velocity of the systems meant to catch it. *Closing that gap requires a fundamental shift from payment integrity as a set of human-scale, retrospective programs to payment intelligence—a real-time, continuously learning function capable of evaluating every claim at the speed and resolution adjudication demands.* Drawing on experience working alongside leading health plans, we identify six structural shifts that will define that transition and the leadership choices that help plans evolve with the future.

Introduction

The United States healthcare system now processes [\\$5.3 trillion](#) a year in spending, and for the plans operating within it, getting payments right is mission critical. As one payment integrity executive put it in a recent conversation with us, "Nearly every claim has something coded wrong. Most of it doesn't matter financially. Our job needs to be—very simply—finding the part that does, at scale, and fast. I worry that the systems and tools we have used for this so far are not the ones we need for the future."

That tension sits at the center of payment integrity (PI) today, and it is getting harder to resolve. Administrative complexity is rising, clinical and payment rules are proliferating, and data volumes are expanding. Providers, meanwhile, are arming themselves with artificial intelligence (AI). [According to a 2024 survey](#), 65% of health system revenue cycle leaders believe generative AI will substantially affect medical coding operations. Provider-side AI is giving providers more sophisticated tools to optimize billing, sharpen appeal strategies, and push documentation to its limits, and it is doing so at a pace most payer PI programs were never built to match. That pace, more than any single tactic providers deploy, is what payers are now racing to catch.

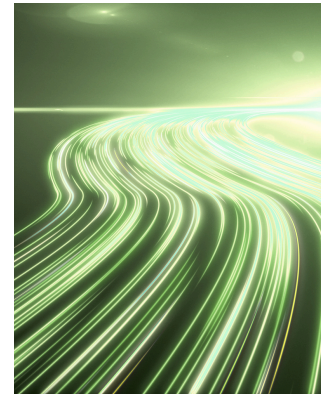
Payer leaders have long asked how to prevent as many incorrect payments as possible. In our conversations with payer leaders, that question is now joined by a deeper one. Can the payment integrity function itself keep pace—operating fast enough, accurately enough, and at the scale claims volumes demand—as providers adopt AI of their own?

After working alongside health plans for years, we've identified four dimensions that determine whether a PI function is built for what's coming:

- **Quality:** Does the program find errors reliably and make accurate decisions, or does volume compress judgment?
- **Scale:** Does the program cover the full claims population, or does it sample a fraction to fit a fixed review capacity?
- **Speed:** Does the program catch errors in near real time, as part of adjudication, or only after the claim has already been paid?
- **Scope:** Does the program surface new and unknown ways money is leaking, or only the patterns it already knows to look for?

Quality is the dimension most programs overestimate. Most PI leaders believe their review quality is solid, because the alternative—that experts are missing things at scale—is hard to see from inside the program. Capacity problems are visible through backlogs, aging inventory, employee feedback, and missed deadlines. Quality erosion is not as visible. It only shows up later as overturned decisions, missed recoveries, and patterns nobody happened to catch, at which point it looks like bad luck rather than a structural symptom.

This is the shift from payment integrity to payment intelligence.



Today's PI model was built for a slower, more static claims environment with fixed rule sets, periodic refreshes, and review capacity sized to a sample of claims rather than the full population. That model cannot keep pace with a provider side generating new billing patterns, documentation strategies, and appeal tactics faster than payer teams can study them, write rules for them, and push those rules through approval. That gap, between the velocity of AI-driven billing behavior and the velocity of the PI response, is the structural breaking point.

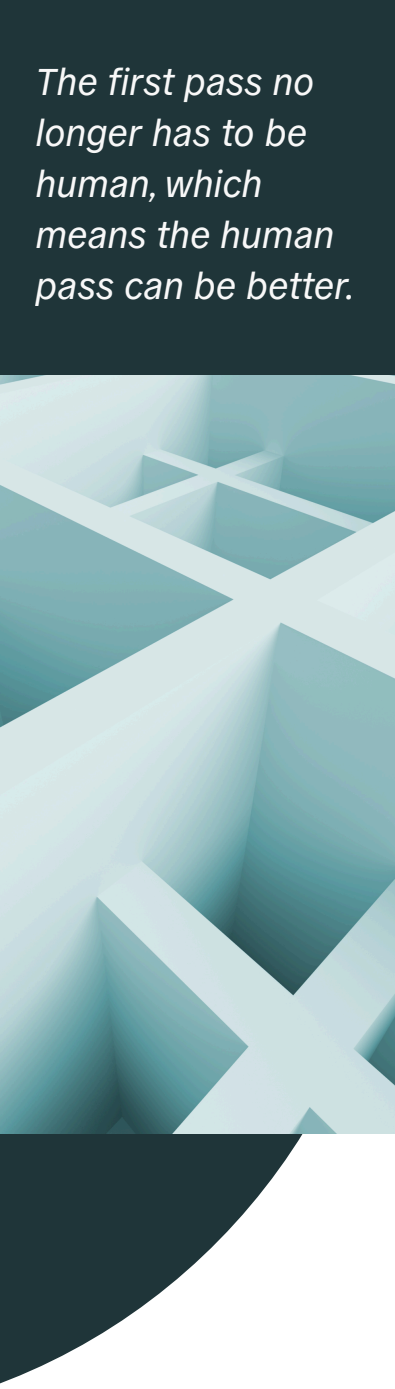
Closing it requires payment intelligence, a model in which AI reads and reasons over unstructured claims and clinical data at a scale no review team can match, surfaces patterns no one has yet thought to look for, and feeds every outcome—missed, caught, overturned, or recovered—back into the system as a learning tool.

That continuous loop lets the function improve on its own, rather than waiting on the next round of expert-authored rules. And once that loop is running, the intelligence no longer needs to live inside the rigid, transaction-focused systems that process claims today. It can surround them, reasoning in real time and routing decisions back into adjudication before a claim is ever paid.

This is the shift from payment integrity to payment intelligence. It's a real-time, intelligence-powered function with continuous learning built into its operations, capable of the quality and scale that veterans of payment integrity have long wanted but never had the tools to achieve.

We believe six structural shifts will define the next decade of payment intelligence.

I. From error-prone manual reviews to quality-assured AI



The first pass no longer has to be human, which means the human pass can be better.

Payment integrity has historically treated complex work as human work. A nurse or clinician reads the record, a coder validates the Diagnosis-Related Group (DRG) code, an investigator builds the subrogation case, a reviewer drafts the letter. Evidence gathering, judgment, and execution are bundled into one manual workflow. That model has delivered real value—most PI programs today return somewhere between 5:1 and 10:1 on investment—but it is worth asking whether it is approaching its ceiling.

That ceiling shows up first as a capacity problem when backlogs grow, aging inventory accumulates, and deadlines slip. The quality problem is harder to see. When reviewers work through high volumes of cases, judgment compresses. Cases that deserve deeper scrutiny get the same initial time as cases that don't. Tired eyes miss patterns that fresh eyes may have detected. The experts doing the work are skilled, but the conditions rarely let them do their best work, and most programs underestimate how much that costs them.

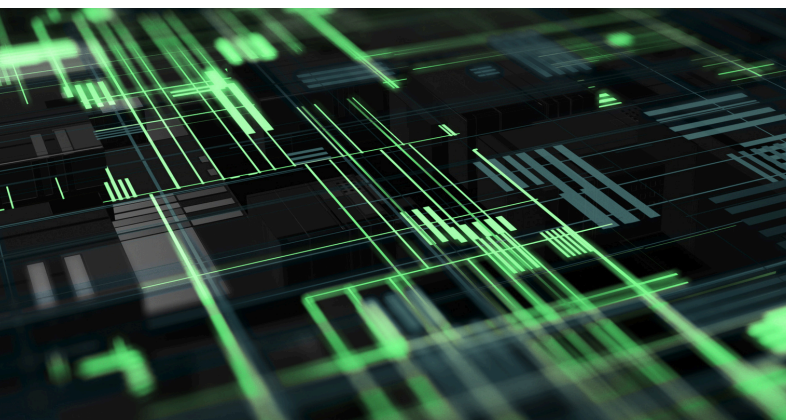
AI changes those conditions. Large language models (LLMs) and agents can read clinical, legal, and claims text; assemble evidence; draft recommendations; and execute repeatable steps. The first pass no longer has to be human, which means the human pass can be better.

Chart review is one of the clearest examples. Instead of asking a coder or clinician to start with a raw chart, AI can summarize the stay, map diagnoses and procedures against what was billed, and flag the specific documentation that actually needs expert judgment. In practice at Machinify, AI-assisted review workflows can achieve recall rates above 90% on findings while correctly routing the bulk of no-finding cases to auto-close. The expert's time shifts to the cases that actually require judgment, rather than the full volume of reviews, producing higher-quality reviews and identifying more errors. Notably, experts unanimously reported that AI-assisted reviews and suggestions pushed them to consider things they wouldn't have otherwise, improving quality significantly.

The financial case for this rests on how much leverage a small quality gain produces against a large existing cost base, not on some multiplying effect. For one payer, our analysts showed the program would remain net positive even if review costs doubled, as long as hit rates improved by just 1%. That reframes the economic goal. The aim is not to drive down the cost per review, but to increase the value captured per review.

Subrogation follows the same pattern. Recovery rights determination, third-party identification, case building, and lien negotiation all depend on messy, unstructured information. AI can assemble the case and surface the moments that require expert judgment, rather than asking investigators to rebuild context from scratch on every file. The result is higher-quality reviews that produce stronger recoveries.

But quality is only part of the story. The other parts are scale and scope. Erroneous claims form a pyramid. At the top sits a small number of errors that are large, obvious, and well-understood. These are the ones existing rules and experienced reviewers are built to catch. Beneath that sits a far larger volume of smaller, subtler, less-obvious errors. These are the ones that don't trip a known rule, don't reach a dollar threshold that earns review time, or simply haven't been named yet. A manual-first model is structurally confined to the top of that pyramid. No amount of additional hiring changes that because the constraint is the operating model, not the headcount.



Expert reviewers track policy adherence, exceptions, clinical appropriateness, and governance with the key claim details surfaced by AI. That shift is what allows quality and scale to improve together.

AI removes that ceiling by changing what a single expert can cover, not just how quickly they cover it. It can review the full population of claims rather than a sample sized to fit review capacity, which means the pyramid's base—the long tail of smaller errors that was previously invisible by design—becomes addressable for the first time. Combined with quality-assured review, this raises the floor on how well every claim gets reviewed while also expanding how much of the pyramid the program can reach at all.

The human role, in this model, shifts from operator to assuring quality. Expert reviewers track policy adherence, exceptions, clinical appropriateness, and governance with the key claim details surfaced by AI. That shift is what allows quality and scale to improve together, rather than trading off against each other as they do in a manual-first model.

II. From expert-written rules to AI-discovered content

Content in PI has long been constrained by expert authoring capacity and the organizational machinery required to act on it. A senior coder, nurse, investigator, or actuary sees a pattern, then translates it into a set of rules. That's only the beginning. From there, it enters a chain of clinical review, coding validation, legal and compliance scrutiny, operational assessment, and committee approval before it ever touches a live claim. The process is rigorous by design—each handoff is a legitimate check—but it is slow, and it depends on human experts who know what questions to ask and which answers matter most.

By the time a new piece of content clears the full approval process, the billing behavior that prompted it may have already evolved. And the pipeline only moves at all when an expert has seen enough of a pattern to name it, document it, and shepherd it through the process. Data mining models, audit selection logic, and clinical policy all face the same authoring bottleneck. Content only gets written when someone thinks to suspect it, and only after it survives evaluation by multiple committees.

AI changes the source of content. Systems can mine claims, contracts, policies, records, and outcomes, surfacing statistical patterns and correlations a human reviewer would not think to look for. Those patterns, once surfaced, can be shaped into proposed edits and policies far faster than expert teams can author them from scratch. The discovery problem starts to disappear because the patterns no longer depend on someone thinking to suspect them.

A recent Machinify review for a leading payer illustrates this well. Claims for a specific class of oncology prescriptions were being billed at the maximum dosage amount regardless of patient weight. This pattern appeared consistently across a large claims population. No rule in any edit library captured it. A human reviewer would have needed to already suspect the pattern to go looking for it, but AI found it by scanning the data directly. The financial magnitude of the pattern was quantifiable within days—faster than a manual investigation would have even gotten off the ground.

A human reviewer would have needed to already suspect the pattern to go looking for it, but AI found it by scanning the data directly.

In at least one domain Machinify has worked in closely, AI-driven pattern discovery has identified over \$100 million in payment integrity findings. The significance extends well beyond the speed of discovery. Findings at that scale suggest there is a much larger body of payment error that has yet to be identified because no one has had reason to suspect it. AI is positioned to bridge that gap in ways that manual, incremental rule-writing never will.



The role of the clinical and coding expert then shifts from authoring every rule to curating, validating, and governing what AI proposes. Traditional edits are deterministic—a claim either meets the criteria or it doesn't, and the rule isn't expected to carry a false-positive rate because it isn't probabilistic in the first place. What AI proposes is different. It starts as statistical patterns, not validated rules, and it needs to earn its way into deterministic, deployable logic. When AI surfaces a proposal, the expert's job is to validate the clinical or coding logic behind it, assess how often it would flag claims that turn out to be correct, review how similar claims have fared on appeal, and decide whether—and how—to convert it into something that can run in production with the reliability a traditional edit has. That validation step is what makes AI-discovered content safe to deploy at the determinism that PI operations require.

Content creation, in this model, starts to function more like a continuous feedback loop than the periodic authoring cycles most payers run today. What doesn't change is the need for a governance layer that decides what gets deployed and why.

III. From pay-and-chase to decision co-processing

Most payers already run a substantial share of editing pre-pay, and coordination of benefits, basic code edits, and eligibility checks have operated before claims finalize for years. The question now is how much further upstream, and how much deeper into complex claims, its logic can now extend, and why so many claims are still resolved after the fact instead of before payment.

Half of the answer is structural. Historically, the technology available for clinical judgment, evidence assembly, and pattern detection wasn't fast enough or capable enough to run before a claim finalized, so anything that needed real judgment got pushed downstream into post-pay review and recovery. But the other half of the answer is economic. Much of the post-pay industry runs on contingency fees, where vendors are paid a percentage of what they recover on paid claims. That model rewards finding errors late, and it gives the parties best positioned to fix the upstream pipeline little financial reason to do so. The result is an inventory of avoidable errors that gets discovered downstream by design, not necessity.

The cost of that design shows up in float, dispute volume, write-offs, and provider abrasion, all of which accumulate the longer an error sits unresolved. Payers have long sensed that the dollar never incorrectly paid is worth more than the dollar recovered later. What's changed is that AI now makes it possible to act on that instinct at a scale that goes well beyond the editing that already runs pre-pay—covering more claim types, more complex judgment calls, and a far larger share of total claims volume than pre-pay logic has ever reached before.



That requires a different architecture than simply expanding the existing edit set. Core adjudication platforms were built for stability, throughput, and determinism, not rapid intelligence, and they need to remain systems of record. They are also typically too legacy, too brittle, and too expensive to extend quickly enough to keep up with this shift. Rather than replacing the adjudication system outright, the more workable path looks like a strangler-fig approach, in which a new layer grows around the existing system, gradually taking on more of the intelligence work without disturbing the core. The adjudication platform remains the transaction backbone, while a modern, AI-enabled decision layer surrounds it, reasoning, learning, and adapting at a different speed. That layer functions less like an additional edit and more like a decision co-processor to adjudication itself.

Coordination of benefits (COB) is a useful illustration of how much further this can go, not because pre-pay COB is new, but because improved underlying intelligence has unlocked coverage of a far larger share of the coordination workload than simple rule-based pre-pay logic was ever capable of reaching. Eligibility signals are typically available before a claim closes, and modern COB logic can resolve coordination in sub-second time inside or adjacent to adjudication, rather than catching only the cases simple rule-based pre-pay logic was built to catch. Plans that have pushed COB further upstream this way have seen return multiples around 8:1 compared with post-pay recovery alone.

Higher-complexity categories show the harder version of the same shift. DRG and clinical review depend on medical record evidence, which has historically made them resistant to anything but post-pay review. Agentic AI is beginning to change that, moving evidence assembly and chart review earlier in the claim lifecycle so that a meaningful share of cases that once had to wait for post-pay audit can be resolved before the claim pays. The implication is that even claim types long assumed to require after-the-fact review can move upstream once evidence-gathering gets fast enough.



The transition still requires careful sequencing. Clean claims should continue to pay quickly. Suspect claims should be surfaced before payment with enough evidence and rationale to support action. The adjudication system remains the system of record, but the judgment that used to happen downstream increasingly happens in the co-processing layer around it. Post-pay becomes cleanup, exception handling, and a learning signal about what should move earlier next time, not the primary mechanism for getting payment right.

IV. From siloed programs to the Intelligence Hub

Most payment integrity organizations are not one system but many programs operating side by side—coding, fraud, clinical review, chart review, recovery, COB, subrogation, vendors, and internal teams—each built around a human-scale process that made sense when every domain required its own specialized reviewers, its own flow, and its own cadence. A claim sits in line for clinical review, then separately enters a fraud queue, then separately gets evaluated for COB, each on its own schedule, each unaware of what the others have found. Each team asks which claims belong in their queue rather than the more important question of what a specific claim actually needs. When every program depended on manual judgment, that fragmentation was largely unavoidable. It makes much less sense now that AI can evaluate a claim against most of PI's logic in the time it takes to adjudicate it, because the constraint that forced those evaluations apart—the speed of human review—is no longer the bottleneck.

Reframing from program-centric to claim-centric operations is possible with AI, and it is a more meaningful shift than mere routing efficiency. When a program is built around human review, splitting it from other programs is almost unavoidable as each domain requires its own specialized reviewers, its own workflow, and its own pace. When the underlying analysis can run in real time, that separation stops being necessary.

The future state is a unified Intelligence Hub. As a claim comes out of adjudication, the Hub applies relevant intelligence across the full payment integrity landscape in sub-second time, determining whether the claim is clean or whether it carries signals for COB, subrogation, secondary code edits, DRG review, data mining concepts, contract logic, or provider policy. Routing the claim efficiently to the next right action is a natural consequence of evaluating everything at once rather than passing a claim through disconnected queues, but that efficiency is a byproduct rather than the goal. The goal is that payment integrity stops being a downstream function that reacts to what adjudication already decided and becomes part of the adjudication itself, operating at the speed the claim demands rather than the speed a set of human-scale programs can sustain.

The goal is that payment integrity stops being a downstream function that reacts to what adjudication already decided and becomes part of the adjudication itself, operating at the speed the claim demands rather than the speed a set of human-scale programs can sustain.



This also builds a much better data foundation. One of the hardest problems in payment integrity today is reconstructing the full x-ray of a claim: what happened, who changed it, why it changed, what evidence supported the decision, and how each program acted on it. Some of that transaction record lives in the adjudication system, but the rationale, evidence, workflow history, and decision trace often do not. Unraveling what actually happened can become a forensic exercise precisely because the programs were never designed to share a single view of the claim.

A unified Intelligence Hub solves that by creating a shared decision trace for every claim, including lifecycle status, decision history, evidence, ownership, program activity, and every material modification in one place. Once this layer is established, payment integrity can become real-time, adaptive, and governable in the way the next era demands.

V. From static rules to continuously learning intelligence

That shared decision trace is also what makes a learning system possible because the system cannot learn from outcomes it cannot reconstruct.

Most payers are familiar with PI content maintenance. Content is written, deployed, refreshed periodically, and eventually overtaken by changes in billing behavior, policy, documentation, and operational reality.



When the market moves slowly, that model works well. It breaks when both provider behavior and payer operations are changing faster than static rules can keep up, as is the case today.

The problem is that payment integrity programs often do not learn enough from their own mistakes. Those mistakes come in two forms. There are errors of omission: claims that should have been flagged but were missed. And there are errors of commission: claims that were flagged, pended, denied, audited, or recovered in ways that created cost, abrasion, or rework without enough value.

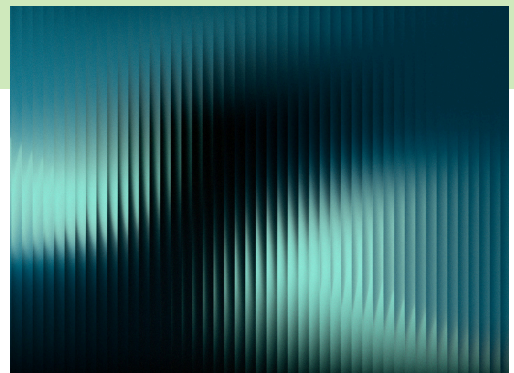
In a learning system, every outcome becomes information. A missed recovery teaches the system where its detection logic was too weak. A no-finding audit teaches where selection was too broad. An overturned denial teaches where the evidence standard, policy logic, or routing decision was wrong. A provider abrasion pattern teaches where the program may be technically correct but operationally destructive. The system gets better not only by adapting to provider behavior, but by continuously improving its own judgment.

That is the shift from static rules to continuously learning intelligence. Every adjudicated, reviewed, appealed, recovered, overturned, or untouched claim becomes part of a feedback loop. The goal shifts from finding issues to improving precision. More of the right claims are selected, fewer clean claims are disturbed, and better evidence is assembled, so that fewer decisions create downstream waste.

AI now makes continuous learning possible. Patterns surface faster, feedback loops close in real time, and the system can propose new content faster than any expert team could write it manually. What makes that loop trustworthy is the validation layer that governs what actually gets deployed. Every proposed rule still requires clinical review, false positive assessment, appeal survivability testing, abrasion monitoring, and performance measurement against both missed opportunity and unnecessary intervention. The speed of learning has to be matched by the rigor of validation, or the loop becomes a liability rather than an asset.

Much of the urgency for payers comes from provider-deployed AI tools. Coding, documentation, and appeal tools are improving quickly, and payer systems need to learn at the same cadence as the market. But shifting toward AI-assisted reviews doesn't signal payer reactivity. Rather, it is a sign of internal growth toward payment intelligence that becomes measurably better from every mistake it makes, whether the mistake was failing to act or acting when it should not have.

Every adjudicated,
reviewed, appealed,
recovered, overturned,
or untouched claim
becomes part of a
feedback loop.



VI. From AI arms race to a cooperative transparency compact

The most obvious response to providers adopting smarter billing, documentation, and appeal tools is for payers to adopt smarter review, denial, and recovery tools. Both sides invest in better capabilities for the same fight—and the fight gets more expensive without getting more accurate.

The result is a system where both sides become incrementally better at contesting payment, but the overall economics don't improve. Appeals grow and administrative cost compounds, and while automation may make each side more capable of winning individual disputes, it doesn't make the payment system more accurate or less expensive to operate.

...payers with a reputation for transparent, predictable adjudication face fewer appeals and experience more productive relationships with their provider networks.

The dynamic is self-reinforcing, and it has a structural economic dimension that makes it hard to escape. The same contingency-fee economics that sustain pay-and-chase also sustain the arms race. The parties best positioned to catch errors late have limited financial incentive to help prevent them early. The arms race persists not only because both sides have learned to expect adversarial behavior from each other, but because the incentive structures in place reward that adversarialism. More intelligent technology on both sides reinforces both the expectation and the incentive.

Breaking it requires a demonstrated willingness to reduce avoidable disputes rather than to win them more efficiently. Early examples of what this looks like are already emerging. Moving COB further upstream is one: the payer resolves coordination before the claim closes rather than disputing it afterward, reducing surprises for providers and friction for both parties. The more consequential version of this shift, though, is what becomes possible when payers make their adjudication logic legible at the claim level. Detailed, record-level annotation shows a provider not just what standard is being applied, but precisely how that standard is being applied to their specific documentation, in real time. That kind of transparency changes the nature of the payer-provider relationship. Providers stop guessing at what a payer will accept and start understanding it, which is a different and more durable form of alignment.

That shift toward a shared goal is what makes trust possible, and trust changes the economics more durably than any individual tool. Machinify's discussions with providers consistently indicate that payers with a reputation for transparent, predictable adjudication face fewer appeals and experience more productive relationships with their provider networks.

Payment integrity has the opportunity to evolve into the kind of function that is woven into the adjudication process, visible to providers, and oriented around accuracy rather than recovery. That is a different ambition than deploying better tools for the same fight, and it is the one most likely to change the system rather than just one side's position within it.

The leadership question

Taken together, these shifts mean that the next decade will reward plans that build intelligence systems, not just scale of review.

Provider-side AI will keep raising the sophistication and speed of billing, documentation, and appeals. That sophistication and speed, on their own, is a good thing. The industry needs for claims to move more quickly, but we also need accuracy. The problem is that the payer-side operating model that was built for a slower, more static claims environment and has not yet been redesigned for one that moves this fast. Payers can respond by building better tools for the same fight, or they can build a different architecture altogether that is oriented around governed AI execution, unified claim truth, continuous learning, and greater transparency across the payment lifecycle.

Payment intelligence is a new operating model for how claims are understood, routed, reviewed, paid, challenged and learned from. It compounds in value the longer it runs, rather than hitting the ceiling that every human-scale program eventually finds.

**Payment intelligence is
a new operating model for
how claims are understood,
routed, reviewed, paid,
challenged and learned from.**

Transforming everything at once is neither possible nor practical. The starting point is to build the data foundation, move one or two mature categories upstream, give expert teams the tools to govern work at scale, and create feedback loops that make the system smarter every time it misses an opportunity or creates avoidable abrasion. The healthcare payment system has spent a generation getting very good at disputing money. Now, we have the chance to get equally good at not needing to—which, for the patients whose care sits behind every claim, is the point that always mattered most.

ABOUT THE AUTHORS



Prasanna Ganesan is Machinify's Chief Technology Officer (CTO). He founded Machinify in 2016 with the goal of unlocking innovation for healthcare organizations with safe and transparent AI solutions. Prior to Machinify, he co-founded VUDU in 2005 (acquired by Walmart in 2010), where his pioneering work as CTO resulted in over 30 patents.



Shri Santhanam is Chief Product Officer at Machinify, where he leads product strategy and the development of technology platforms transforming healthcare payments. Prior to Machinify Shri was at Experian, where he was GM of software and platforms and Chief AI officer leading the company's AI strategy and transformation plan. An active innovator and speaker on the application of data, technology, and artificial intelligence in business transformation, Shri has appeared in Bloomberg and Nasdaq TradeTalks, among other industry forums. Shri is an engineer by education and earned his Bachelor of Engineering from the Indian Institute of Technology Madras, where he received the Governor's Gold Medal for academic excellence and leadership, and later completed his Master of Engineering at Stanford University.

ABOUT MACHINIFY

Machinify is *the* healthcare payment intelligence company with an AI-native, human-validated operating system that fixes healthcare payment *errors before they happen*, reducing systemic waste and administrative complexity to make the system better for everyone.

Connect with our team at info@machinify.com
or by visiting www.machinify.com/contact.