



The Payment That Runs Itself

How AI is quietly taking over the back office of payments, and what the smartest banks and fintechs are doing about it

CONTENTS

01 The Big Shift - Payments used to be about speed. Now it is about intelligence..... 3

02 Why Now - Three things changed at once, and they all landed on operations..... 4

03 The Hidden Cost - The quiet tax of an operation that cannot keep up..... 5

04 The New Operating Volume - The self-running payment: let AI carry the 95%, keep experts on the 5%..... 6

05 Proof It Works - The largest players are already doing it, with public results..... 8

06 The Shift, Stage By Stage - From a wall of queues to an operation that runs itself..... 9

07 Where You Are Today - The Autonomy Ladder: four rungs to a self-running operation..... 10

08 The Next Test - When the customer is a bot, trust becomes the product..... 11

09 What To Do - Five moves to build a self-running operation..... 12

10 The Bottom Line - The rails will keep getting faster. The advantage moves up a level..... 13

11 Putting It To Work - Sutherland FinPay360, the self-running model in practice..... 14

THE BIG SHIFT

Payments used to be about speed. Now it is about intelligence.

For over 20 years, the race in payments was simple: move money faster. Faster rails. Faster checkout. Faster settlement. That race is nearly over. In the United States, less than 10% of consumer spending still happens in cash, and instant payments moved more than \$2 trillion in 2025.

Speed, in other words, is no longer a differentiator. It is the baseline.

The prize is still huge. Payments generated about \$2.5 trillion in revenue in 2024 and is expected to reach \$3 trillion by 2029, making it the most valuable business in financial services.¹ But the easy money is gone. After years of near 9% growth, revenue is now expected to grow around 4% a year as fees compress and interest-rate tailwinds fade.² So the question every payments leader now faces has quietly changed. It is no longer **“how do we move money faster?”** It is a harder one:

How do we run a payment operation that is instant, intelligent, and safe all at once, at a cost that still works when margins are thinner and volumes keep rising?



The answer is not a new rail. It is a new way of running the operation on top of the rails, one where the everyday work increasingly runs itself.

WHY NOW

Three things changed at once, and they all landed on your operations team

Payments did not drift into this moment. Three forces arrived together, and none of them hit the rails. They hit the people and systems that process, check, and settle every transaction.

The force	What is happening	Where it hits you
1 Money is now instant and final	Real-time rails like FedNow and RTP moved over \$2 trillion in 2025. Once an instant payment clears, it cannot be pulled back.	Every fraud and compliance check must happen in milliseconds before the money leaves, not overnight.
2 AI agents are becoming customers	Visa and Mastercard both launched agent-payment programs in 2025. Mastercard completed its first live AI-agent purchase that September.	A new kind of buyer, one that is software, now hits your checkout and does not behave like a human.
3 Rules got stricter and faster	Suggested answers, real-time guidance, next-best action, summarization, entity extraction, automated QA, predictive CSAT/NPS/churn signals.	You must prove what your systems decided and why, in real time, not reconstruct it at audit.

Exhibit 1 • The three forces reshaping payments, and where each one actually lands.

Instant money, machine buyers, and always-on compliance all land in the same place: the operation. And most operations were built for a slower, batch-based world. That is the gap this paper is about.

THE HIDDEN COST

The quiet tax of an operation that cannot keep up

A payment operation built for batch processing taxes the business in three ways. Most leaders feel these costs without ever seeing them added up.

1. Too many false alarms.

Old rule-based systems flag far more good transactions than bad ones, often 10 to 20 false alerts for every real fraud caught. This wastes analyst time and drives away good customers. Research cited across the industry shows banks can lose 10% to 12% of customers in the year after a single wrongful decline.³

2. Growing exception queues.

Onboarding repairs, sanctions hits, disputes, and chargebacks all fall out of automation and into manual queues. At real-time volumes, those queues grow faster than any team can hire. Turnaround slows, and customers feel it.

3. Five teams, five systems, no single view.

Onboarding, processing, fraud, disputes, and servicing usually run in separate silos with separate data. When an AI agent needs a yes-or-no answer in a fraction of a second, an operation stitched across five silos simply cannot respond in time.

The scale of the problem is not small. **Organizations lost an average of \$60 million each to payment fraud last year, and the global cost of fraud passed \$485 billion in 2024.**

In a batch world, a slow operation was merely expensive. In a real-time, AI-driven world, a slow operation is a liability the moment the money leaves.



THE NEW OPERATING VOLUME

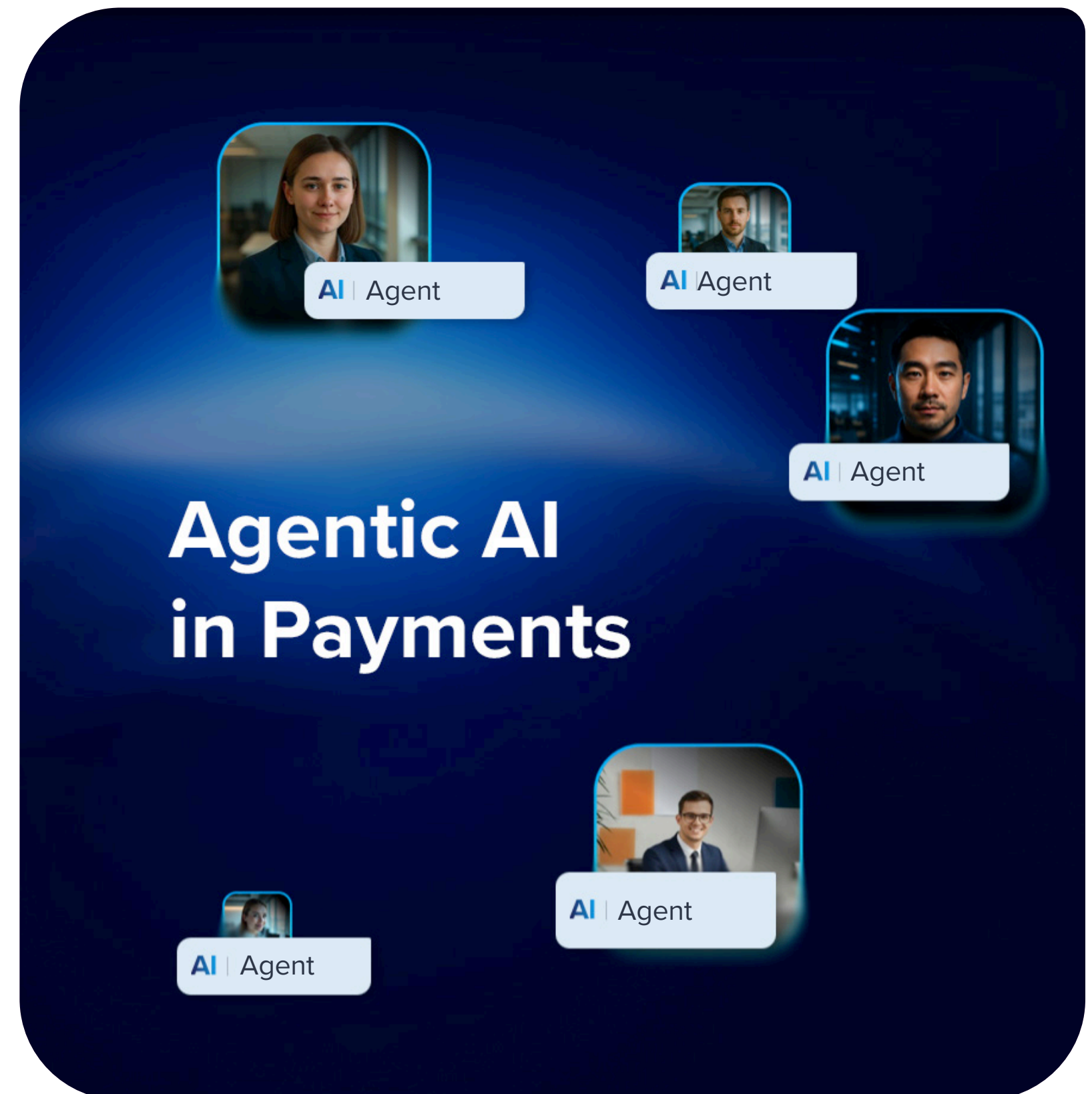
The self-running payment: let AI carry the 95%, keep experts on the 5%

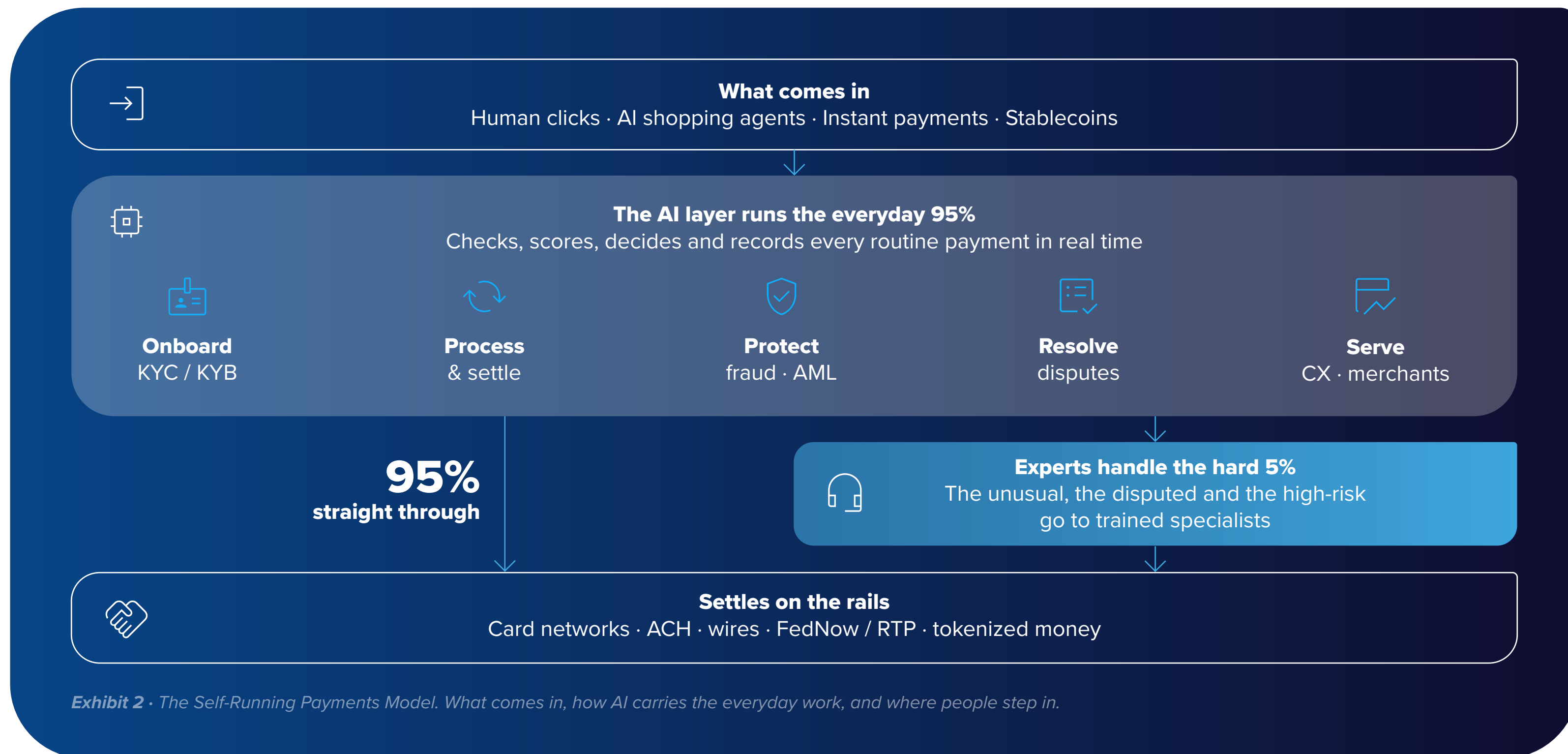
Here is the shift in one sentence. Stop running payments as a wall of manual queues, and start running it as an operation that mostly runs itself, with people focused only where their judgment matters most.

We call this the **Self-Running Payments Model**. The principle is simple:

AI runs the everyday volume; specialists govern the exceptions.

Intelligent systems handle roughly 95% of payments that are clean and routine, at machine speed. Trained people are reserved for the hard 5%: the unusual, the disputed, and the high-risk, where accountability and a clear, auditable reason really matter.





This is not about removing people. It is about moving them off the treadmill of routine checks and onto the decisions that actually protect the business and the customer.

PROOF IT WORKS

This is not a theory. The largest players are already doing it.

The self-running model is not a vision of the future. The institutions with the most to lose have already moved, and the results are public.

When AI carries the everyday volume, fraud is caught in milliseconds, good customers stop getting declined, and analysts finally spend their day on the cases that matter.

Institution	What they did	Result
JPMorgan Chase	Built AI fraud detection that weighs thousands of signals per transaction and scores in real time	About \$1.5B in yearly savings; fraud caught roughly 300x faster than old systems
HSBC	Deployed AI to monitor 900 million transactions a month across 40 million accounts	60% fewer false positives; hundreds of millions in attempted fraud stopped
Mastercard	Runs generative-AI “Decision Intelligence” across its network to score transactions in context	Screens 159 billion transactions a year; 83% of leaders say AI cut false positives
Stripe	Acquired stablecoin firm Bridge and launched agentic-commerce tools for AI-driven checkout	Positioned to run AI-agent and stablecoin payments as core, not experiments

Exhibit 3 • What AI-run payment operations are delivering, in named, public examples.

The pattern is consistent. When the routine work is automated and the exceptions are concentrated with experts, three things move at once: cost falls, speed rises, and quality improves.⁴ These are not competing goals anymore. The self-running model delivers them together.

THE SHIFT, STAGE BY STAGE

From a wall of queues to an operation that runs itself

The change plays out across the whole payment lifecycle. Here is what it looks like at each stage.

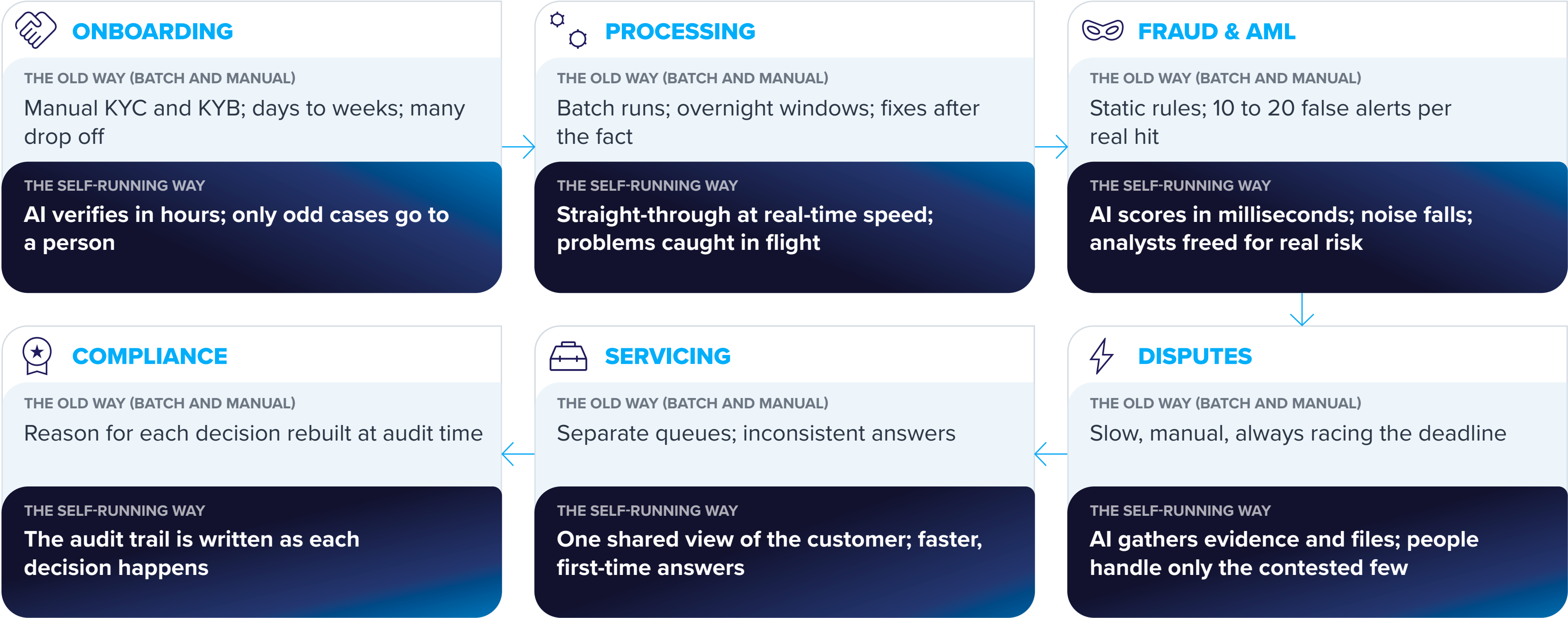


Exhibit 4 • The old operation versus the self-running operation, across the payment lifecycle.

WHERE YOU ARE TODAY

The Autonomy Ladder: four rungs to a self-running operation

No one jumps straight to a fully self-running operation. Progress happens in stages. Use the ladder below to find your rung honestly, then decide where you need to be given how fast your rails and your customers are moving.



 Rung	 How it runs	 What it looks like	 What holds you back
1 Self-Running	Runs itself, well governed	AI carries the 95%; experts govern the 5%; audit built in	The frontier; needs steady investment and discipline
2 Orchestrated	AI leads, people oversee	AI decides across stages; exceptions routed to experts	Data and governance still catching up to the automation
3 Assisted	Tools bolted on	Some automation and AI in pockets; the rest still manual	Automation stops at the easy tasks; exceptions still pile up
4 Manual	People and queues	Analysts work alerts; batch processing; siloed teams	Cannot scale with real-time volume; high cost to serve

Exhibit 5 • The Autonomy Ladder. Four rungs from manual queues to a self-running, well-governed operation.

Most banks and processors sit today between rung 2 and rung 3. They have automated the easy, rules-based tasks. What is left is the hard middle: fraud investigation, dispute resolution, and complex servicing, where old automation runs out of road. Rung 4 is not about cutting people. It is about placing them exactly where their judgment creates the most value.

THE NEXT TEST

When the customer is a bot, trust becomes the product

Agentic commerce breaks a quiet assumption inside most fraud systems: that a real person is on the other end. That assumption is now wrong, and the numbers are moving fast.

+805%

AI-agent shopping traffic, one year

\$67B

Sales influenced by AI agents, one week

Sept 2025

First live AI-agent card purchase

Exhibit 6 • Agentic commerce is arriving fast. Figures from Adobe Analytics, Salesforce and Mastercard, 2025.

The card networks have responded by giving AI agents special, scoped credentials with built-in spending limits.⁵ But issuing a credential is the easy part. Someone still has to decide, in real time, whether a given agent-driven payment is a genuine purchase or a bad actor in disguise. Even the networks building the fastest rails admit the unsolved problems are agent identity and dispute resolution, the judgment questions, not the settlement ones.⁶

This is exactly the work a self-running operation is built for: fast, contextual, well-governed decisions, with a clear reason recorded every time. Whoever does this well will own the advantage in AI-driven commerce.

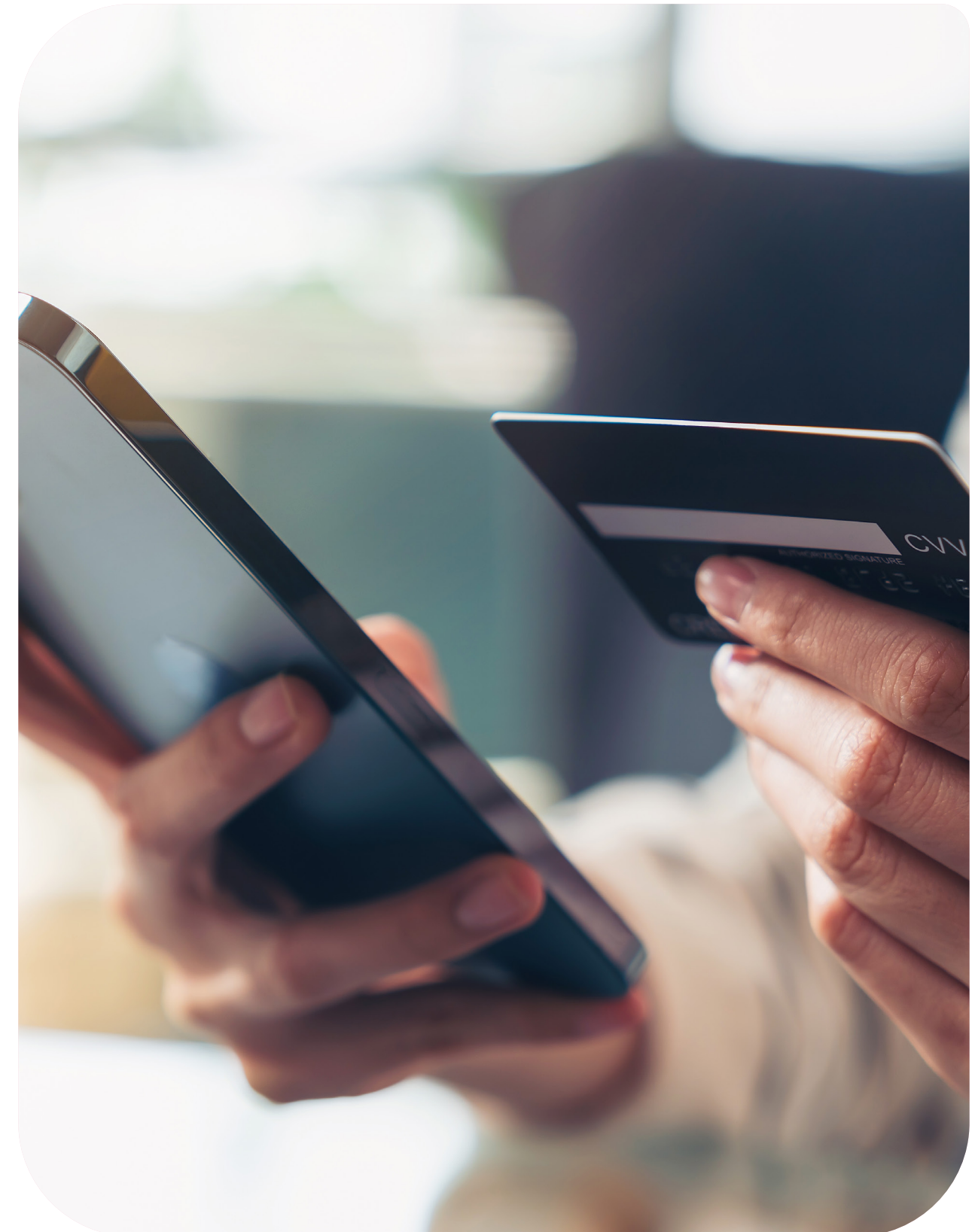


WHAT TO DO

Five moves to build a self-running operation

Strategy has to become a plan. Five moves separate the institutions that will run an intelligent payment operation from those that keep paying the batch-era tax.

- 1 Join up the lifecycle first.** AI cannot run an operation split across five silos. Put onboarding, processing, fraud, disputes, and servicing on shared data before you automate. Fragmentation is the one thing automation cannot fix.
- 2 Design around the exception.** Let AI carry the clean 95% at machine speed, and build clear, well-instrumented paths that hand the hard cases to specialists. The exception path is where risk and customer trust are won or lost.
- 3 Build compliance in, not on.** Record the reason for each decision as it happens. With instant money and machine buyers, rebuilding the story at audit time is no longer good enough for regulators or boards.
- 4 Get ready for AI agents now.** Treat agent identity and real-time agent-versus-fraud decisions as a 2026 capability, not a someday project. The demand curve is already bending up.
- 5 Measure outcomes, not activity.** Cost to serve, straight-through rate, false-positive rate, dispute time, and first-time resolution are the scoreboard. Build the transformation, and the partnerships behind it, around moving those numbers.



THE BOTTOM LINE

The rails will keep getting faster. The advantage moves up a level.

Before long, every institution will have the same instant rails, the same agent credentials, the same real-time infrastructure. Those become commodities. What will not become a commodity is the intelligence that decides what to do with all of it: the operation that runs the volume, governs the exceptions, and proves its reasoning on demand.

Sutherland's view is simple. The winners in the AI era of payments will stop treating operations as a cost center bolted under the rails, and start treating a self-running operation as the product itself: AI-run, human-governed, and built to make payments smarter, faster, and safer in the same motion. That is the loop we build and run: intelligent systems carrying the volume, specialists governing the exceptions, and results you can measure across cost, risk, growth, and experience.

Which rung of the Autonomy Ladder is your operation on today, and what would it take to climb one before your rails and your customers force the question?

PUTTING IT TO WORK

Sutherland FinPay360

Everything in this paper points to one operating model: AI runs the everyday volume, and trained specialists govern the exceptions. That is exactly what Sutherland FinPay360 was built to deliver. FinPay360 is an AI-native, rail-agnostic payments solution that spans the full lifecycle, from onboarding and processing to fraud, disputes, and servicing. The wedge is the same one this paper argues for: **AI handles the volume of screening, monitoring, and processing, while experts own disputes, chargebacks, and complex escalations.** It is not a tool bolted onto a queue. It is the operation, run as an outcome.

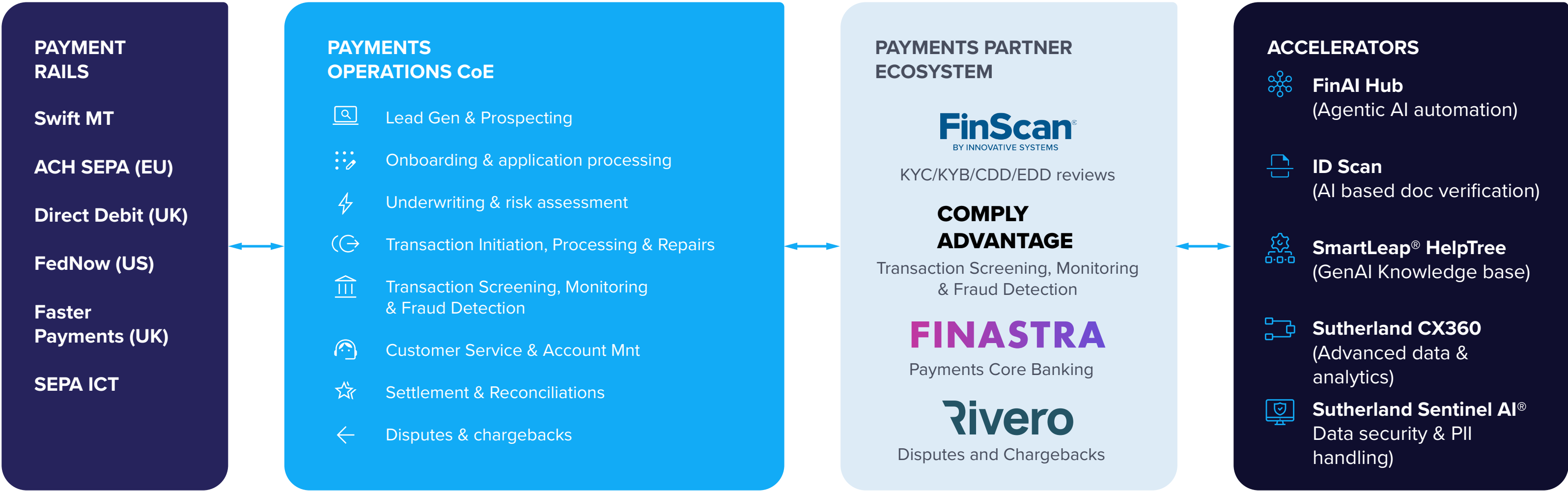
New operating model	FinPay360 does	The result
Let AI carry the routine 95%	A library of ready-to-deploy payment agents for processing, repair, fraud, AML, monitoring and servicing	99% of payments screened in under 0.5 seconds, 24x7
Keep experts on the hard 5%	Trained specialists own disputes, chargebacks and complex escalations	Up to 70% faster repair and exception resolution
Cut the false-alarm tax	ML-tuned screening across thousands of watchlist sources	About 70% fewer false positives in monitoring
Build compliance in, not on	Real-time screening and PCI-compliant data protection built into the flow	100% real-time payment and customer screening
Get ready for AI-agent commerce	Operational support for stablecoins, tokenization, wallets, and programmable payments	Next-gen rails ready as demand arrives

Exhibit 7 • How FinPay360 maps to the takeaways in this paper.

Behind the model sits real depth: more than 7,400 payments specialists and 25-plus payments clients across cards, banking payments, and remittances, proprietary accelerators including FinAI Hub, Sutherland Sentinel AI®, ID Scan, Sutherland CX360®, and SmartLeap® HelpTree, and outcome-based commercials that tie Sutherland's compensation to your savings, efficiency, and customer-experience gains. Sutherland was named a Leader in Everest Group's Payments Business Process Services PEAK Matrix Assessment 2025.

Sutherland FinPay360: AI-Driven, E2E Payments Offering

Comprehensive payments offering across payment rails built on bedrock of domain, AI, and Digital leveraging Sutherland IP and Industry Partnerships



ADVISORY SERVICES

DIGITAL ENGINEERING & SYSTEMS INTEGRATION SERVICES

References

All figures from sources published within the last 24 months. Accessed August 2026.

1. **McKinsey & Company, Global Payments Report 2025. Payments revenue about \$2.5T in 2024, on track for \$3T by 2029; cash below 10% of US consumer spending. Summarized via PaySpace Magazine, October 2025.**
<https://payspacemagazine.com/news/mckinsey-report-raises-stakes-for-global-payments-revenues-forecast-to-surpass-3t-by-2029/>
2. **Boston Consulting Group, Global Payments Report 2025: The Future Is (Anything but) Stable. Revenue growth easing from about 8.8% to about 4% a year; real-time A2A payments up about 40% in 2025.**
<https://www.bcg.com/press/22september2025-reshape-global-payments-landscape>
3. **FluxForce AI (citing McKinsey Financial Services research), Fraud detection ROI analysis, 2026. Banks can lose an estimated 10% to 12% of customers in the year after a false decline; 10 to 20 false alerts per genuine fraud.**
<https://www.fluxforce.ai/blog/fraud-detection-api-integration-add>
4. **Reuters, Mastercard and BusinessStats, AI in payments results, 2025 to 2026. JPMorgan about \$1.5B in annual fraud savings and roughly 300x faster detection; HSBC 900M transactions a month monitored with a 60% false-positive cut; Mastercard Decision Intelligence screening 159B transactions a year; average \$60M fraud loss per organization and \$485B global fraud impact in 2024.**
<https://www.mastercard.com/global/en/news-and-trends/Insights/2026/ai-is-helping-banks-save-millions-by-transforming-payment-fraud-prevention.html>
5. **TechInformed and Digital Commerce 360, Visa Intelligent Commerce and Mastercard Agent Pay, 2025 to 2026. Mastercard launched Agent Pay (April 2025) and completed its first live agentic transaction (September 2025); Visa launched Intelligent Commerce and the Trusted Agent Protocol.**
<https://techinformed.com/visa-opens-one-integration-for-ai-agent-payments/>
6. **Forbes Digital Assets, Visa, Mastercard and Coinbase Are Fighting Over How AI Agents Pay, June 2026. Networks name agent identity and dispute resolution as the unsolved governance problems of agentic commerce.**
<https://www.forbes.com/sites/digital-assets/2026/06/07/visa-mastercard-and-coinbase-are-fighting-over-how-ai-agents-pay/>
7. **Adobe Analytics and Salesforce (via Paz.ai analysis), 2025 holiday commerce data, December 2025. AI-agent shopping traffic up about 805% year over year on Black Friday 2025; AI agents influenced about \$67B in Cyber Week sales.**
<https://www.paz.ai/blog/the-payment-networks-are-all-in-what-visa-mastercard-and-paypals-q4-moves-signal>
8. **Arnold & Porter and Congress.gov, The GENIUS Act of 2025. Signed into law July 18, 2025; the first US federal framework for payment stablecoins.**
<https://www.arnoldporter.com/en/perspectives/advisories/2025/07/new-stablecoin-legislation-analyzing-the-genius-act>
9. **IntelliPay and Federal Reserve Financial Services, US real-time payments data, 2025. FedNow and RTP together processed more than \$2 trillion in 2025; adoption of ISO 20022 on faster rails.**
<https://intellipay.com/real-time-payments-are-here-what-fednow-and-rtp-actually-mean-for-your-business/>
10. **Stripe and Forbes Digital Assets, Stablecoin volume and infrastructure, 2025 to 2026. Stablecoin transfer volume reached about \$27.6T in 2024; Stripe acquired Bridge; stablecoin rails moving into mainstream payments.**
<https://stripe.com/resources/more/stablecoin-trends-businesses-need-to-understand-in-2026>

SEE IT END TO END

Explore Sutherland FinPay360

The full offering, the agent library, the accelerators and the outcomes. Scan the code or visit the page.

● Get in touch



SCAN TO EXPLORE

Sutherland is an AI-driven business transformation partner to enterprises, including several Fortune 500 clients. We design, run, and automate business operations at scale, helping clients improve efficiency, strengthen customer experiences, and deliver measurable outcomes.

Most companies either build AI-led capabilities or run operations. Sutherland does both. We run complex, mission-critical processes while engineering the AI-led solutions that sit within them. This closed loop helps clients move beyond AI pilots to reliable, governed outcomes that improve and scale over time.

With deep operational expertise, advanced products, platforms, and engineering, we help clients modernize operations and build future-ready businesses.

