PASTBEHAVIOR RESEARCH

When Software Becomes the Operator

What happens when software stops helping people do the work and starts being answerable for the outcome.

11 CASES · 6 STILL STANDING · 5 RUNGS · 1 UNRESOLVED QUESTION

01 / THE ANSWER

Operational complexity does not disappear. It changes owner.

Every company in this study crossed the same line. It stopped selling a tool that a customer used to do the work, and became the party that performs the work and answers for it. Payroll gets filed. Money moves. A container clears customs. When it goes wrong, the vendor is the one holding it.

The lazy version of what follows is that these companies discover they were secretly service businesses all along. That is not what the record shows, and the best-documented case here contradicts it directly. What the record shows is narrower and more useful.

Companies that made this transition intact tended to build the system of record before they took on the obligation, and to grow the obligation slowly enough that the operating apparatus could be built alongside it. Companies that ruptured either took the obligation on as the business model, before any apparatus existed, or took it on top of real infrastructure and then let the volume outrun the people.

> **Sequencing lowers the starting risk. Compression decides whether that advantage survives.**

That second half is where the record thins out badly. Nearly every company that got through this grew its responsibility-bearing book over years, so the question facing the current generation of AI-native operators is one the past cannot answer. It is being tested right now, in public, by companies that have not finished.

Two other findings came out of the work that were not obvious going in. The ladder runs in both directions, and one of the largest companies here climbed back down. And the recurring failure across a decade of cases is not an operational failure at all. It is a measurement failure, and it repeats with unusual precision.

> **The repeated mistake was measuring how much work the software touched instead of whether the hard work actually got resolved.**

---

02 / THE LADDER

Five different businesses hiding under one word

"Software company" covers at least five distinct positions, each with a different cost structure and a different failure mode. Companies move between them without announcing that they have, which is why the transition is usually visible only in hindsight. This is a lens for reading the cases rather than a finding proved by them. No company is obliged to climb.

---

01 **Selling software**

The customer does the work and owns the outcome. Errors belong to them.

What attaches: engineering, sales, support. **Failure mode:** churn.

02 **Assisting the work**

The software recommends, drafts, calculates, or flags. A human at the customer signs off.

What attaches: accuracy expectations, deeper support. **Failure mode:** distrust, then churn.

03 **Administering the work**

The vendor performs the work on the customer's behalf. Runs the payroll, ships the pallet, processes the enrollment.

What attaches: implementation, exception handling, service operations. These scale with transaction volume rather than with revenue. **Failure mode:** cost to serve, and backlog.

04 **Legally or operationally responsible**

The vendor is the named party. It holds the license, signs the filing, employs the worker, clears the shipment.

What attaches: licensing, certification, audit, controls, legal exposure. **Failure mode:** fines and enforcement, which do not announce themselves as gradually as churn does.

05 **Bearing financial risk**

The vendor's balance sheet absorbs some part of the outcome.

What attaches: capital requirements, reserving, pricing cycles, loss ratios. **Failure mode:** a capital requirement discovered rather than planned.

The distinction between rungs three and four matters more than it looks. At rung three, an unhandled exception is a support ticket. At rung four, the same exception is a late filing with a statutory penalty attached. The work is identical. The consequence is not.

Most of the mechanism in this study lives at rungs three and four. Rung five is a smaller and more specialized problem, handled separately near the end.

03 / THE COHORT

Who crossed, when, and what happened

Eleven cases, chosen so that a majority did not rupture. A cohort assembled only from wreckage proves its thesis by selection, and postmortems get written while quiet competence does not. Six came through intact. Two corrected course and are still operating. Three ruptured.

COMPANY	MOVED	DOMAIN	OUTCOME	WHAT IT ESTABLISHES
Zenefits	1 → 4	HR / insurance brokerage	**Ruptured**	Responsibility as the original business model, with no apparatus underneath
Rippling	1 → 4	HR / payroll / PEO	Intact, strained	Sequencing solved the compliance failure and not the delivery one
TriNet	5	PEO	Steady state	What the mature version costs, published quarterly
Justworks	1 → 5	PEO	Intact	The slow path, and what it costs in time
Gusto	3	Payroll / benefits	Intact	Channel built as an operating layer, not a lead source
Deel	1 → 4	Global employment	Intact	Fast climbing is possible, and prices itself in headcount
Shopify	3 → 1	Commerce / logistics	Reversed	The ladder runs down as well as up
Flexport	4	Freight / customs	**Corrected**	Operator headcount scaled into a spike and could not unwind
Olive AI	2 → 3	Healthcare admin	**Ruptured**	An automation claim measured against the wrong denominator
Babylon Health	2 → 5	Healthcare	**Ruptured**	Skipping rungs three and four to reach five
Klarna	4	Consumer credit	**Corrected**	Current-generation AI at the exception boundary, and the walk-back

Bright Health Group and Oscar Health are treated separately in the rung-five section. Their evidence bears only on companies carrying balance-sheet risk, which most operators do not.

04 / THE FIRST MISTAKE

When the obligation arrives before the machinery

Zenefits sold free HR software to small businesses and made its money as a licensed insurance broker on the health plans those businesses bought through the platform. That structure put it on rung four in year one, before it had built anything that belonged there.

The consequence was that growing revenue and growing regulatory exposure were the same activity. Sales headcount was the growth lever. Sales headcount was also the licensing obligation. There was no configuration in which one could safely outrun the other, and nothing in the product would catch the gap, because the product was not the thing that was breaking.

Zenefits

RUNG 4 FROM YEAR ONE

FACT An internal software macro let sales employees skip pre-licensing education requirements. Former employees told BuzzFeed News it was in use as early as 2013 and as recently as 2015. In February 2016 the same outlet reported that 83 percent of the company's Washington State insurance deals through August 2015 had been transacted by employees without the required state licenses.

FACT Parker Conrad resigned as CEO and left the board on February 8, 2016. Roughly 250 employees were cut on February 26. In November 2016 the California Department of Insurance concluded an enforcement action with a $7 million fine, half suspended, which the department described as among the largest licensing-violation fines in its history. An SEC administrative proceeding followed in October 2017. Conrad surrendered his California insurance license by settlement in May 2018.

FACT TriNet completed the acquisition of Zenefits on February 15, 2022, at a total purchase price of $209 million disclosed in its Form 10-Q. The company had been valued at $4.5 billion in a 2015 round. The two figures rest on different accounting bases and are useful for magnitude only.

> "We sell insurance in a highly regulated industry. In order to do that, we must be properly licensed. For us, compliance is like oxygen. Without it, we die. The fact is that many of our internal processes, controls, and actions around compliance have been inadequate, and some decisions have just been plain wrong."
>
> DAVID SACKS, INCOMING CEO, IN AN EMAIL TO ZENEFITS EMPLOYEES, FEBRUARY 2016

Conrad's own statement on the way out named the same thing from the other side: the company's management infrastructure and policies had not kept pace with its growth. Both men are describing an apparatus problem rather than a software problem. That framing recurs in every rupture in this study, which is the first thing worth carrying forward.

---

05 / THE SECOND ATTEMPT

Zenefits to Rippling, and what the comparison can prove

The same founder built a second company in the same category and did not repeat the same failure. That is irresistible as a story and treacherous as evidence, so it is worth separating what the record supports from what hindsight supplies.

What is documented

FACT Rippling was founded in 2016 and launched in 2017, starting with employee onboarding. Conrad described the architecture to TechCrunch in October 2022: an employee graph at the bottom of the stack, middleware above it for reports, permissions, and workflow automation, and products on top of that. Rippling PEO was announced to the broker market around November 2020, four years in, on top of payroll and benefits that already ran. Global payroll came in October 2022, six years in.

FACT The November 2020 PEO launch was aimed at brokers rather than around them. It offered brokers real-time visibility into client data and renewal information so they could advise clients on whether a PEO remained the right fit, one-click data export so a client could be taken back to market, and moving clients off the PEO in days rather than months. Zenefits had made itself the broker and taken the commission. Rippling made itself something the incumbent channel could place, and built the exit into the product at launch.

> **A company that builds the off-ramp on day one is designing for the case where its own responsibility does not fit the customer.**

What hindsight is supplying

INFERENCE No public statement from Conrad establishes that PEO was sequenced later because of Zenefits. The compound-startup thesis he has argued repeatedly produces exactly this ordering as a byproduct of building many products on one data model, so the compliance benefit may be a consequence rather than a motive. Conditions also differed enormously: roughly $1.85 billion raised into a mature software market, against a company that had been building the category.

And Rippling is not a clean success

FACT Rippling conceded in writing three separate times in its 2026 Better Business Bureau responses that capabilities confirmed during the sales process failed after migration had already begun. It was recruiting foundational members of a Global Implementation team at approximately $1 billion in annual recurring revenue while selling enterprise deals, and in February 2026 launched a forward-deployed engineering organization to internalize work its own announcement described customers otherwise hiring third-party consultants to do. Conrad has publicly acknowledged discovering that the 90th percentile support experience was far worse than the median.

FACT Rippling does not appear on the IRS list of active Certified Professional Employer Organizations dated April 15, 2026. Justworks has been certified since January 1, 2017. Careful sequencing and full regulatory credentialing are separate decisions, and these two companies made them differently.

The supported lesson is narrower than the story and more useful. Compliance risk and delivery risk are different operator problems with different causes. At Zenefits, responsibility was embedded in the revenue model, regulated exposure scaled with the sales floor, and controls failed. At Rippling, regulated responsibility was layered onto a mature system of record, the compliance failure did not repeat publicly, and implementation complexity surfaced anyway, later and at far greater scale.

> **Sequencing can materially reduce one operator failure mode without eliminating the operator problem.**

06 / THE LAG

Why the apparatus tends to arrive after the revenue

Across multiple cases here, the licensing, implementation, underwriting, or exception-handling function was built after the volume that required it. The mechanism is unglamorous. Revenue-generating functions get resourced first because they are visible and measurable. Back-office functions get resourced when they break. The gap between those two moments is set by growth rate rather than by company size, which is why it shows up at a billion in revenue as readily as at fifty million.

This is a tendency and not a law, and the exceptions matter. Justworks does not show it. Deel does not appear to, though it publishes nothing that would let anyone check. Presenting the lag as universal would be wrong.

The reliable tell is documentary. In several of these cases the company was asking new hires to write the process documents for work already running at scale. Implementation checklists, underwriting playbooks, standard operating procedures, and onboarding templates normally predate volume. When a job posting asks a new hire to build them, the volume got there first.

What separated the survivors was rarely whether they had the lag. It was how the gap surfaced. In the clearest rupture cases, the operating gap became visible externally before management had demonstrated it was under control. Zenefits learned its licensing was broken from a journalist. Olive's customers reported the promised savings had not materialized before the company conceded it. Bright Health's cost structure emerged through reserve development. Internal awareness is not resolution and guarantees nothing, but the sequence in which a problem becomes public is itself informative when reading any company's current position.

---

07 / THE HUMAN TAIL

Automation removes the middle of the distribution, not the edges

The question is not whether automation works. In every case here it worked on the thing it was pointed at. The question is which work goes first, which stays human longest, and how much the labor curve actually bends once the vendor is answerable for the result.

Regulated, correctness-critical work has a long tail where the rule is ambiguous, the data is missing, or two obligations conflict. Automation handles the well-specified center of that distribution efficiently. The tail needs judgment. Because cost per item is highest in the tail, removing half the item count removes considerably less than half the labor.

CURRENT-GENERATION CASE

Klarna

FRONTIER MODELS, AND THE WALK-BACK

FACT Klarna launched an OpenAI-powered customer service assistant globally in February 2024. In its first month it handled 2.3 million conversations, work the company said was equivalent to roughly 700 full-time agents. Klarna reported resolution in under two minutes against eleven for human agents, a 25 percent drop in repeat inquiries, and a projected $40 million profit improvement for 2024. Headcount fell from roughly 5,000 to roughly 3,500 through 2024, largely by attrition under a hiring freeze.

FACT On May 8, 2025, Sebastian Siemiatkowski told Bloomberg the company had cut too far and was reopening hiring. His words: "As cost unfortunately seems to have been a too predominant evaluation factor when organizing this, what you end up having is lower quality." By June 2026 he had settled on a durable framing, that human customer service would come to be seen as a premium tier while AI handled the simplest work.

REPORTED The caveat matters and cuts against the simple reading. The assistant was deployed into a function whose quality had already been degraded by a preceding outsourcing transition, and the public record does not cleanly separate the two. Klarna was still running significant AI volume and reporting material productivity gains through late 2025. This is not a story about automation failing.

The mature comparison is instructive because nothing was walked back. TriNet launched an AI assistant in spring 2026. By its second-quarter earnings call on July 30 it reported that the assistant had addressed 50 percent of customer-initiated chat sessions and lowered service case volumes. In the same quarter the company was growing its sales force by approximately 20 percent, running more than 100 new representatives through a training program, and accelerating investment in its service model.

"AI and HR is most valuable when combined with judgment rooted in deep domain expertise and a strong service orientation."

MIKE SIMONDS, PRESIDENT AND CEO, TRINET GROUP, SECOND QUARTER 2026 EARNINGS CALL, JULY 30, 2026

Half the sessions handled, service capacity redeployed, and the organization still growing. That is what a working deployment looks like at the operator boundary. It is a real gain, correctly sized.

Deel supplies the blunt version of the same point. A company reported to have crossed a billion dollars in annual recurring revenue employs on the order of nine thousand people to hold employment liability in a hundred and fifty countries. REPORTED Those figures come from secondary sources and are unaudited, and they are doing real analytical work here, so treat them as directional. What they suggest is that fast operator scaling is survivable, and that the price is paid in headcount.

> **The recurring failure is not that automation underperforms. It is that the metrics chosen at launch do not measure what determines whether it worked.**

This finding repeats with unusual precision. Klarna measured coverage, speed, and repeat-contact rate. What ultimately determined the outcome was quality on complex and emotionally loaded cases, which those metrics did not capture. Rippling published first response time. Every 2026 customer complaint in the record concerns time to resolution and ownership. Olive AI told hospitals its technology would cut administrative spending by a factor of five, and an April 2022 Axios investigation reported that the savings calculations rested on rough estimates and were tracked only when a client asked.

Three companies, three different technologies, a decade apart. In all three, the measured variable was throughput and the failing variable was resolution.

Can good sequencing survive compressed scaling?

Sequencing has good evidence behind it. Compression does not, and pretending otherwise would be the fastest way to make this study useless.

The question is whether a company that built the infrastructure first is still protected when responsibility-bearing volume arrives as a step change rather than over years. Every counter-case in this study grew slowly. Justworks took a decade at roughly a third annually. Rippling added PEO in year four and grew the book across the following five. TriNet has had thirty years. None of them ran the compressed version, so none of them can say what happens to a company that does.

Two cases bear on it directly and neither is clean.

Flexport

INFRASTRUCTURE FIRST, THEN A SPIKE

FACT Flexport was founded in 2013 as a digital freight forwarder and licensed customs broker, which is rung-four responsibility by construction. It held the platform and the licenses well before the pandemic. REPORTED Gross revenue went from $3.3 billion in 2021 to approximately $5 billion in 2022, then fell to roughly $1.6 billion in 2023, with gross revenue down about 70 percent year over year in the first half of 2023.

FACT The company cut 20 percent of its workforce in January 2023 and a further 20 percent, roughly 600 people, beginning October 13, 2023. Dave Clark joined as co-CEO in September 2022 and resigned in September 2023, after which founder Ryan Petersen returned as CEO and rescinded dozens of outstanding offers. Petersen told FreightWaves the company had twice as many desks as it needed.

CONFOUND The revenue collapse was driven substantially by falling freight rates, which is a price effect rather than a responsibility-scaling effect. What survives the confound is narrower and still useful: the operator headcount built into the spike could not be unwound at the speed the spike reversed. Fixed operating capacity lags in both directions.

Deel is the other, and it points the opposite way. It reached rung four across a hundred and fifty countries in roughly six years without a public rupture, apparently by converting capital directly into operating headcount at a ratio no software business would recognize. That is one data point, its figures are unaudited, and one data point is not a base rate.

What would settle it is a case where a company with mature owned infrastructure took a step change in regulated volume and then published enough operating detail to see whether cost to serve, exception rate, and implementation time held. No such case was found in this research. The absence is weak evidence, because the category producing these companies is young enough that the sample may simply not exist yet.

UNRESOLVED Whether good sequencing survives compressed scaling. The honest position is that this is currently being tested rather than known, and anyone claiming a confident answer in either direction is working from something other than the historical record.

Gross margin falls.
Operating margin can rise.
Both are in the same filing.

The lazy claim is that operator businesses have service economics and are therefore worse. The best-documented case in this study says something more specific, and Shopify says it about itself.

Shopify

RUNG 1 AND RUNG 3 IN ONE COMPANY

FACT From Shopify's Form 10-K and subsequent 10-Q filings: gross profit margins on Shopify Payments, the largest driver of merchant solutions revenue, are typically lower than on subscription solutions because of the associated third-party costs, and the continued growth of merchant solutions has caused and may cause a decline in overall gross margin percentage.

FACT The same filings state the other half directly: "We view this revenue stream as beneficial to our operating margins, as Shopify Payments requires significantly less sales and marketing and research and development expenses than Shopify's core subscription business."

REPORTED Third-party analysis of the second quarter of 2026 puts merchant solutions gross margin at 38.4 percent against subscription solutions at 79.7 percent. Merchant solutions grew 37.4 percent year over year against 22.3 percent for subscription.

At Shopify, the gross-margin difference between the two revenue lines is roughly a factor of two. The point is not that operator revenue always cuts gross margin in half. It is that adding responsibility can materially alter the economics of the revenue line while still improving operating leverage elsewhere, and that a company can hold both effects at once without either one settling whether the move was correct.

What actually changes by rung is more specific than a margin. At rung three, labor intensity and cost to serve become the binding constraints, and they scale with transaction volume. At rung four, the new costs are licensing, certification, audit, controls, and legal exposure, and they behave like a fixed floor that has to be in place before the volume rather than after it. At rung five, working capital and reserving appear, and a bad year becomes a capital event rather than an income statement event.

The mature reference point is TriNet, which publishes the full picture quarterly. In the second quarter of 2026 it reported total revenue of $1.2 billion, of which professional services revenue, the fee for administering the work, was $159 million. Adjusted EBITDA was $128 million at a 10.9 percent margin. Most of the top line is pass-through. Any read of an operator's gross revenue as though it were software revenue will be wrong by close to an order of magnitude.

---

10 / BOTH DIRECTIONS

The ladder runs down, too

Almost every discussion of this transition treats it as one-way. The largest company in this study demonstrates otherwise, and did it deliberately.

Shopify announced the Shopify Fulfillment Network in June 2019, acquired 6 River Systems that year for $450 million, and acquired Deliverr in May 2022 for $2.1 billion, its largest acquisition. That is a software company climbing to rung three and administering physical logistics on merchants' behalf.

On May 4, 2023 it sold the logistics business to Flexport in exchange for roughly 13 percent of Flexport and a board seat, sold 6 River Systems to Ocado, and cut 20 percent of its workforce. The transaction completed on

June 6, 2023.

"Building logistics infrastructure is a side quest every e-commerce entrepreneur is eventually pulled into. Shopify was the perfect place to bootstrap this effort from 0 to 1 and we have done this. The next step is to take what we have and take it from 1 to N as a main quest."

TOBI LÜTKE, CEO, SHOPIFY, MEMO TO EMPLOYEES, MAY 4, 2023

FACT Shopify reported a first-quarter operating loss of $193 million on the same day, against $98 million in the comparable prior period. Shares rose sharply following the combined earnings, restructuring and logistics announcement, gaining as much as 24 to 29 percent in the following session depending on the exchange and the window measured. The three items were announced together, so the move cannot be attributed to the divestiture alone.

Harley Finkelstein's framing at the time was that the services could be delivered more effectively through integration than through ownership.

> **Descending a rung is a strategic option with a documented precedent, not an admission of failure.**

The uncomfortable symmetry is that Flexport, the receiving party, absorbed this expansion in the middle of its own contraction, and was cutting 20 percent of its own staff five months later. The work did not become easier by changing hands. It went to a party whose cost structure was built for it, which is a different claim and the one that matters.

11 / DISTRIBUTION

Once you own the outcome, the channel is part of operations

In a rung-one business, a channel produces leads and the analysis stops at cost per acquisition. Once the vendor owns the outcome, the cost of a customer includes getting their data right, setting them up, and handling their exceptions for as long as they stay. A channel that delivers customers who are already advised, already structured, and already carrying clean records reduces all three. That makes channel design an operating decision.

TriNet is the clearest instance. FACT On the July 30, 2026 earnings call, Simonds noted that more than 90 percent of small and medium businesses offering health care obtain it through a health insurance broker. TriNet's broker channel represented 32 percent of new sales in the quarter, with requests for proposals up 54 percent year over year. He described retention-based incentives to brokers as critical to alignment, and the operating design around it as putting the right people at the local level against the right broker producers, giving brokers trusted-adviser access as standard procedure, including them in renewal discussions, and assigning dedicated client service staff against the broker block rather than customer by customer. That is a delivery architecture rather than a marketing program.

Gusto shows the same mechanism outside the PEO model. REPORTED BY THE COMPANY More than 23,000 accounting firms held active partner accounts as of April 30, 2026. The program performs implementation work directly: partners receive onboarding help, free payroll transfers, automated import from prior providers, and a multi-client dashboard. The accountant becomes the first line of exception handling for their own book and holds the client relationship through the year.

Rippling's 2020 PEO launch gave brokers renewal visibility and one-click export, which reduces friction at entry and at exit. An exit path lowers the service burden of clients who should not be on the product in the first place. **Zenefits** is the negative instance: by becoming the licensed broker itself, it collapsed distribution and regulatory exposure into a single activity, so the channel could not absorb any operating load because the channel was the operating load.

The transferable principle is that in an operator business, the best channel may lower the cost of delivering the outcome as well as sourcing it. Mechanisms observed across these cases include structured data arriving at intake, a trusted adviser absorbing first-line questions, better-qualified customers, standardized implementation, easier renewals, and a cleaner path out for customers who do not fit. The limit is worth stating. TriNet expects direct sales to remain more than half of new acquisition, so the broker channel is an addition rather than a replacement, and the mechanism is strongest where the reason for purchase is itself intermediated.

---

12 / THE METRICS

What actually tells you whether the transition is working

No thresholds are offered here, because the cohort does not support them. What the cohort does support is which quantities matter, and they are consistently not the ones companies publish at the moment they cross the line.

MEASURE	WHY IT MATTERS	EVIDENCE BEHIND IT
Resolution, not first response	The published metric is usually the one the company is good at. Complaints concentrate on ownership and time to close.	Rippling's public support metric against the substance of its 2026 complaints
Coverage and accuracy, separately	Automation coverage can succeed while quality on hard cases falls. One number hides the other.	Klarna, 2024 to 2025
The 90th percentile, not the median	Operator failures live in the tail. Medians are structurally blind to them.	Conrad's own acknowledgment
Delivery headcount per unit of retained revenue	The clearest signal of whether the labor curve is bending or the book is simply being staffed.	Deel's ratio; TriNet's concurrent AI gain and headcount growth
Share of the book that has completed a full annual cycle	In any business with renewals or periodic obligations, a book that has not cycled has not been tested.	Simonds on two renewal cycles before pricing aligns with risk

MEASURE	WHY IT MATTERS	EVIDENCE BEHIND IT
Attrition decomposed by cause	Aggregate churn conflates price decisions with service failures, which have opposite remedies.	TriNet split price-driven and service-driven attrition and found them moving separately
Gross margin by revenue line, not blended	A shifting mix drags the blended number regardless of execution quality.	Shopify's own filing language on mix
Cost to serve by acquisition channel	Tests whether a channel is lowering delivery cost or only sourcing volume.	The distribution cases, as a hypothesis rather than a proven result
Data completeness at onboarding	Upstream data quality determines downstream exception rate, and is measurable before the exceptions arrive.	Inferred across the implementation cases; the weakest-evidenced item here

The pattern running through the top half of that table is the same measurement failure described earlier. Throughput is easy to instrument and gets reported. Resolution quality is hard to instrument and is what breaks.

---

13 / WHAT AI CHANGES

Separating what history establishes from what is still being proved

Much of the automation evidence in this study is a generation old. Olive AI ran robotic process automation. Babylon ran probabilistic symptom modeling built in the 2010s. Treating those as evidence about current models would be a category error, so the split has to be explicit.

What history establishes

That exception-heavy regulated work has a long tail, that the tail carries the cost, and that automation reaches the tail last. That claim is supported across a decade and across technologies, and current systems do not obviously escape it. Klarna is the strongest current-generation instance and points the same way, with the qualification that its assistant did work and the walk-back concerned scope rather than capability.

What is genuinely open

How fast the tail shrinks with systems that can read primary sources, reason across conflicting rules, cite what they relied on, and abstain when the inputs are thin. Nothing in the historical cohort had those properties. The most interesting design pattern appearing in current operator companies is not higher automation coverage but explicit abstention, versioned prompts, evaluation suites, and decision logs built so an AI-assisted output can be reproduced for an auditor. Whether that produces a materially different labor curve is not yet visible in any published operating metric this research could find.

What has not changed at all

Rungs four and five are legal and financial positions rather than technical ones. A model cannot hold a license, post a bond, or be the named party on a filing. Certification timelines, audit requirements, carrier relationships, and capital adequacy move at the speed of institutions. Whatever automation does to the cost of the work, it does nothing to the calendar of becoming permitted to do the work.

> **Automation can compress the labor. It cannot compress the institutions.**

---

14 / RUNG FIVE

The financial risk case, kept separate on purpose

Rung five applies to a minority of operators, so it is held apart rather than folded into the general argument. Where a company's balance sheet absorbs the outcome, additional mechanisms appear, and two well-documented cases sit on either side of them.

FACT Bright Health Group announced in January 2022 that it was positioned to pass one million members. It reported a loss of roughly $1.2 billion for 2021, raised $175 million in October 2022 to carry the business to profitability, then exited the individual market entirely, giving up approximately 970,000 commercial members and releasing roughly $250 million in regulated capital. That last figure measures how much balance sheet the book had been consuming without anyone having planned for it.

FACT Oscar Health faced the same mechanism in the same market in overlapping years. Its medical loss ratio deteriorated to 87.4 percent for full year 2025 and 95.4 percent in the fourth quarter alone, producing a $443.2 million net loss against a $25.4 million profit in 2024. It repriced rather than exiting, and reported second-quarter 2026 net income of $361.8 million with full-year operating earnings guidance raised to between $500 and $700 million.

What separated them was capital sized to absorb a full reset year, a market with a scheduled annual repricing mechanism, and the willingness to take the loss year rather than leave. Fourteen years passed between Oscar's founding and that result.

The general point is the one worth carrying: financial risk changes the severity and the economics of the operator transition. It does not change whether the transition has occurred. Zenefits failed at rung four carrying no claims risk at all.

---

15 / THE DECISION

Seven questions, in the order the evidence suggests asking them

Not a recommendation about whether to climb. A set of questions the cases indicate are load-bearing, arranged so the cheap ones come first.

1. **What rung are we on?** Most companies in this study crossed without deciding to. Naming the rung costs nothing and changes what counts as underinvestment.
2. **Does the system of record precede the responsibility?** The survivors had one. The clearest ruptures took the obligation first and had to acquire the data, the process, and the controls at the same time.
3. **How fast is responsibility-bearing volume growing, separately from total revenue?** This is the compression question. The historical record cannot answer it for you. It can only establish that it matters.
4. **Are we measuring resolution or throughput?** Three companies, three technologies, a decade apart, all measured the second and were broken by the first.
5. **What does the tail look like?** If nobody can produce the 90th percentile from a dashboard, customers are discovering it instead.
6. **Is the channel lowering cost to serve, or only sourcing volume?** Cost to serve by acquisition cohort answers it. Most companies do not cut the data that way.
7. **What is the off-ramp?** An exit path for customers this responsibility does not fit appears in the record as a design choice of the companies that did well, and it reduces the service burden of the worst-fitting accounts.

Two further questions sit behind these and are worth holding in mind rather than scoring. The institutional calendar for licensing, certification, and carrier relationships runs on its own clock and is the constraint automation does not touch. And descending a rung remains available, with precedent, though it is almost never in the strategy deck.

METHOD

What this study can and cannot support

Supported. That software companies do not escape operational complexity by becoming responsible for the work. That the successful cases sequenced responsibility behind a mature system of record and built the apparatus alongside the book. That automation reaches the exception tail last. That distribution and delivery are coupled once the vendor owns the outcome. That the ladder runs down as well as up. That the recurring failure is a measurement failure.

Not supported. That operator businesses necessarily have service economics, which the Shopify filings contradict directly. That every software company should climb, or should not. That current models behave like the automation in the older cases. That any specific outcome follows from occupying this structural position, since six of eleven cases here occupy it and are fine.

Unresolved. Whether good sequencing survives compressed scaling. The counter-cases all grew over years. Deel is the single partial exception and its figures are unaudited. This is the largest evidentiary gap in the study and it sits directly under the question the current generation is asking.

Selection. The cohort was built so that a majority did not rupture, because postmortems get written and quiet competence does not. Any study assembled this way still tilts toward failure and should be discounted accordingly. Three of the ruptures reached their end state in 2023 inside a severe digital health funding collapse, which is a confound and is treated as one.

Evidence standard. Every claim above is traceable to a dated public source. Where a claim rests on secondary reporting rather than filings, calls, or regulatory records, it is labeled in place rather than in a footnote. Company self-reported figures are marked as such. Contested causes are presented as contested.

PRIMARY SOURCES

TriNet Group Q2 2026 earnings call, held 30 July 2026 · TriNet Form 10-Q, quarter ended 30 June 2026 · TriNet Form 10-Q, quarter ended 31 March 2026 · TriNet Form 8-K and Form 10-Q on the Zenefits acquisition, 2022 · Shopify Form 10-K FY2025 and Form 10-Q FY2026 · Shopify press release on completion of the Flexport sale, 6 June 2023 · Justworks Form S-1, December 2021 · IRS list of active Certified Professional Employer Organizations, report date 15 April 2026 · California Department of Insurance press release, 9 May 2018 · SEC administrative proceeding 33-10429, 26 October 2017 · Oscar Health Form 8-K, February and August 2026 · BuzzFeed News reporting on Zenefits, November 2015 and February 2016 · Bloomberg interview with Sebastian Siemiatkowski, 8 May 2025 · TechCrunch interview with Parker Conrad, 20 October 2022 · Rippling PEO broker launch materials, November 2020 · Axios investigation of Olive AI, April 2022 · Healthcare Dive and Fierce Healthcare on the Olive wind-down, 1 November 2023 · Forbes on the Babylon US shutdown, 9

August 2023 · Fierce Healthcare and Healthcare Dive on Bright Health exits, 2022 · Supply Chain Dive, FreightWaves and CNBC on Shopify and Flexport, May and October 2023 · Gusto partner program documentation, April 2026

PASTBEHAVIOR · RESEARCH · AUGUST 2026 · PREPARED FROM PUBLIC SOURCES