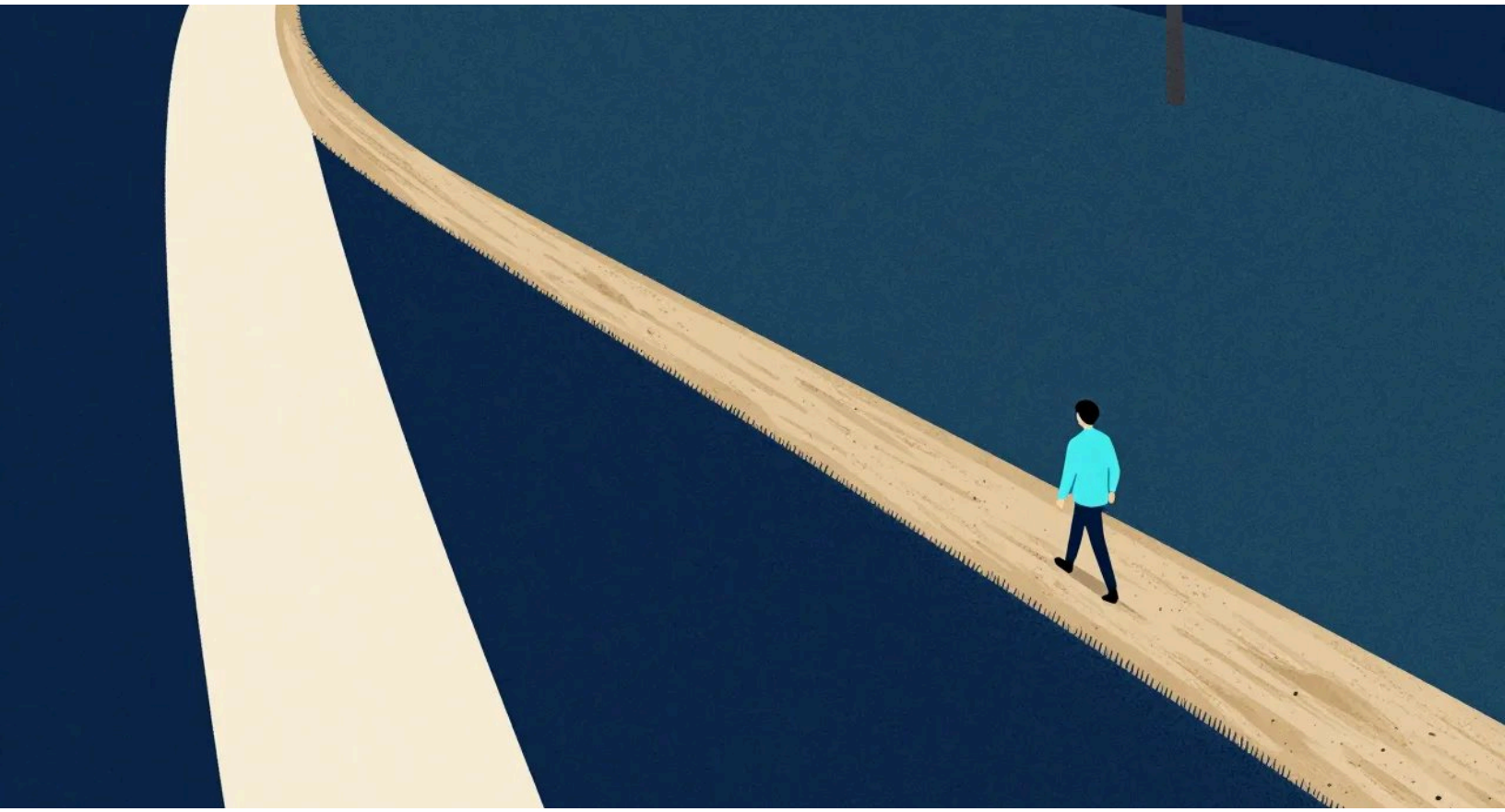


EXECUTIVE PAPER · DISCOVERY

Automate the Process You Actually Have

Automation and AI copy whatever process they are given at machine speed. Executives who fund them against the process manual instead of the work as people really do it are paying to scale the gap between the two.

AUTHOR

Chris Watkins

SERVICE

Discovery

FOR

COO, CEO, CFO, CIO · 12 min read

AT A GLANCE

The problem. Most automation business cases describe the process as written in a policy, a flowchart or a system specification. The real work includes exceptions, judgment calls and workarounds that nobody wrote down. When software replaces that work, the undocumented parts either break or get decided wrongly at scale, as Australia found in 2025 when an automated system cancelled payments for more than 900 job seekers under a law it had never been updated to follow.

Why it persists. Discovery looks like delay, and the people who know the exceptions are rarely in the room when the business case is written. Vendors price the happy path. Nobody owns the gap between the rule and the code.

The recommendation. Make a measured baseline of the work as done a condition of funding. Build it from interviews, direct observation and system event logs. Ask for cycle time, exception rate, rework rate and a list of workarounds before any money moves, and keep measuring the same numbers after go-live.

The Decision in Front of You

Every operations leader now has a queue of automation proposals: bots for accounts payable, AI agents for customer service, rules engines for eligibility and claims. Each comes with a process map and a savings figure. The decision is whether to fund against that map or to first require evidence of how the work is performed today.

Automation that matches the real process removes cost and delay. Automation built on an idealized version handles the easy cases, pushes the hard ones to people who no longer have the time or context, and at worst makes the wrong decision thousands of times before anyone notices. The cost of finding out is a few weeks of discovery. The cost of not finding out has, in public cases, run to hundreds of millions of dollars and a government system taken offline for repair.

The rule this paper argues for: no automation or AI funding for a workflow without a measured baseline of the work as done, which later becomes the yardstick for whether the project worked.

What's Happening

When the rule changes and the code does not

Australia's Targeted Compliance Framework, 2022 to 2025. Australia's employment department runs an automated system that enforces the obligations job seekers must meet to receive income support. In 2022 the law was amended so that a human decision-maker had to consider an individual's

circumstances before cancelling a payment [1][2]. That discretionary step was never built into the system. When a failure was recorded, the payment was cancelled automatically [2].

The Commonwealth Ombudsman reported in August 2025 that these cancellations were unlawful, in a report titled "Automation in the Targeted Compliance Framework: When the Law Is Changed but the System Isn't" [3]. The error ran from April 2022 until September 2023, when outside lawyers advising on an unrelated matter noticed it [2]. The department took almost ten more months to pause cancellations, which the Ombudsman called unacceptable [1]. The report found the people who drafted the law change assumed it needed no rework of systems or processes, and that the drafters were wrongly assumed to understand how the automated system worked [2]. A separate review found the system's IT had become unstable and identified five errors dating back to 2018 [2]. In mid-2026 the department told a Senate committee the system would stay offline for several months for changes to IT systems, procedures and training, with 964 job seekers affected [4].

Even the review that confirmed the problem needed correcting: Deloitte agreed in October 2025 to partially refund its roughly A$440,000 fee after its report on the framework was found to contain fabricated references, produced partly with a generative AI model [5].

The price of encoding a misunderstanding

Robodebt, settlement approved 2026. Between 2015 and 2019 the Australian government raised welfare debts by averaging annual tax office income across fortnights, although the social security law assesses income in the fortnight it was received [6]. The Royal Commission found in 2023 that the scheme was neither fair nor legal [6]. In June 2026 the Federal Court approved an additional A$548.5 million settlement, including A$475 million in compensation for about 125,000 registered claimants, on top of an earlier settlement reached in 2021 [7]. About 450,000 people are estimated to have been affected [7].

The Post Office Horizon inquiry, 2025. Volume 1 of the inquiry's final report, published in July 2025, describes hundreds of postmasters, managers and assistants wrongly accused of theft and fraud because of faults in the Horizon system [8]. Among the questions the inquiry examined was how the Post Office handled discrepancies that postmasters reported [8]. Some people were held liable for small sums the system said were missing, and others were wrongly imprisoned [8]. The discrepancy reports were the one place where the gap between the system and the branch showed up, and the process treated them as losses to recover.

Klarna, 2025. In February 2024 Klarna said its AI assistant was handling 75% of customer chats [9]. In May 2025 its chief executive told Bloomberg the AI-focused path had not been the right one and began recruiting human agents again, so that customers could always reach a person [9]. Klarna has not published which kinds of cases went wrong. The 75% figure measured how many chats the assistant took, which says nothing about how it handled the ones that needed judgment.

The industry pattern

The failure rate for automation programs has been high for a decade. EY reported in 2016 that as many as 30 to 50% of initial robotic process automation (RPA) projects fail, based on the projects it was called in to rescue [10]. The figure is old and comes from a firm that sells the remedy, so treat it as an indication of scale. Deloitte's 2022 survey of 479 executives in 35 countries found process fragmentation was the top barrier to scaling automation in each of its previous four surveys, and 52% cited an inability to change processes or ways of working [11]. Only 23% of those organizations were using process mining, the discipline of reconstructing processes from system logs, although 82% agreed it produced better outcomes [11].

AI agents are inheriting the same problem with a larger budget. Gartner predicted in June 2025 that more than 40% of agentic AI projects will be canceled by the end of 2027, citing rising costs, unclear business value and inadequate risk controls [12].

Why the Usual Approach Fails

The documented process is a minority of the real one

Process mining research shows how far real work drifts from its description. The 2019 Business Process Intelligence Challenge released the purchase-to-pay event log of a multinational paints and coatings company: 251,734 purchase order items, 1.6 million events and 42 activities, with every item assigned to one of four documented matching flows [13]. One published analysis of that log found 14,333 distinct paths through those four flows [14].

The labels did not describe the work either. In the flow defined as "invoice before goods receipt," 91% of cases actually had the goods receipt first or at the same moment [14]. In 21% of those cases staff still performed a payment-block removal step the case did not need, which raised the median time from order to payment from 72 to 91 days [14]. A bot built from the flow definitions would have automated the wrong sequence for nine cases in ten.

Exceptions are where the cost lives

This is an old finding. In Michael Hammer's 1990 account of Ford's accounts payable department, more than 500 people spent most of their time on mismatches between purchase orders, receiving documents and invoices [15]. Ford's first plan, new computer systems on the existing process, aimed for a 20% headcount cut. Redesigning the process so that mismatches could not arise produced a 75% reduction and cut the items matched from 14 to three [15]. The source is 36 years old. The same economics apply to a bot or an agent today, because the routine path is usually the cheapest part of a workflow already.

Three habits keep the gap hidden

Business cases are written from the policy, the vendor demo or a workshop with managers, all of which describe work as imagined. The clerk who keeps a spreadsheet of vendors whose invoices always need correcting is rarely interviewed.

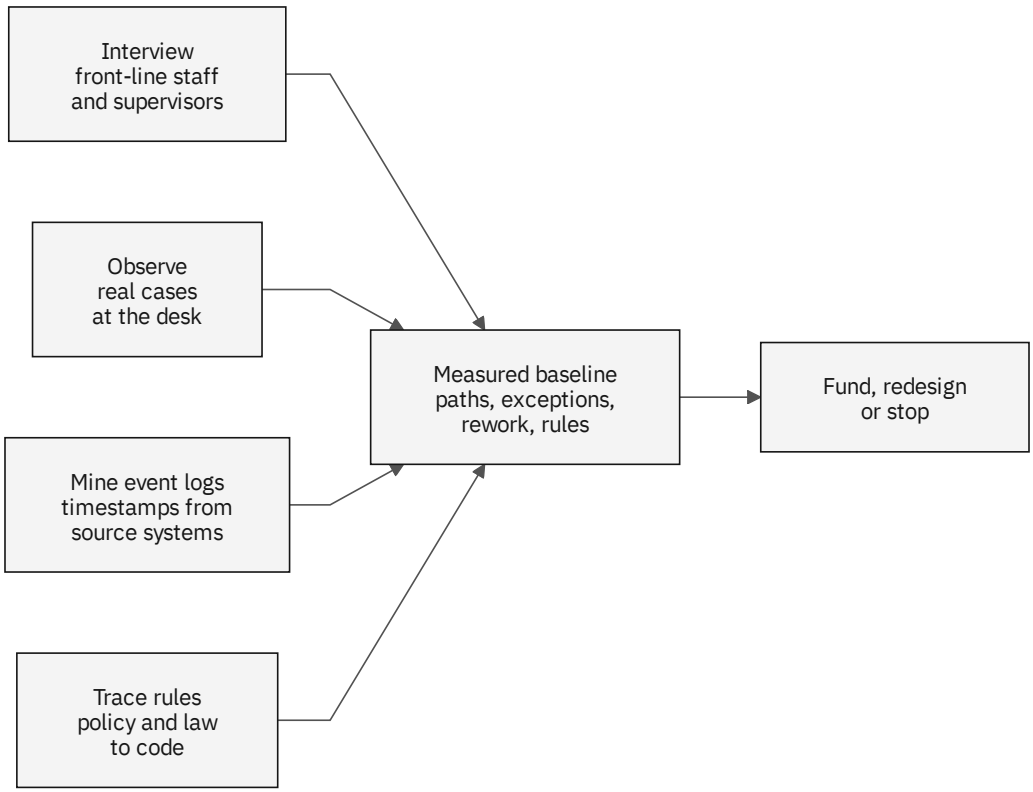
Rules and code drift apart without an owner. The Targeted Compliance Framework failed at exactly this joint: a law changed, and nobody was responsible for checking that the system changed with it [2][3].

Success is measured as hours removed from the happy path. Few business cases count the exceptions pushed back to staff, the rework they create or the wrong decisions made at volume. The NIST AI Risk Management Framework asks organizations to understand and document the intended purposes, context-specific laws and deployment settings of an AI system, and to define processes for human oversight [16]. Most organizations cannot meet that expectation for a workflow they have never measured.

The Framework: The Work-as-Done Baseline

Before funding, require a baseline built from four kinds of evidence. Each catches what the others miss.

Exhibit 1. The Work-as-Done Baseline: four evidence sources, one measured picture, then the funding decision



Source: author's framework, drawing on process mining practice [17] and NIST AI RMF MAP 1 and MAP 3.5 [16].

Interview. Talk to the people who do the work and the people who handle its failures, separately from their managers. Ask what they do when the system is wrong, which cases they set aside and what they keep outside the system.

Observe. Sit with staff for a sample of real cases, including the slow ones. Observation finds the copying between systems and the phone calls that no log records.

Mine the logs. Most ERP, case management and ticketing systems record a timestamp for every step. Process mining reconstructs every path cases actually took and compares them with the documented model, a technique called conformance checking [17]. This gives volume and distribution, which interviews cannot.

Trace the rules. For any workflow that applies policy, law or contract terms, list each rule and the place in the system that implements it, with a named owner. This is the check the Targeted Compliance Framework never had [2].

The baseline ends in one of three decisions. Fund automation of the paths that are stable and high-volume. Redesign first where exceptions dominate, as Ford did. Stop when the process depends on judgment the organization cannot yet specify.

The numbers to ask for

Exhibit 2. Seven baseline measures an executive should see before approving automation spend

Measure	Definition	Why it matters
Volume	Cases per month, by path	Sizes the prize and shows which paths are worth automating
Cycle time	Median and 90th percentile, start to finish	The 90th percentile shows where exceptions sit; the median alone hides them
Path concentration	Share of cases on the most common path	Low concentration means the "standard process" is a minority, as in the 2019 log [14]
Exception rate	Share of cases that leave the standard path or need manual handling	The ceiling on what automation can remove without redesign
Rework rate	Share of cases where a step repeats, a record is reversed or a case reopens	Rework automated is rework multiplied
Workaround inventory	Spreadsheets, side systems and manual checks found in interviews and observation	Each one is a rule the automation will not know
Rule traceability	Share of governing rules mapped to a system implementation and an owner	Untraced rules fail silently when policy or law changes [3]

Source: author's framework. Measures are derivable from event logs [17] plus interviews and observation.

The same seven numbers become the acceptance test after go-live. Falling cycle time with a rising exception rate means cost moved to another team.

Costs, Risks, and Tradeoffs

Exhibit 3. Tradeoffs of requiring a work-as-done baseline

Area	Cost or risk	Mitigation
Time	Four to eight weeks of discovery before build	Run discovery on the next candidate while the current build proceeds
Staff time	Front-line staff spend hours in interviews and observation	Schedule around peaks; pay for backfill on small teams
Log access	Event data may be incomplete, scattered across systems or restricted	Start with the system of record; fill gaps with observation and sampling
Uncomfortable findings	The baseline exposes non-compliance, side systems and poor controls	Agree up front that findings are for redesign, and route control failures to risk owners
Tooling	Process mining software adds license cost	Small workflows can be mined with open-source tools or a spreadsheet of timestamps
When this is wrong	A low-volume, low-risk task with a short, stable path, such as moving a monthly report between two systems, costs less to automate and fix than to study	Use a one-page baseline: volume, one interview, one observed run

Source: author's analysis.

How to Act

Exhibit 4. A one-year roadmap

Window	Actions	Owner	Milestone
First 30 days	List every automation and AI initiative funded or proposed, with the workflow each touches. Pick the two with the largest spend or the highest decision risk. Name a process owner for each.	COO with CIO	Inventory complete; two workflows, two named owners
First 30 days	Add a baseline requirement to the investment approval template: the seven measures in Exhibit 2.	CFO	Template revised; no new automation funding approved without it
First 90 days	Run the full baseline on both workflows: interviews, observation of at least 30 real cases each, event log extraction, rule tracing.	Process owners with internal audit or an operations analyst	Both baselines delivered with all seven measures
First 90 days	For any workflow applying law, policy or contract terms, assign an owner for each rule and a change trigger linking policy changes to system changes.	General counsel with CIO	Rule register in place; every rule mapped to code and an owner
First 365 days	Re-measure the seven numbers quarterly on every automated workflow. Retire or redesign any automation whose exception or rework rate rose after go-live.	Process owners, reported to the executive committee	Quarterly report covers all automated workflows; exception and rework trends shown against baseline
First 365 days	Extend baselines to all workflows above a spend or risk threshold set by the executive team.	COO	All in-scope workflows have a current baseline less than 12 months old

Source: author's recommendation.

Questions to Ask Your Team

1. For the workflow we are about to automate, what share of last quarter's cases followed the documented process from start to finish?
2. What is the 90th percentile cycle time, and what happens to the cases that take that long?
3. Who on the front line was interviewed for this business case, and what workarounds did they describe?

4. Which spreadsheets or side systems does this process depend on today, and what happens to them after go-live?
5. Which rules from policy, law or contract does this automation apply, and who updates the system when one of them changes?
6. Where do exceptions go after automation, and has that team's workload been costed?
7. What exception and rework rates will tell us, six months after launch, that this project failed?

Closing

Automation repeats the process it was given, faster and at greater volume. When that process is the version in the manual, the automation repeats the manual's errors and drops the judgment that staff were quietly supplying. Require a measured baseline of the work as done before you fund automation or AI on any workflow. Measure exceptions, rework and the rules that govern each decision, assign an owner for every gap, and judge the project afterward against the same numbers.

Sources

[1] The Canberra Times. "Unlawful welfare cancellations: Ombudsman report findings." Coverage of the Commonwealth Ombudsman's report. August 6, 2025. https://www.canberratimes.com.au/story/9033710/unlawful-welfare-cancellations-ombudsman-report-findings/.

[2] iTnews. "Job seekers had payments cancelled unlawfully by gov IT system." Coverage of the Commonwealth Ombudsman's report and the Deloitte review. August 6, 2025. https://www.itnews.com.au/news/job-seekers-had-payments-cancelled-unlawfully-by-gov-it-system-619327.

[3] Austaxpolicy. "Commonwealth Ombudsman finds cancellation of income support payments under Target Compliance Framework unlawful." Summary of the Ombudsman's report "Automation in the Targeted Compliance Framework: When the Law Is Changed but the System Isn't." 2025. https://www.austaxpolicy.com/news/commonwealth-ombudsman-finds-cancellation-of-income-support-payments-under-target-compliance-framework-unlawful/.

[4] AAP News. "System that cancelled welfare payments taken offline." 2026. https://aapnews.aap.com.au/news/system-that-cancelled-welfare-payments-taken-offline.

[5] The Register. "Deloitte refunds Aussie gov after AI fabrications slip into $440K welfare report." October 6, 2025. https://www.theregister.com/2025/10/06/deloitte_ai_report_australia/.

[6] JURIST. "Australia Robodebt royal commission hands down findings, recommends referrals for prosecution." Coverage of the Royal Commission into the Robodebt Scheme report. July 2023. https://www.jurist.org/news/2023/07/australia-robodebt-royal-commission-hands-down-findings-recommends-referrals-for-prosecution/.

[7] AAP News. "Australia's biggest class action settlement approved." June 2026. https://aapnews.aap.com.au/news/australia-s-biggest-class-action-settlement-approved.

[8] Post Office Horizon IT Inquiry. "Volume 1 of the Post Office Horizon IT Inquiry's final report: Rapid Read." July 2025. https://www.postofficehorizoninquiry.org.uk/sites/default/files/2025-07/Post_Office_Horizon_Inquiry_Volume_1_Rapid_Read.pdf.

[9] Tech.co. "Klarna Reverses AI Customer Service Overhaul." Coverage of the chief executive's Bloomberg interview. May 2025. https://tech.co/news/klarna-reverses-ai-overhaul.

[10] EY. "Get ready for robots: Why planning makes the difference between success and disappointment." 2016. https://www.eyfinancialservicesthoughtgallery.ie/wp-content/uploads/2016/11/ey-get-ready-for-robots.pdf.

[11] Deloitte Insights. "Intelligent automation 2022 survey results." Survey of 479 executives in 35 countries. 2022. https://www.deloitte.com/us/en/insights/topics/talent/intelligent-automation-2022-survey-results.html.

[12] The Star (Reuters). "Over 40% of agentic AI projects will be scrapped by 2027, Gartner says." Coverage of Gartner press release. June 25, 2025. https://www.thestar.com.my/tech/tech-news/2025/06/25/over-40-of-agentic-ai-projects-will-be-scrapped-by-2027-gartner-says.

[13] van Dongen, B. "BPI Challenge 2019." Dataset, Eindhoven University of Technology, 4TU.Centre for Research Data. DOI 10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1. 2019. https://research.tue.nl/en/datasets/bpi-challenge-2019/.

[14] Porter, A., Masse, D., Visatemongkolchai, N., Seneviratne, J., Deokule, T., and Hartman, N. "Balancing Efficiency and Risk in Procure to Pay: Safely Realizing Cost Savings Using Process Mining Techniques." BPI Challenge 2019 submission, CKM Analytix. 2019. https://ckmnet-website-blog.s3.amazonaws.com/CKM_BPI_Challenge2019_FINAL+(SUBMIT).pdf.

[15] Hammer, M. "Reengineering Work: Don't Automate, Obliterate." Harvard Business Review, July to August 1990. https://folk.idi.ntnu.no/thomasos/paper/hammer_reengineering.pdf.

[16] National Institute of Standards and Technology. "Artificial Intelligence Risk Management Framework (AI RMF 1.0)." NIST AI 100-1, Table 2, MAP 1.1 and MAP 3.5. 2023. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[17] van der Aalst, W., et al. "Process Mining Manifesto." IEEE Task Force on Process Mining. 2011. https://www.win.tue.nl/ieeetfpm/shared/process_mining_manifesto-small.pdf.

The views in this paper are the author's own and do not represent any employer. Scenarios are hypothetical. This paper is not legal, financial, or compliance advice.