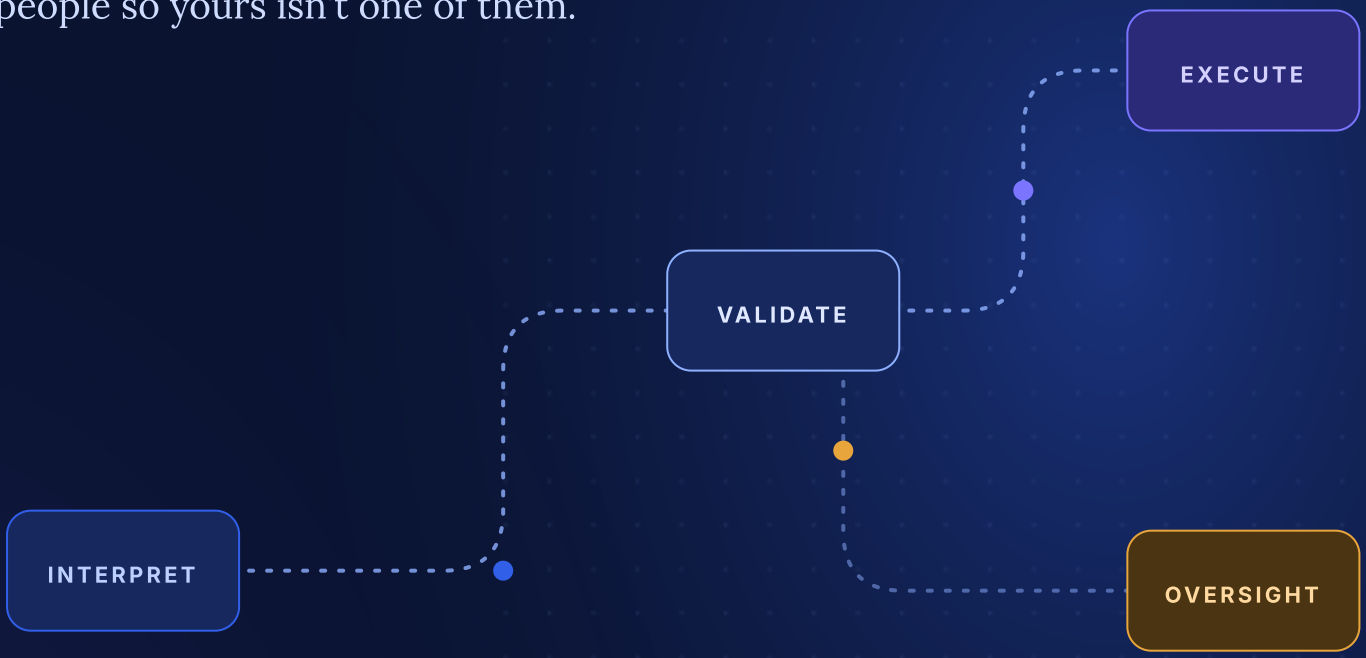
LTS | ENTERPRISE AUTOMATION

WHITEPAPER · NOVEMBER 2026

AI AGENTS VS. RPA · A DECISION PAPER FOR CEOS, CIOS AND CTOS

# Before You Scale AI Agents

Gartner expects more than 40% of agentic AI projects to be cancelled by the end of 2027, on cost, unclear value and weak controls. Here is how to place agents, RPA and people so yours isn't one of them.




LTS · Leo TechnoSoft
www.leotechnosoft.net

EXECUTIVE SUMMARY

# The question is no longer what to automate. It is what you are prepared to let software decide.

For a decade, automation meant scripting what people already did. Agents now interpret documents, weigh context and choose actions. That is a different class of capability — and a different class of risk. Choosing between "RPA" and "AI agents" frames the decision at the wrong altitude.

**88%**
of organizations use AI in at least one function

**62%**
are at least experimenting with AI agents

**~6%**
attribute 5%+ of EBIT to AI and report significant value

Source: McKinsey, *The State of AI in 2025* (Nov 2025; 1,993 respondents, 105 countries).

## Three conclusions

1 **The constraint is redesign, not adoption.**
Nearly every enterprise is using AI; very few are earning from it. Value follows leaders who rebuild the workflow around the technology rather than bolting it onto the old one.

2 **No single technology should own a process end to end.**
Agents interpret. Deterministic automation executes. People decide where the consequence of error is material. Each step gets the minimum flexibility it needs — and the maximum control it can justify.

3 **Autonomy must be earned, not granted.**
Start every agent one rung below where you hope it will end up. Promote it on evidence — accuracy, exception rates, override rates — not on a vendor demonstration.

WHAT WE RECOMMEND

① Audit the estate: find bots that break often or queue exceptions.

② Decompose two workflows by step, not by tool.

③ Pilot agents **upstream** of existing RPA, at low autonomy.

④ Set an autonomy ladder and named owners before go-live.

⑤ Fund on run-cost and error-cost, not build-cost alone.

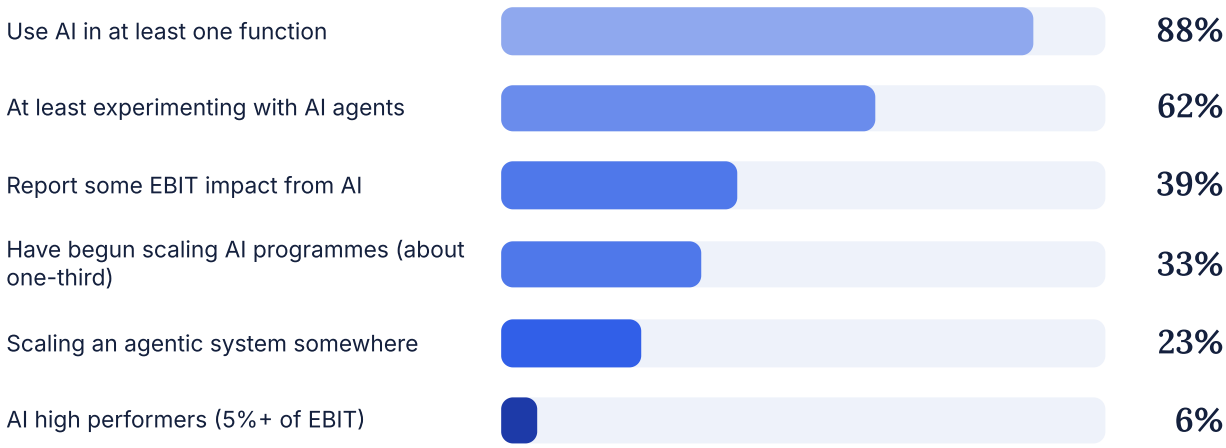
⑥ Treat the workforce transition as part of the design.

01 · THE ADOPTION–VALUE GAP

# Technology is arriving faster than the operating model.

The headline numbers on AI adoption are impressive. The numbers beneath them are the ones that matter to a board. Each bar below is a share of the same survey population; the drop from top to bottom is the distance between *using* AI and *being changed by it*.

SHARE OF ORGANIZATIONS, MCKINSEY STATE OF AI 2025




## What the gap tells us

Experimentation is cheap and spreads quickly. Scaling is neither. It requires clear process ownership, integration into systems of record, controls that satisfy risk functions, and a workforce that trusts the output. McKinsey's high performers are nearly three times as likely to have fundamentally redesigned individual workflows.

That is the central argument of this paper: the choice of automation *pattern* matters more than the choice of automation *product*.

ANALYST CAUTION

**40%+**

of agentic AI projects will be cancelled by the end of 2027, Gartner forecasts, citing escalating costs, unclear business value and inadequate risk controls.

Of the thousands of vendors marketing "agentic" products, Gartner estimates only about **130** offer genuine agentic capability — a practice it calls *agent washing*.

LOOKING AHEAD

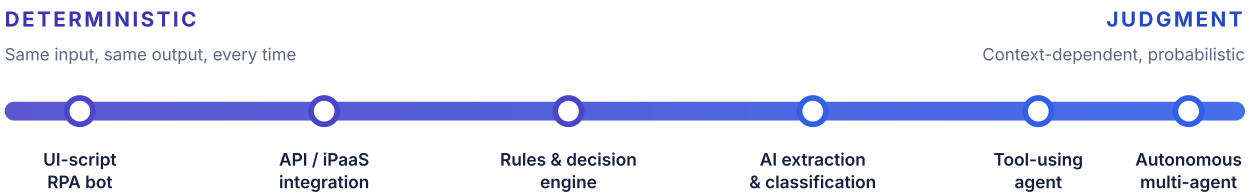
Gartner expects about one-third of enterprise applications to embed agentic AI by 2028, and 15% of day-to-day work decisions to be taken autonomously, from effectively zero in 2024. Decisions are moving to software whether or not governance has caught up.

Sources: McKinsey, The State of AI in 2025 (5 Nov 2025). Gartner press release, 25 Jun 2025, as reported by CRN Asia. Percentages are shares of all survey respondents. McKinsey reports that most respondents with EBIT impact attribute less than 5% of EBIT to AI.

02 · TWO MACHINES, TWO LOGICS

# RPA and AI agents are not rivals. They sit at opposite ends of one spectrum.

RPA follows instructions exactly. An agent pursues a goal and chooses its own path. Every real enterprise process contains steps that suit each — and steps that suit neither without a person involved.




Auditability and predictability fall to the right; tolerance for messy input and change rises.

| DIMENSION | RPA / DETERMINISTIC | AI AGENTS |
|---|---|---|
| **How it decides** | Fixed rules and scripted steps. No discretion. | Interprets intent, selects tools and sequence to reach a goal. |
| **Input tolerance** | Structured, consistent data and screens. | Free text, PDFs, email, voice, ambiguous documents. |
| **Response to change** | Stops or errs when a screen or schema shifts. Needs repair. | Often adapts to variation — but may adapt *wrongly*, and silently. |
| **Auditability** | Fully traceable; behaviour reproducible from the script. | Requires logged inputs, model version, reasoning trace and confidence to reconstruct. |
| **Cost behaviour** | Low and predictable per transaction once built; maintenance is the hidden cost. | Variable per-case inference cost; evaluation, monitoring and review are the hidden cost. |
| **Characteristic failure** | Loud: the bot breaks and a queue builds. | Quiet: a plausible but wrong action completes. |

> RPA fails loudly. Agents can fail quietly. Your control model must be designed for the failure mode, not the brochure.

03 · REALITY CHECK

# Where the "agents replace RPA" story breaks down.

There is a genuine problem behind the narrative. Bots that depend on screen scraping break when an application's interface changes, and the repair work competes with new automation for the same engineers. Agents that read intent and call APIs can remove much of that fragility. But three claims travel further than the evidence supports.

CLAIM 1

"Agents make RPA obsolete."

**Deterministic execution is still the right tool for a large class of work.**

High-volume, zero-variance transactions — especially into legacy or mainframe systems that expose no API — are cheaper, faster and easier to audit when a script does exactly what it is told. Replacing a reliable bot with a reasoning model adds cost and variance where none was needed. **The better pattern: let the agent decide, let the bot do.**

CLAIM 2

"Self-healing means no maintenance."

**Maintenance does not vanish. It moves.**

Script repairs give way to new work: prompt and tool-definition changes, evaluation suites, model-version drift, guardrail tuning and review queues. That work is real and needs an owner. Compare the *total* run burden of both approaches on your own workflows rather than assuming the new one is lighter.

CLAIM 3

"More autonomy means more return."

**Return comes from fewer exceptions and cheaper errors — not from removing people.**

Unbounded autonomy raises the cost of a wrong action. The strongest business cases come from agents that resolve routine exceptions quickly, with consequential decisions routed to a person who has the full picture in front of them.

> "Most agentic AI propositions lack significant value or return on investment (ROI)."
>
> Anushree Verma, Senior Director Analyst, Gartner (June 2025)

A NOTE ON BENCHMARKS

Widely circulated figures on RPA maintenance burden and agent ROI vary enormously by source, and most come from vendors. We treat them as hypotheses to test, not as inputs to a business case. Page 12 shows which four numbers to baseline in your own estate first.

04 · A PRACTICAL ASSESSMENT

# Five questions that decide the pattern for every step.

Run each workflow through these questions *step by step* — not process by process. The answers determine which technology fits, how much autonomy it receives and where a person must stand.

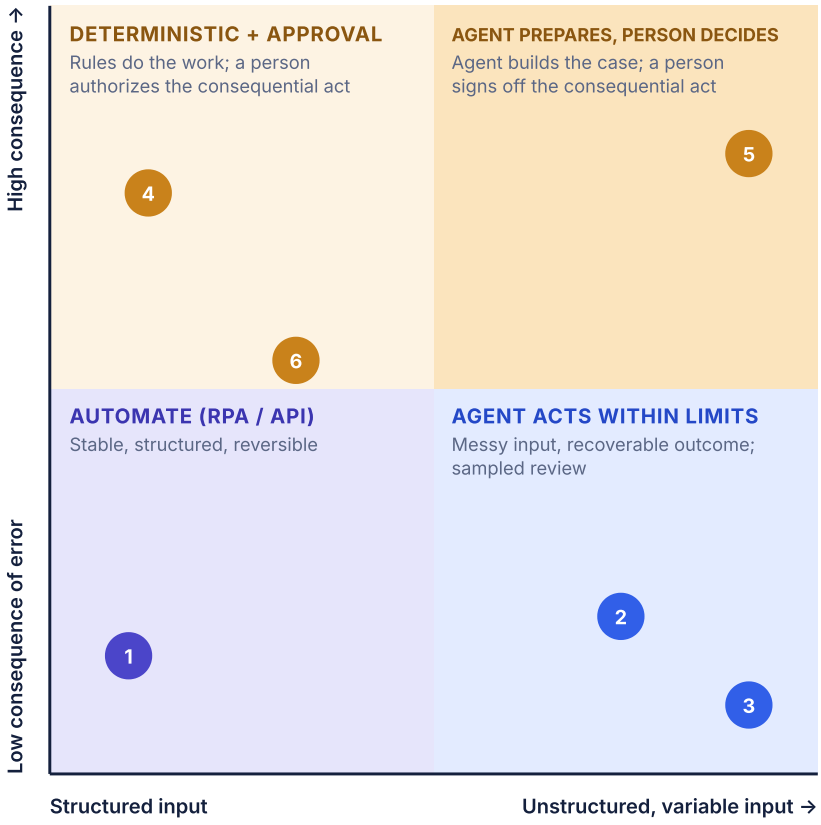
| | | |
|---|---|---|
| 1 | **Inputs & context**<br>How varied is the information? | Are inputs structured and consistent, or does the task require interpreting varied information across sources? **Variable input points toward interpretation (AI); stable input toward rules.** |
| 2 | **Rules & exceptions**<br>Can the logic be written down? | If steps and decision rules can be made explicit, encode them. If the exception rate is high or the rules conflict, that is where agents earn their cost — **with a defined escalation path.** |
| 3 | **Impact & reversibility**<br>What does a wrong action cost? | A mis-filed document is recoverable; a released payment, a denied claim or a changed entitlement may not be. **The less reversible the action, the more approval it needs — whatever technology performs it.** |
| 4 | **Access & evidence**<br>What can it touch, and what is recorded? | Define systems and records in scope, the narrowest permissions that work, and the logs a reviewer or regulator would need. **If you cannot reconstruct a decision, you cannot defend it.** |
| 5 | **Ownership & change**<br>Who answers for it? | Name the process owner, control owner and technical owner — and how changes to prompts, models, rules and scripts are tested and approved. **An automation without an owner is a liability with a license fee.** |

Questions 1–2 choose the technology.
Questions 3–5 choose the control model.

Most failed pilots answered the first two and skipped the rest. The next page shows how the answers translate into a placement.

05 · PLACEMENT MAP

# Two variables place most steps: input mess and error cost.




## ILLUSTRATIVE PLACEMENTS

**1 Batch entry into a legacy system.**
Fixed format, no API, easily reversed. A bot is the right tool.

**2 Invoice exception triage.**
Variable documents; agent classifies and resolves within tolerance.

**3 Inbound email routing.**
High variability, cheap to correct. Ideal first agent pilot.

**4 Payment release above threshold.**
Rules validate; a named approver authorizes.

**5 Adverse-media / KYC review.**
Agent assembles evidence; analyst decides and signs.

**6 Access-rights changes.**
Policy rules plus approval; agents never self-grant privilege.

### Read the map by quadrant, not by tool

The same bot can sit in two quadrants if you change the surrounding controls. A payment script is safe at low value and unacceptable at high value without an approval step.

### Placements move over time

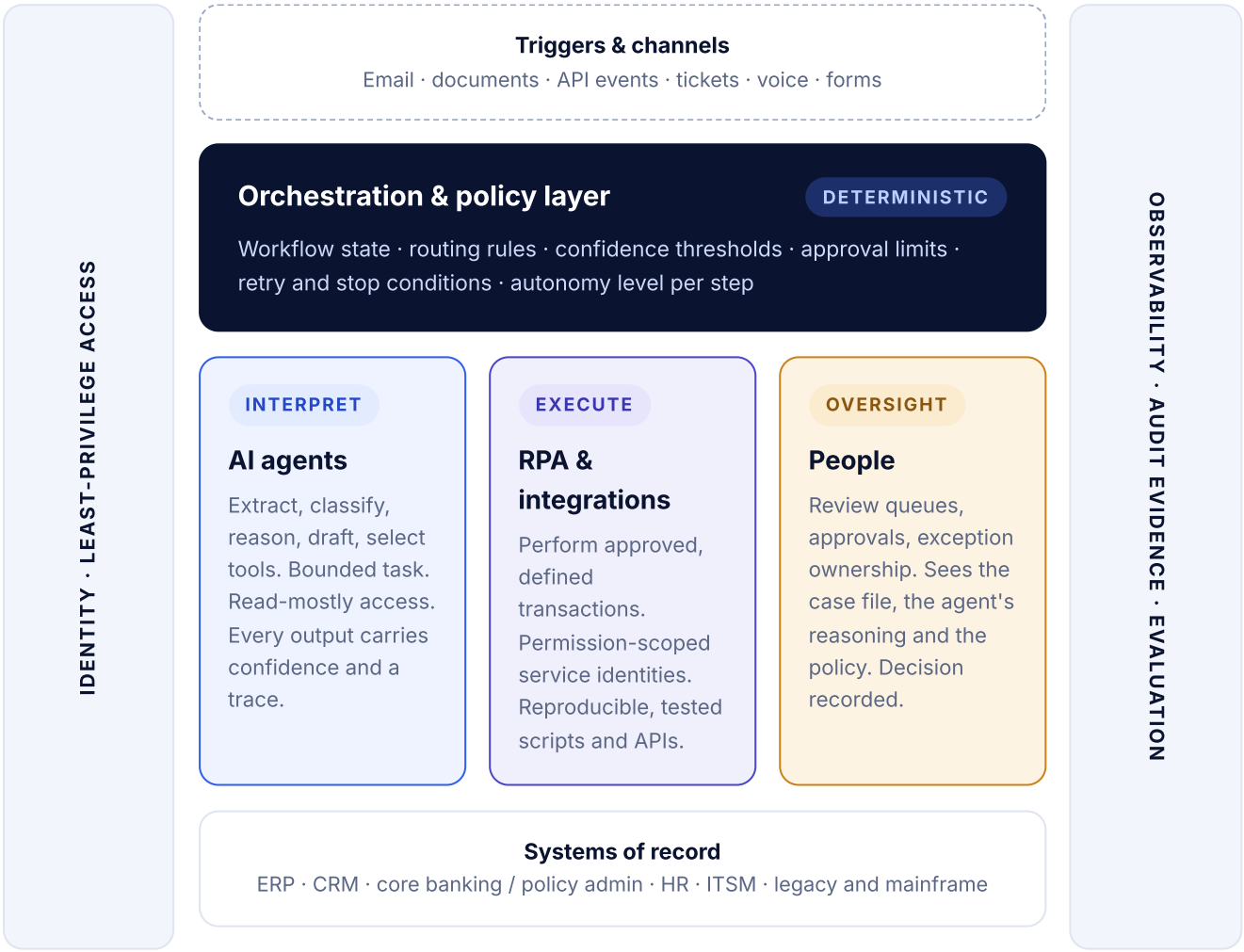
As evidence accumulates, steps migrate right and down — from "proposes" to "acts within limits." Moving a step is a governance decision, not a configuration change.

Illustrative positions only. Place your own workflows using the five questions on page 6 and your organization's risk appetite.

06 · REFERENCE ARCHITECTURE

# Put the orchestrator in charge — not the agent.

The pattern that scales has a deterministic coordinating layer that owns workflow state, policy and routing. Agents, bots and people are *workers* it calls on. This keeps the audit trail, permissions and escalation logic in one place that risk and audit teams can inspect.




## Four design rules

**1 · Agents request; they do not write.**

Changes to systems of record flow through scoped tools or bots the orchestrator controls — never through an agent holding broad credentials.

**2 · Every decision leaves evidence.**

Log the inputs, model and prompt version, output, confidence, who approved and what changed. Retain according to policy.

**3 · Confidence is a routing input.**

Below-threshold, conflicting or out-of-policy cases go to a named owner, never to a retry loop.

**4 · Build a stop switch.**

Any step must be pausable, rolled back or reverted to manual — with the fallback process rehearsed, not documented and forgotten.

07 · THE OPERATING PATTERN IN PRACTICE

# One workflow, five steps, three kinds of worker.

Consider accounts payable exceptions — the invoices that do not match their purchase order. Most are routine; a few are fraud signals. A single technology cannot treat both well.

| | | |
|---|---|---|
| 01 INTERPRET | **Read the vendor email and invoice PDF**<br>Agent extracts PO number, quantities and prices, identifies why the match failed (price variance, short delivery, duplicate, missing PO) and drafts a case summary. | **Owner**<br>AP process owner |
| 02 VALIDATE | **Check against policy — with rules, not judgment**<br>Three-way match tolerance, vendor master data, duplicate detection, approval limits. Deterministic and reproducible. | **Owner**<br>Finance control owner |
| 03 ROUTE | **Send each case to the right place**<br>Within tolerance → proceed. Outside tolerance → AP analyst with the case file. **Any change to vendor bank details → treasury control with call-back verification. Never automated.** | **Owner**<br>AP analyst / Treasury |
| 04 EXECUTE | **Post the approved correction**<br>An API call or bot posts to the ERP under a scoped service identity. Terminal steps remain deterministic. | **Owner**<br>Technology owner |
| 05 VERIFY | **Record, sample and learn**<br>Outcomes and overrides logged; a sample is reviewed weekly; override patterns feed back into prompts and rules. | **Owner**<br>Internal control |

### WHAT CHANGES FOR THE BUSINESS

- Analysts open a prepared case, not a raw email thread.
- Routine variances clear without a queue.
- Fraud-adjacent cases get *more* human attention, not less.

### WHAT CHANGES FOR RISK

- Every automated decision has a reproducible rule check.
- Agents never hold payment or master-data write access.
- A single control point can halt the flow.

Illustrative example. Tolerances, approval limits and owners must be set by the organization's finance and risk functions.

# Automation succeeds at the speed people trust it.

Every pilot that stalls does so for one of two reasons: the controls were not credible to risk, or the tool was not credible to the people using it. Both are design failures. Human oversight is therefore not a brake on automation. It is the mechanism that allows autonomy to expand.

TODAY
**Exception analyst**
Re-keys data, chases missing information, clears queues.

→

TOMORROW
**Case owner & agent supervisor**
Judges prepared cases, handles what agents cannot, and teaches the system through every override.

TODAY
**Process owner**
Owns SOPs and throughput targets.

→

TOMORROW
**Workflow designer & outcome owner**
Decides which steps get which worker, sets autonomy levels and answers for the result.

TODAY
**Risk & internal audit**
Periodic sampling, after the fact.

→

TOMORROW
**Continuous assurance partner**
Reviews live evidence, tests guardrails and approves promotions up the autonomy ladder.

## Five commitments that earn workforce trust

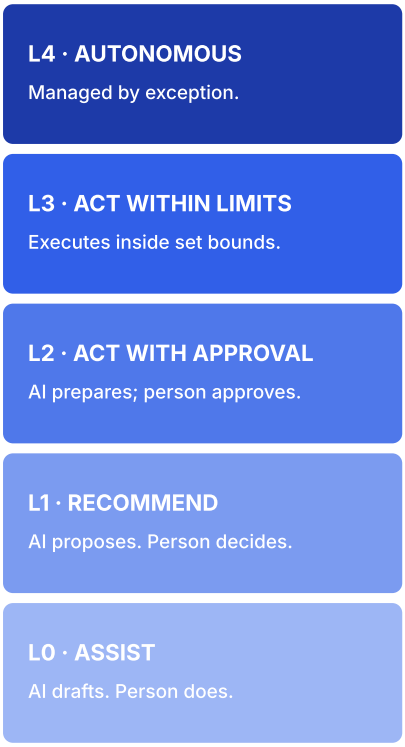
- **Show the reasoning.** People override what they can inspect.
- **Make the override one click.** And make it count — every correction improves the system.
- **Say what the work becomes** before the pilot starts, not after.
- **Invest in the transition.** Fund reskilling alongside licences.
- **Measure review effort honestly.** A faster task is not a better process if it creates more checking.

> The goal is not fewer people making decisions. It is better-prepared people making the decisions that matter.

09 · GOVERNANCE

# An autonomy ladder: let agents earn each rung.

Treat autonomy as a promotion with criteria. Every agent step has a declared level, an owner and evidence for why it holds that level. Promotion — or demotion — is a recorded decision.




| LEVEL | MINIMUM CONTROLS | EVIDENCE TO PROMOTE |
|---|---|---|
| **L4** | Continuous monitoring; drift alerts; periodic re-certification; documented fallback. | Reserved for low-consequence, reversible work. Rarely the goal. |
| **L3** | Hard limits (value, scope, count); sampled QA; kill switch; reversibility. | Stable error rate, cost of errors within appetite, clean audit sample. |
| **L2** | Approval step with case file; confidence thresholds; named approver. | Low override rate sustained over agreed volume. |
| **L0–L1** | Logged outputs; human makes every decision; no write access. | Accuracy against a labeled sample on real cases. |

## Anchor the programme to recognized frameworks

### NIST AI RMF

Govern, Map, Measure, Manage: a common vocabulary for AI risk that boards and auditors increasingly expect.

### ISO/IEC 42001

An AI management-system standard: roles, controls and continual improvement that can be certified.

### EU AI Act

Requires effective human oversight (Art. 14) for high-risk systems. Application dates were revised in 2026; confirm timing with counsel.

The level model is LTS's working framework, informed by common practice; it is not a regulatory classification. This paper is not legal advice.

# Fund on run-cost and error-cost, not build-cost.

Most automation cases are won on implementation estimates and lost in operation. Compare the two approaches across their full life, line by line, using your own data.

| COST LINE | RPA / DETERMINISTIC | AI AGENTS |
|---|---|---|
| **Build** | Process analysis and scripting; well-understood effort. | Use-case design, tool integration, evaluation sets; effort often underestimated. |
| **Run** | Licence and infrastructure; near-fixed per transaction. | Per-case inference and orchestration cost; varies with volume and complexity. |
| **Maintain** | Script repairs when interfaces or schemas change. | Prompt, model-version and guardrail upkeep; regression testing. |
| **Exceptions** | Cases fall out to people; often the largest queue. | Agent resolves many; the remainder arrive better prepared. |
| **Review & control** | Periodic testing; deterministic audit. | Approval workload, sampling, monitoring and evidence retention. |
| **Cost of error** | Rare but systematic when a script is wrong. | Distributed and harder to detect; depends on autonomy level. |

**A SIMPLE NET-VALUE TEST PER WORKFLOW**

Net value = (cases × cost avoided per case) − (run + maintenance + review + governance) − (error rate × cost per error)

## Four numbers to baseline before you buy anything

01

**Exception rate**

Share of cases that leave the happy path today.

02

**Cost per exception**

Fully loaded handling time and rework.

03

**Bot maintenance load**

Engineer hours per bot per quarter, from tickets.

04

**Cost of an error**

Financial, regulatory and customer impact.

If a proposal's return depends on a benchmark rather than on these four numbers, ask for the source and the sample. Treat vendor benchmarks, including ours, as hypotheses until your pilot confirms them.

11 · ROADMAP

# Four phases, each with a gate you can say no at.

Sequence matters: audit before pilot, pilot upstream of existing bots before replacing any, and migrate only what evidence supports. Each gate is a decision a steering group makes with data.

1 · AUDIT | 2 · HYBRID PILOT | 3 · CORE MIGRATION | 4 · SCALE

Weeks 0–6

### Process audit

Inventory bots. Rank by break frequency, exception rate and unstructured-input barriers. Decompose two workflows by step. Baseline the four numbers.

**Gate:** baselines agreed; two candidates chosen; owners named.

Weeks 6–18

### Hybrid pilot

Place an agent upstream to resolve exceptions before existing RPA runs. Start at L1–L2. Test failure, missing data and application change on purpose.

**Gate:** exceptions and cycle time improved without added review load or error cost.

Months 5–12

### Core migration

Replace the most maintenance-heavy UI bots with API-based integrations. Keep deterministic bots where no API exists. Promote steps on evidence.

**Gate:** run-cost and error-cost below the audit baseline.

Month 12 +

### Scale with governance

Stand up an agent governance function and shared orchestration platform. Extend to adjacent workflows; re-certify autonomy levels on a cycle.

**Gate:** continuous assurance operating; audit satisfied.

## The pilot scorecard: service, effort and control together

### Straight-through completion

Cases completed with no manual intervention.

### Exception & rework rate

Cases needing correction, review or recovery.

### End-to-end cycle time

Elapsed time to verified completion.

### Review & control effort

Human workload and control adherence.

**AGREE STOP CONDITIONS IN ADVANCE**

Define the thresholds — error rate, override rate, an unexplained incident, a control breach — at which an agent is demoted a level or paused. A pilot that cannot fail is a demonstration.

# One decision for each seat at the table.

The automation programme crosses technology, finance, operations and risk. These are the questions each leader should be able to answer within ninety days.

| LEADER | THE DECISION TO TAKE | THE QUESTION TO ASK YOUR TEAM |
|---|---|---|
| **CEO & Board** | Set the risk appetite for delegated decisions: which categories of action software may never take alone. | "Who is accountable when an agent is wrong, and how would we know?" |
| **CIO / CTO** | Mandate an orchestration layer with centralized identity, logging and kill-switches. Prioritize API integration over new screen scraping. | "Which of our bots break most, and what would it cost to replace them with an API?" |
| **CFO** | Approve funding on total run-cost and error-cost with stage gates, not on a build estimate. | "What is our baseline cost per exception, and where did this benchmark come from?" |
| **COO** | Choose two workflows to decompose and own the redesign, including the workforce transition. | "Which steps need judgment, and who will hold it once agents prepare the case?" |
| **CISO / CRO** | Define the autonomy ladder, least-privilege rules and evidence requirements before the first pilot goes live. | "Can we reconstruct and reverse any agent action within an hour?" |

THE STRATEGIC IMPERATIVE

Stop spending capital to repair fragile scripts, and stop buying autonomy you cannot govern. The advantage will go to the enterprises that decide, step by step, what to interpret, what to execute, and what a person must own.

## Sources

McKinsey & Company, *The state of AI in 2025: Agents, innovation, and transformation*, 5 November 2025. Survey fielded 25 June – 29 July 2025; 1,993 participants, 105 countries.

Gartner, "Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027," press release, 25 June 2025 (figures and quotation as reported by CRN Asia).

NIST, *AI Risk Management Framework 1.0*; ISO/IEC 42001:2023; Regulation (EU) 2024/1689 (EU AI Act), Art. 14. EU application dates were amended by the 2026 Digital Omnibus on AI; verify current dates with the Official Journal and legal counsel.

Process examples, the autonomy ladder and the placement map are illustrative LTS frameworks, not benchmarks.

MAKE THE NEXT DECISION SPECIFIC

# Bring a real process. Leave with a better-framed decision.

Start with the work, its exceptions and its operating constraints — not with a predetermined technology. In a working session with your process, technology and risk owners, we will:

1. Frame the process and the business objective.
2. Decompose it by step and place each on the map.
3. Surface exceptions, access needs and control points.
4. Define pilot criteria, owners and gates for a next step.

---


CONTACT

**RinTchen Kang**
rintchen@leotechnosoft.net
www.leotechnosoft.net


ABOUT LTS

LTS (Leo TechnoSoft) helps enterprises decide where RPA, AI agents and human oversight fit across business workflows — workflow first, designed around controls.

© 2026 LTS (Leo TechnoSoft). All rights reserved.