

— INSIGHT · ORG DESIGN FOR AI

# The Cost of **Autonomy.**

Why four autonomy levels produce four different business cases — and why pilots priced at Level 1 quietly become unbudgeted Level 3 systems in production.

---

AUTHOR

SERIES

EDITION

PUBLISHED READING TIME

HandsOn Insights Pillar 02 · Org Design for AI Vol. 01 · N° 05 April 2026 9 minutes

## EXECUTIVE SUMMARY

# Companies buy autonomy without pricing it.

Five findings senior leaders should weigh before signing the next AI business case.

---

**01****AI spending is doubling. The organizational investment isn't.**

BCG's 2026 AI Radar puts AI spending at **0.8 % of revenues, rising to 1.7 % in 2026**. In the same dataset, **70 % of the value** from AI comes from people and process change — only 10 % from algorithms. The two lines are not moving in lockstep.

---

**02****Most business cases price compute, not autonomy.**

A typical CFO sign-off captures licences, computational cost, and technical implementation. The operating cost that appears once a system runs autonomously — governance, monitoring, accountability, capability — is missing from the model and scales non-linearly with the autonomy level.

---

**03****Four autonomy levels produce four cost structures.**

L1 Human-in-the-Loop, L2 AI Decides & Human Reviews, L3 AI Decides & Human Notified, L4 Human-in-the-Exception. Each forces a different governance, monitoring, and capability load. Building infrastructure for L3 with the governance of L1 is the modal failure mode.

---

**04****Level 2 is the trap. It looks like oversight; it usually isn't.**

Deloitte's 2025 Human Capital data: **62 % of executives say AI already influences a majority of their decisions; only 5 %** report meaningful progress on decision-making governance. The gap sits almost entirely at Level 2 — the zone organizations assume is safe because someone is technically reviewing the output.

---

**05****Three decisions turn the Cost of Autonomy from surprise to line item.**

Classify every AI system by *target* autonomy level. Price the organizational load per level as explicit budget. Install classification governance — who can promote a system from one level to the next, against what evidence standard. None of the three require new technology.

---

## CONTENTS

# What's inside.

---

<b>01</b>	<b>The Pricing Error</b> Why the technical implementation cost line is not the whole truth.	<b>05</b>
<b>02</b>	<b>Four Levels, Four Cost Profiles</b> The Human–AI Interface and what each autonomy level forces on the organization.	<b>08</b>
<b>03</b>	<b>Where the Business Case Breaks</b> Level 2 looks like oversight. Most of the time, it isn't.	<b>11</b>
<b>04</b>	<b>Real Review, Not Decorative Review</b> Authority, exception triggers, and the design of an oversight layer that holds.	<b>14</b>
<b>05</b>	<b>Pricing Autonomy Before You Buy It</b> Three decisions for the Vorstand or executive committee this quarter.	<b>17</b>
<b>06</b>	<b>About HandsOn &amp; Sources</b> References, methodology, and how to engage.	<b>20</b>

---

## SECTION

# 01

## The pricing error.

The most expensive AI mistake in the enterprise right now is a pricing error in the business case. Companies buy autonomy without pricing it — and the organizational system that has to carry it is nowhere to be seen.

---

## The technological implementation cost line is not the whole truth.

**W**

hen a CFO signs off on an AI business case today, the P&L impact model usually has three lines: licences,

computational costs, and technical implementation. What is missing is the operating cost that appears once the system is live and takes on autonomy.

That cost is structural, not incidental. It scales *non-linearly* with the autonomy level at which the system runs.

A system that purely recommends is cheap to govern because a human makes every decision. A system that executes end-to-end inside a policy is expensive to govern because the organization has to build a monitoring, escalation, and recalibration architecture that is defensible under EU AI Act scrutiny and operationally embedded into everyday work.

None of that lives in the infrastructure budget. It lives in headcount, role design, training hours, governance forums, and a reporting line.

The failure mode is predictable. A Level 1 pilot delivers a proof point at a flattering cost. The business case for scale assumes the cost structure of the pilot. The system is then promoted to a higher autonomy level — because that is where the ROI actually sits — and the organizational load appears as incidents, stalled deployments, and the CFO question nobody can answer: *why did this pilot cost three times the business case?*

To answer that question in advance, you need a vocabulary for autonomy.

# The budget is doubling. The organizational investment is not moving with it.

BCG’s 2026 AI Radar, based on a survey of 2,400 executives, puts numbers on the pattern. AI spending is expected to rise from 0.8 % to 1.7 % of revenues in 2026 — roughly a doubling across the sample. In the same dataset, 70 % of the value from AI comes from people and process change; only 10 % comes from algorithms themselves. If the budget is doubling and seven-tenths of the value lives in the organization, the organizational investment has to move in lockstep. In the companies we work with, it rarely does.

## 0.8→1.7%

Of revenue. Expected AI spending in 2026 vs. 2025 — roughly a doubling.

BCG AI RADAR 2026 · 2,400 EXECS

## 70%

Of AI’s value comes from people and process change.

BCG AI RADAR 2026

## 10%

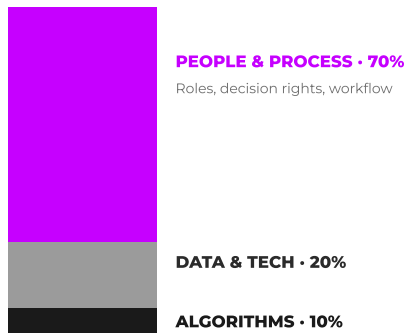
Comes from algorithms themselves. The remainder is data and tech.

BCG AI RADAR 2026

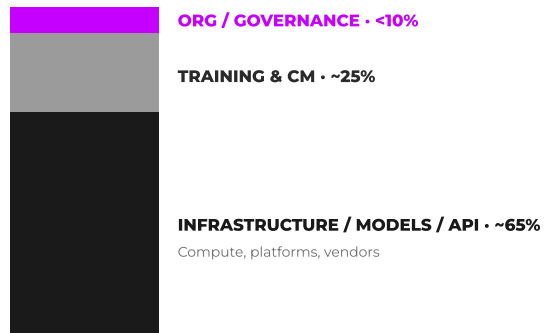
### Figure 01 · Where the value lives vs. where the budget goes.

Two stacks should match. They don’t.

SOURCE OF AI VALUE



TYPICAL BUDGET ALLOCATION



SOURCE · BCG AI RADAR 2026; HANDSON ANALYSIS OF EU MID-MARKET BUDGET COMPOSITION

---

## Article 14 prices itself in headcount, training, and authority.

The EU AI Act is explicit on what oversight requires. Human oversight must be assigned to a person with the necessary competence, training, and authority to intervene in or stop a high-risk AI system. That authority is not a paragraph in a policy document. It is headcount, role design, training hours, governance forums, and a reporting line — none of which appear on the infrastructure budget.

---

***“Human oversight shall be assigned to a natural person who has the necessary competence, training and authority to effectively exercise oversight of the high-risk AI system.”***

ARTICLE 14 · EU AI ACT · REGULATION (EU) 2024/1689

---

### DEFINITION · THE COST OF AUTONOMY

**The Cost of Autonomy is the delta between what a business case prices and what an autonomous system actually costs to run.**

It is the governance, monitoring, accountability, and capability load the organization has to absorb once the system stops recommending and starts deciding. It is structural — not incidental — and it scales non-linearly with the autonomy level. Pricing it in advance turns surprise overruns into a budget line. Skipping it produces incidents, stalled deployments, and a CFO question with no answer.

The architecture this whitepaper proposes — the **Human-AI Interface** — is the organizational answer. Every AI-enabled decision type runs at exactly one of four autonomy levels. The next section is the vocabulary.

## SECTION

# 02

## **Four levels. Four cost profiles.**

The HandsOn AI Operating Model defines the Human–AI Interface as the organizational architecture of decision-making and accountability when humans and AI share responsibility. Every AI-enabled decision type runs at exactly one of four autonomy levels.

# Each level forces a different cost structure on the organization.

The two foundation levels keep humans in or close to every decision. The two activation levels move humans up the stack — from per-decision review to system-level oversight. The error most organizations make is to build the technological infrastructure for Level 3 and the governance infrastructure for Level 1.

**LEVEL 01** FOUNDATION

## 01

---

### Human-in-the-Loop

AI recommends; a human approves every output. Safe by design. Governance load low, capability load moderate, monitoring load negligible. Unscalable in high-volume processes.

<b>GOVERN.</b> Low	<b>MONITOR</b> Low	<b>CAPAB.</b> Med
-----------------------	-----------------------	----------------------

**LEVEL 02 · THE TRAP**

## 02

---

### AI Decides, Human Reviews

AI executes inside a defined scope; a human reviews a sample and handles exceptions. Looks cheapest on paper. Carries the heaviest hidden cost in the model.

<b>GOVERN.</b> High	<b>MONITOR</b> High	<b>CAPAB.</b> High
------------------------	------------------------	-----------------------

**LEVEL 03**

## 03

---

### AI Decides, Human Notified

AI runs end-to-end inside policy limits. Humans monitor at system level. Named accountability — AI Owner, AI Steward, override authority — becomes load-bearing.

<b>GOVERN.</b> High	<b>MONITOR</b> High	<b>CAPAB.</b> Med
------------------------	------------------------	----------------------

**LEVEL 04**

## 04

---

### Human-in-the-Exception

AI orchestrates multi-step workflows. Humans set objectives; intervene only on exceptions. Heaviest governance load. EU AI Act documentation obligations scale here.

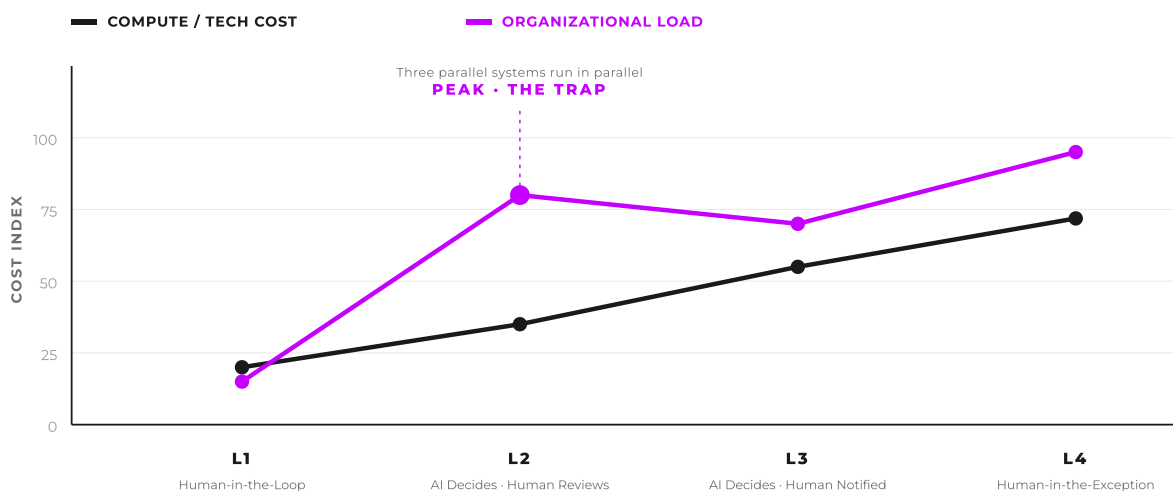
<b>GOVERN.</b> V-High	<b>MONITOR</b> V-High	<b>CAPAB.</b> High
--------------------------	--------------------------	-----------------------

# Compute cost is roughly linear. Organizational cost is not.

The technical cost line bends gently as a system gains autonomy. The organizational cost line bends sharply — and crosses the technical line somewhere between Level 1 and Level 2. Most business cases price the gentle line. The Cost of Autonomy is the gap between the two.

**Figure 02 · Cost of Autonomy by level.**

Indexed cost to run one AI-enabled decision type at each autonomy level. Compute (gray) vs. organizational load (purple).



SOURCE · HANDSON ENGAGEMENT DATA · INDEXED COST PER AI-ENABLED DECISION TYPE

Read this carefully: the organizational cost line peaks at **Level 2**, not Level 4. That is because Level 2 is the only level at which an organization pays for three parallel systems at once — the AI execution layer, the human review layer, and the interface that keeps them coupled.

## SECTION

# 03

## Where the business case breaks.

Levels 1, 3, and 4 force clarity on the organization. Level 2 is the trap — it looks like oversight because a human is formally in the loop, and the design work that would make that real often doesn't happen.

## Three things have to be written down and resourced. Skip any one and the review is a rubber stamp.

Levels 1, 3 and 4 force clarity on the organization. Level 1 has no autonomy to govern. Levels 3 and 4 are obviously autonomous — you cannot deploy them without sampling sizes, monitoring thresholds, kill switches, and an escalation path. The risk is visible, so the design work happens.

Level 2 is the trap. It looks like oversight because a human is formally in the loop. The oversight only works if three things are written down and resourced: **the sampling rate, the exception trigger, and the authority to override or retrain the model.**

Skip any one of them and the review becomes a rubber stamp. That is where most AI incidents in mid-sized organizations come from — AI that is nominally reviewed by humans who have neither the time nor the mandate to actually check and, if necessary, reject the output.

Article 14 of the EU AI Act does not treat decorative review as real either.

Competence, training, and authority are the standard the regulator will measure against, and a queue of flagged items with no one authorized to reject them does not meet it.

The cost consequence is specific. At Level 2, an organization pays for three parallel systems at the same time: the AI execution layer, the human review layer, and the interface that keeps them coupled.

If the review layer is built at the capacity and skill level required to actually reject outputs, Level 2 is often more expensive per decision than Level 3 — because at Level 3 the humans have moved to system-level oversight and the per-decision labour has been engineered out.

---

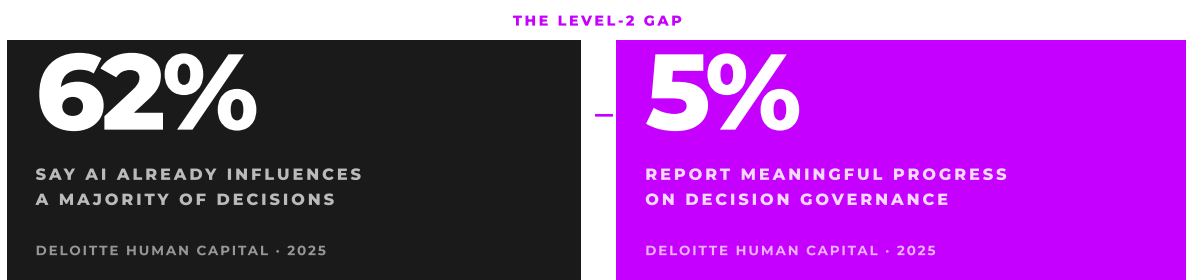
## 62 % are deciding with AI. 5 % have governed it.

Deloitte’s 2025 Human Capital data quantifies the gap. The delta between “AI shapes our decisions” and “we have decision-making governance” sits almost entirely at Level 2 — the zone organizations assume is safe because someone is technically reviewing the output.

---

### Figure 03 · The Level-2 gap.

Executives reporting AI’s influence on decisions vs. executives reporting decision-making governance progress.



A 57-point gap between AI-shaped decisions and AI decision governance.

SOURCE · DELOITTE 2025 GLOBAL HUMAN CAPITAL TRENDS · EXECUTIVE COHORT

---

***“The autonomy level that feels cautious at first is frequently the most expensive one to run.”***

SECTION 03 · THE COST OF AUTONOMY

---

## SECTION

# 04

## **Real review. Not decorative review.**

Two variables determine whether Level 2 oversight is real: **the authority of the reviewer** and **the design of the exception trigger**. Both are organizational and governance variables — not technical ones.

## Two variables. Both organizational. Both consistently underbuilt.

**Authority is about decision rights.** If the reviewer cannot overrule the model, request retraining, or suspend the system without a three-layer escalation, the review workflow is functioning as throughput management, not oversight.

The HandsOn Decision Rights Registry treats this as a design artifact: for every AI-enabled decision type, there is a named authority, an evidence standard, and a recalibration trigger. A company that cannot produce this registry for its top five AI use cases does not have Level 2 oversight; it has a queue.

**The exception trigger is the second variable.** A review gate that flags 1 % of cases and is resourced to review 100 % is expensive but functional. A review gate that flags 20 % of cases and is resourced to review 5 % is a compliance exposure — because 15 % of outputs are passing through unreviewed while the organization tells itself a human is in the loop.

The trigger rate, the resourcing, and the statistical sampling design all have to be published, monitored, and adjusted as model performance drifts. That work is invisible in most business cases. It is also the single largest driver of Level-2 run-rate cost in our engagements.

McKinsey's State of AI data indicates that roughly **80 %** of AI-using organizations have not redesigned a single workflow around their AI deployments. That is a stark way of saying the same thing: the review architecture exists in slideware, not in the operating model.

As long as that is true, the Cost of Autonomy will show up as incidents, rework, and regulatory exposure rather than as a line in the business case.

# A 20 % trigger resourced for 5 % review is a compliance exposure, not oversight.

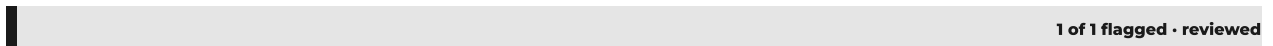
The arithmetic of an exception trigger is unforgiving. Two parameters — the rate at which the system flags cases for human review, and the capacity the organization has resourced to actually review them — determine whether the oversight layer holds or leaks.

## Figure 04 · Three review architectures.

100 outputs per scenario. Flagged for review (purple) vs. actually reviewed (dark) vs. passed unreviewed (gray).

### SCENARIO A · FUNCTIONAL

Trigger 1% · Capacity 100% — expensive per case, but defensible.



### SCENARIO B · COMPLIANCE EXPOSURE

Trigger 20% · Capacity 5% — 15% of outputs pass through unreviewed.



### SCENARIO C · CALIBRATED

Trigger 5% · Capacity 5% — trigger rate matches resourced review capacity.



■ REVIEWED    ■ FLAGGED · NOT REVIEWED    ■ UNFLAGGED · PASSED

SOURCE · HANDSON ENGAGEMENT DATA · ILLUSTRATIVE MODEL OF TRIGGER-RATE / REVIEW-CAPACITY MISMATCH

## SECTION

# 05

## **Price autonomy before you buy it.**

Three decisions turn the Cost of Autonomy from a post-hoc surprise into a priced line item. None require a new tool stack. All three are within reach of any Vorstand or executive committee this quarter.

---

## Three decisions for the next thirty days.

The practical question is what a COO or CFO should actually do. Three decisions, in this order, turn the Cost of Autonomy from an after-the-fact surprise into a priced line item.

---

### DECISION 01

# 01

#### **Classify every AI system by target autonomy level — not the level it runs at today.**

A forecasting system with a human approving every output is Level 1; if the ROI depends on approving 10,000 outputs a week, the real target is Level 2 or 3. This exercise alone reliably surfaces 5–15 systems per mid-sized enterprise that are planned at one level and budgeted at another. Until classification is by target, the business case is fiction.

---

### DECISION 02

# 02

#### **Price the organizational load per level as explicit budget lines.**

For each target level, estimate governance, monitoring, capability, and accountability cost as separate items — review headcount at Level 2, dashboard and drift tooling at Level 3, exception engineering at Level 4, training programs calibrated to the target level rather than generic AI literacy. The exercise forces the conversation out of the compute budget and into the operating budget, which is where the real cost lives.

---

### DECISION 03

# 03

#### **Install classification governance — who can promote a system from one level to the next.**

The most frequently skipped step, and where most regulatory exposure gets built. A Level-2 system that quietly becomes Level 3 without a governance review is a non-compliance event waiting to be discovered. One page is usually enough — but it has to name an authority, an evidence threshold, and a cadence.

---

All three decisions fail politely if the organization has not already decided who is accountable for AI outcomes at the board level — which is, separately, the precondition for any of this to work.

---

## A line item, not a surprise.

The Cost of Autonomy decides whether your AI portfolio earns its business case or embarrasses it. Four autonomy levels; four organizational cost structures. A pilot priced at Level 1 and scaled to Level 2 or 3 without rebuilding the cost model is the single most common cause of AI business cases that look strong on paper and fail in production.

If you run an AI portfolio: take the five largest initiatives, classify each by target autonomy level, and ask your CFO to price the organizational load per level. **If the delta between compute budget and full cost is less than 2x, the exercise is incomplete.**

If you are at Vorstand or board level: put the three decisions on the agenda — classification, organizational load pricing, classification governance. None of them require new technology. All of them are cheaper to take now than after the first incident.

Article 14 will be enforced in courtrooms, not in slideware. The organizations that price autonomy properly will discover their AI investments earned what was promised. The organizations that don't will discover the Cost of Autonomy the same way every previous generation of unpriced organizational risk has been discovered: as a line item on next year's remediation budget.

---

***“If the delta between your compute budget and your full cost is less than 2x, the exercise is incomplete.”***

CLOSING ARGUMENT · THE COST OF AUTONOMY

---

### ENGAGEMENT

**If you're classifying AI systems by autonomy level, or pricing the organizational load on a portfolio, reach out directly.**

HandsOn works with leadership teams on exactly this transition — classifying AI decision types by target autonomy level, pricing the operating-model load per level, and installing the classification governance that keeps the business case honest as systems scale.

## ABOUT THE PUBLISHER

## HANDSON

# HandsOn Insights.

Pillar 02 · Org Design for AI

HandsOn is an independent advisory firm working with European mid-market and enterprise leadership teams on AI organizational design. The Insight Series publishes original research and field-derived frameworks on the strategy, structure, and governance of AI inside large organizations.

The HandsOn AI Operating Model defines the Human–AI Interface as the core design object of an AI-enabled organization. This whitepaper formalizes the cost model that follows from it — a pricing vocabulary for the organizational load that every level of autonomy demands.

## SOURCES

---

– **BCG** · AI Radar 2026 · survey of 2,400 executives.

---

– **Deloitte** · 2025 Global Human Capital Trends.

---

– **McKinsey** · The State of AI 2025 (QuantumBlack).

---

– **EU AI Act** · Regulation (EU) 2024/1689 · Article 14, Human Oversight.

---

– **HandsOn AI Operating Model** · Human–AI Interface framework.

---

– **HandsOn** · Decision Rights Registry · engagement data 2024–2026.

---

## CONTACT

info@wearehandson.de  
wearehandson.de

## SERIES

HandsOn Insight  
Pillar 02 · Org Design for AI

## EDITION

Vol. 01 / N° 05  
April 2026