

Forty-eight reports a day instead of twenty.
One evening hour to approve them.

What AI Tests in Your Rules

Ten questions for introducing AI,
seen from the organization

Guido Bosbach

Organization under AI

October 2026

Executive Summary

AI makes producing cheap. Checking and taking responsibility have not become cheap to the same degree. When the volume of drafts, reports, and proposals rises, the work does not disappear. It moves to whoever signs last, without anyone having decided it should.

Many AI policies answer one question well: who is responsible? They rarely ask the second: can that person actually carry the load, with the time, access, and authority available at that point? This paper treats AI adoption as a test of your organization's rules. It walks through ten questions, one for each part of an organization's operating rules, and gives you a measure for each.

Key takeaways

- **A set of rules holds only if it promises no more checking than people can deliver.** Every rule that requires a check is a promise someone has to keep with their time.
- **Reversible decisions may run at machine pace; binding ones may not.**
- **An assigned responsibility is not yet a carried one.** Whoever is responsible needs authority, access, and time.
- **Decide where freed-up time goes before the first saving is booked.** Otherwise it disappears, and with it the room to adapt.
- **Protect the learning years.** Whoever no longer does anything simple today will not be able to check anything difficult tomorrow.
- **Set the way to change your rules before you cut any of them.**

1 • Plus 140 Percent

Most AI rollouts begin with good news. At Thrushwood Home Care, a home care agency, the news was this: the nurses barely had to write anything on their rounds anymore. Two sentences into the phone, and the system turned them into a complete care report. Because dictating now takes almost no effort, every visit gets its own report; before, the nurses wrote mainly when something stood out.

Three months later there are forty-eight reports a day instead of twenty. That is 140 percent more. The director of nursing has to sign off on every one of them. The agency wrote that into its documentation policy after a finding in its last quality inspection. The approvals now reach her at 7:30 in the evening, after the rounds are billed and the phones have gone quiet.

If she has an hour for them, each report gets seventy-five seconds. That may be enough to read it. That the wound on a patient's lower leg looked different today than yesterday, the nurse had mentioned in half a sentence. The system turned it into a tidy paragraph in the middle of the page. Only

someone who put it next to yesterday's entry or called the nurse would see whether it is right, and seventy-five seconds leave hardly any time for either.

Thrushwood, one evening



Constructed case; the numbers are not evidence. Assumption: one hour of approval time in the evening.

Everyone involved follows the intended process. The nurses use a tool that was introduced to relieve them, and the director does what the policy requires. Management's numbers work out too, because the rounds go faster. The nurses really do have less to write. Even so, part of the work has moved: to the desk of the person who signs last, and into an hour of the day that appears on no schedule.

The load comes from a chain: the system makes writing easy, so more gets documented, and every additional report goes through the same approval as before. Nobody thought about that combination when the system was introduced.

Thrushwood does not exist. The agency is constructed, and its numbers are not evidence. What happens there, though, happens in real organizations every day, usually spread across more people and therefore harder to see.

This paper asks what an organization needs to settle so that a rollout like this keeps its promise. The answer lies in the organization's rules: the rules by which it decides, and the way those rules get changed. The technology itself is not the focus. It changes faster than any paper can be revised. The framework in which people and technology work together, however, urgently needs an update.

2 · The Bill That Comes With It

Producing has become cheap with AI. Checking and taking responsibility have not become cheap to the same degree. The economics behind this are well known: in 2018, Ajay Agrawal, Joshua Gans, and Avi Goldfarb described how a component that gets cheaper raises the value of whatever is

needed alongside it and does not get cheaper itself.

What is new is the volume. Thrushwood has forty-eight reports where it used to have twenty. A question of price becomes a question of capacity. A person reads English nonfiction at about 238 words per minute, as the psychologist Marc Brysbaert established across many studies in 2019. Checking takes longer than reading, because it means comparing what you read with what you already know. Good tools can make that comparison easier, for example by flagging what changed since yesterday. They do not take the judgment off anyone's hands.

When volume grows and the checking process stays as it was, part of what is saved in producing shows up again in checking, just somewhere else and usually in a different budget line. In *Cheap to Make, Costly to Check* I call this the checking load.

It lands where the signature is due, with people who have a name and a role. There it looks like a personal problem, like overload or slipping diligence, and whoever reports it seems to be reporting their own inadequacy.

In September 2025, BetterUp Labs and the Stanford Social Media Lab surveyed 1,150 office workers in the U.S.; BetterUp sells coaching to companies and so has an interest of its own in the topic. Forty percent had received work in the previous month that they believed was machine-generated and that looked like work without saving any. By their own account, they needed almost two hours per case to make it usable.

Every rule is a promise

A second bill runs alongside the first. Every rule that says something must be checked promises that someone will look before something takes effect, and someone has to keep that promise with their time, their skill, and in the end their reputation. An organization's rules are therefore also a sum of such checking promises. As long as volume stayed flat, hardly anyone noticed how many of them were no longer being kept.

Where AI multiplies the volume, every one of these promises becomes more expensive at once. That leads to the sentence I consider the most important: **A set of rules holds only if it promises no more checking than people can deliver.** If the checking workload grows and time and staff do not, the rules are hollowed out from within. For a long time nobody notices, because on paper the checking continues.

When AI multiplies what arrives at a checkpoint and the people there do not get the means to do their job, checking becomes faster, shorter, and more superficial. The risk of errors and bad decisions rises with it.

Two questions at the same point

For every AI rollout, and increasingly for software updates that look harmless, two questions follow. **Who owns the new checking load?** That is the question of ownership. **And can they carry it where it lands?** Are the resources, the time, the knowledge, and the authority to decide actually there? That is the question of capacity.

Many software and AI policies answer the first question thoroughly and the second only in principle. I checked two key frameworks, the NIST AI Risk Management Framework and Article 14 of the EU AI Act, against the full text in *Cheap to Make, Costly to Check*. Both assign responsibilities, and the NIST framework expressly calls for the necessary resources to be taken into account. Neither names a number for incoming volume or review time. But an assigned responsibility is not yet a carried one. Thrushwood's director of nursing is responsible; that much is settled. Whether she can do in seventy-five seconds what the policy expects of her, nobody asked when the system was introduced.

The reflex and its counterpart

When something goes wrong, the answer is very often the same. Instead of looking at how the process actually runs, the organization adds another check: one more signature, because machine-generated drafts call for more caution. I call this the control reflex. It is understandable, and it makes worse the very problem it is meant to solve, because every new checkpoint is one more promise drawn on the same scarce hours.

Behind it sits a view of people that is rarely spoken aloud: people make mistakes, so we need rules to catch them. That is not wrong. It becomes wrong only when it is the only thing a set of rules knows about people. Then the rules waste the judgment of the people on the spot, and that judgment is exactly what an organization needs for everything no system anticipated.

The counterpart is a deliberately designed trust architecture that starts at the rollout, long before anything goes wrong. It means rules that state explicitly whom they trust with what, people and systems alike. Trust then becomes a decision with four parts: who trusts, whom, with what, and who withdraws it if it does not hold up.

In *Cheap to Make, Costly to Check* I put this into one sentence: *AI is well placed where it makes preparation cheaper and leaves judgment with the person who answers for it.* Where that person sits and what they need is decided by the rules the system is introduced into.

3 · What AI Tests in Your Rules

In most organizations, the decision to use AI was made long ago, often without anyone deciding it. Features that pre-sort, rate, or draft arrive through updates in the software people already use. If you only review your list of AI projects, you see a fraction of them.

The open question is whether the rules hold that these features flow into. Every organization has rules, even if they are written down nowhere in full, and one question sits above them all: **By what rules does an organization decide, and who may change those rules?**

I sort the answer into nine topics. Beneath them lies a foundation: the question of whom these rules trust. Together, these are the ten questions of this paper. In a workbook now in preparation, these become the articles of an organizational constitution. This is not a legal document like bylaws

or a works agreement. It is a set of rules an organization gives itself. I developed its basic form with colleagues: a few hard limits, an order of precedence, stated reasons, and a procedure for change.

For an AI rollout, the articles can serve as a diagnostic grid, because AI raises a distinct question at each of them. I group them under three questions I put at the start of every organizational analysis.

Does it arrive? Does what was intended reach the place where the work is done? **Does it hold there?** Does that place have what it needs? **What is still open?** Where is it not yet settled who decides in a dispute and when rules change?

Diagnostic question	Article	The question AI raises there
Does it arrive?	Foundation	Whom do your rules trust, people or systems?
	1 · Purpose	What are you using AI for?
Does it hold?	2 · Membership	Who can do what AI requires, today and ten years from now?
	3 · Limits and Precedence	What still applies when the machine would be faster, and how do you still recognize performance?
	4 · Deciding	What may run at machine pace?
	5 · Voice	Who may stop the line?
	6 · Knowing and Checking	Can the person who should check actually check?
What is still open?	7 · Allocating	Where does the time AI frees up go?
	8 · Conflict	Who decides when two good reasons collide?
	9 · Amending	Is a software update a change to your rules?

Ten questions AI asks of your rules

What is still open?

8 • Conflict

Who decides when two good reasons collide?

9 • Amending

Is a software update a change to your rules?

+ Your own question, as an eleventh

Does it hold?

2 • Membership

Who can do what AI requires, today and ten years from now?

3 • Limits and Precedence

What still applies when the machine would be faster?

5 • Voice

Who may stop the line?

4 • Deciding

What may run at machine pace?

6 • Knowing and Checking

Can the person who should check actually check?

7 • Allocating

Where does the time AI frees up go?

Articles 4 and 6 are where the bill from Cheap to Make, Costly to Check sits: whatever is produced cheaply, someone has to be able to check at that pace.

Does it arrive?

1 • Purpose

What are you using AI for?

Foundation • Whom do your rules trust, people or systems?

Based on the framework of a workbook on organizational constitutions (in preparation) and Cheap to Make, Costly to Check (2026).

4 • Does It Arrive? Foundation and Purpose

Foundation • Whom do your rules trust?

Every set of rules carries an image of the people it was written for. It is rarely written down. You can still read it: in the number of signatures on a purchase order, and in who is excused from providing proof and who never is.

AI adds a second image, the one of the system. In experiments, lay people follow advice more readily when they believe it comes from an algorithm, as Jennifer Logg and colleagues showed in 2019. Once people see the system make a mistake, though, trust drops faster than it would for a person making the same mistake, as Berkeley Dietvorst and colleagues showed in 2015. If you demand every receipt from people and believe the system unchecked, you are allocating trust by habit. Neither has much to do with who is actually more reliable.

So before the rollout, write down what you trust the system to do, and give that assumption a date on which it will be reviewed. The system will change in between, with every update.

Look also at the sentences that steer a rollout before it begins: “AI can’t do that yet.” “AI is forcing us to cut jobs.” “The AI did that, not me.” Each of them presents itself as a fact. They are beliefs that you can test before they shape everything else.

Article 1 · Purpose: What are you using AI for?

Two questions get mixed up in almost every rollout: what AI can do technically, and what the organization wants to use it for. A vendor can describe the first, and the organization is well advised to check whether that holds for its own use. Only the organization can answer the second, and for that it needs a purpose against which a decision can be aligned.

A usable purpose can be a single sentence, as long as it creates clarity. The test is simple: what did your last difficult decision actually align with? If the honest answer is “the quarterly numbers,” that is your real purpose, and every AI application will be measured against it, whether anyone decided so or not.

So keep the two questions apart. First establish what already exists, including the use that has long been running unobserved in daily work. Only then turn to what you want, and measure every planned use against your purpose.

Purpose also helps you look outward. At least once a year, bring what is happening outside deliberately inside: review what is foreseeable, ideally together with someone from outside the organization. Foreseeable means something narrow here: decided, ordered, or measured. A signed system contract counts, and so does a feature the vendor has announced for the next release.

5 · Does It Hold? Articles 2 to 7

Article 2 · Membership: Who can do what AI requires?

Membership in this article means more than a box on the org chart: whoever belongs to a task also needs the means to carry it out. When an AI policy assigns responsibility, it usually names a role or a person. That alone achieves little. The person can do the job only if they have three things: the authority to decide, access to the information and systems they need, and the time to actually do it.

So establish that a task counts as accepted only once the person has all three. If one is missing, they may decline the task with reasons, and management has to close the gap. This is a rule you give yourselves, not a legal entitlement. This prevents anyone from being responsible without the means to live up to that responsibility.

Before a task passes to a system, take a knowledge inventory. What knowledge does this task contain that appears in no handbook and shows up only when something goes wrong? Who holds it, and when do they have time to pass it on? Without these questions, an apparently pure efficiency

task easily becomes a transfer of experience that nobody notices.

If an AI-supported process depends on one person's ability, for example the ability to judge its results, that ability should exist in two places, with a trained backup. This failover capacity does not require staffing any role twice.

The learning years

The question of whether someone can check reaches further into the future than it first appears. **Whoever no longer does anything simple today will not be able to check anything difficult tomorrow.**

Where AI takes over tasks, entry-level positions are hit first: the roles whose visible tasks a system seems to handle just as well, such as research, first drafts, and data preparation. An analysis of U.S. payroll data by the Stanford Digital Economy Lab shows a widening gap for 22- to 25-year-olds in the occupations most exposed to AI. In the August 2026 revision, their employment stands 19 percent below where it would be had it kept pace with less-exposed peers. Experienced workers show no such gap, and the authors find no widespread displacement. They expressly do not claim to have identified a cause. Yet these are the roles that would produce, ten years from now, the judgment needed to evaluate AI output at all.

Much of this learning was never a separate cost item. It was paid for by the production beginners did along the way, and whoever did that work learned from it. When that production costs nothing anymore, the currency that paid for training disappears. What remains is an expense this quarter whose return lies in another decade.

Ford shows what it costs when experienced judgment is missing, even though the cuts there were neither about entry-level roles nor about AI. For cost and restructuring reasons, the company cut administrative and engineering positions in three rounds: about 7,000 salaried positions worldwide in 2019, roughly one in ten; about 3,000 positions in 2022, mostly in the U.S., Canada, and India; and 3,800 more in Europe in 2023, mostly in product development, about 2,300 of them in Germany.

In June 2026, Charles Poon, Ford's vice president for vehicle hardware engineering, acknowledged that the company had believed it was enough to introduce AI and feed it its design requirements. By then Ford had hired 350 experienced engineers, many of them former employees, others from suppliers. According to TechCrunch, they are to train younger staff, hunt for failure points before a part reaches the plant floor, and reprogram the AI tools.

My reading, not Ford's: with the eliminated positions, a great deal of experienced judgment left before it reached the systems that were supposed to build on it. Ford itself has not drawn a connection between the cuts and its quality problems. Ford could still bring people back. Where entry-level roles disappear today, the pool of people who could be brought back in ten years shrinks.

Cutting an entry-level position costs a salary. Buying it back in ten years costs the price of a market in which the fundamentals have become scarce.

The mechanism is older than any language model. In 1983, Lisanne Bainbridge described in a paper on the ironies of automation how automation takes the practiced parts of a job away from people and leaves them the hard ones. Whoever steps in only when the system is stuck is expected to master exactly the cases they have had no practice with.

If you have never produced a flawed draft yourself, it is hard to develop a feel for what a flawed draft looks like. Guided error analysis can replace part of that practice. The work psychologist Winfried Hacker, in a 2026 paper published by Germany's Federal Institute for Occupational Safety and Health, calls the consequence burdensome unlearning and names it, next to the burdensome responsibility for checking and approving AI results, as a second source of psychological strain. The link has not yet been measured rigorously.

For apprentices, interns, and new hires, the rule is: first be able to do it yourself, then add the tool. The University of Chicago Law School is piloting a similar approach in its first-year legal writing course in the 2026–27 academic year. Writing without AI is the foundation there, and writing with AI is layered onto it in the same year. Checking is a skill that has to be practiced. Authorization to sign does not replace that practice.

The old entry-level work does not need to be preserved for its own sake; some of it was monotonous and taught little. But someone has to make its learning function visible and rebuild it. Sometimes that means having people do a task deliberately even though the machine would be faster, and showing the cost openly. It is paid from the reserve described in Article 7.

No single organization can solve this alone, because the benefit often accrues to another employer. Germany and its neighbors have an institution that carries exactly this burden jointly: the dual apprenticeship system. Whether it can do so a second time remains open.

Article 3 · Limits and Precedence: What still applies when the machine would be faster?

Frank Ostrowski, also invented, runs a trade business with forty people. One Monday in February his estimator asks whether the system may write their quotes from now on. Ostrowski says yes, and that evening he realizes he had no yardstick for that answer beyond a feeling that it would probably be fine.

Two weeks later, the first sentence in his document that applies even when it gets expensive reads: *"No quote goes out unless its calculation has been checked and one of us stands behind it."* You recognize a hard limit by the fact that, in the case it was written for, it costs something: a contract, say, or a month's head start.

For AI there is almost always an obvious limit: decisions about hiring and dismissal, pay, and development are made by a person. A named individual stands behind the reasons, toward the person affected and not only toward their own manager. Three such limits are plenty to start with, because each one has to be defended when it counts.

The second part of the article is precedence. Below the hard limits, an order of precedence settles what wins in case of doubt, such as safety over speed, or a promise to customers over an internal metric. For an AI rollout, this means: when fast processing competes with thorough checking, the order of precedence says in advance which goal wins. Whoever departs from it gives written reasons.

Two mistakes are tempting. The first is turning a method into a limit. Ostrowski checks every price himself today. That is his way of keeping the limit, and it may change as soon as there is a better way; the limit stays untouched. The second mistake is copying the law. A limit that merely repeats

what the EU AI Act says has to be revised with every amendment, and your own decisions get lost among the copied obligations.

How you still recognize performance

The limit says who decides about evaluation, promotion, and pay. It does not say how that person can still recognize performance. AI has opened that question.

A result looks the same whether someone worked it out or merely accepted it. With the right tool, a new hire can produce a market analysis in an hour that looks like what an experienced manager needed days for in 2024. Whether it carries the same judgment, you cannot tell by looking.

Organizations long read effort as a proxy for diligence, and the finished result as a proxy for skill. If you keep measuring performance by the visible result, you no longer see who checked, improved, or rejected a draft. Under AI, that is often the work that matters most.

I do not have a ready answer. One pointer does help: judgment shows in a conversation about the path to the result, about what someone discarded, where the system was wrong, and why the number is right. An evaluation that makes room for this conversation sees again what it is supposed to evaluate.

One temptation is close at hand, and it is best ruled out in advance: using the systems' logs to measure performance by counting who accepts, edits, or rejects how much. In many jurisdictions this requires consultation with employee representatives, and in every organization it damages the open feedback that Article 5 is about. Anyone who knows that every rejection counts in their evaluation will reject differently. The rejection rate from Article 6 therefore belongs to the checkpoint and to no individual. If a checkpoint is staffed by one person, the rate can still be traced to them. Then aggregate it across several checkpoints, or rule out its use in evaluations in a written agreement.

Article 4 · Deciding: What may run at machine pace?

The rule that is easiest to remember is this: **Reversible decisions may run at machine pace; binding ones may not.** A decision is reversible only under two conditions: an error in it shows early, and the damage can still be repaired by then, even if the same error sits in a thousand cases. If either condition is missing, the decision is binding and is made slowly, in writing, and by more than one head.

How much machine a decision can tolerate is sorted by four zones. In **Zone A**, the system may act alone where an error shows and costs little, as with a rough translation used only for your own orientation. A summary belongs here only if its reader knows the original; otherwise they will not notice its errors. In **Zone B**, the system proposes and a person checks before the proposal takes effect: the decision already exists, and the person confirms or rejects it. In **Zone C**, the person weighs the essential reasons and alternatives and gives reasons for the decision; the system only supplies groundwork. This zone covers decisions with significant consequences for people, such as hiring, pay, or development, and any other decision that hits those affected hard and whose appropriateness cannot be fixed in advance. **Zone D** cuts across the other three. It describes a kind of decision more than a division of labor, the moral borderline case, and can therefore also apply to a decision

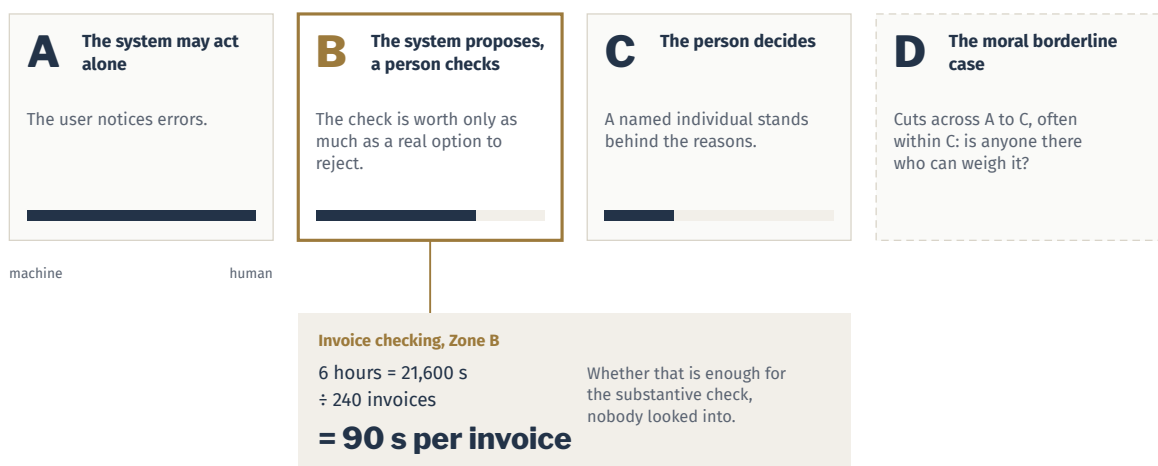
in Zone C. Two legitimate concerns collide, the system supplies material at most, and human judgment is mandatory. The question there is whether anyone is present who can make that judgment and has practiced it.

The idea is not new. In 1978, Thomas Sheridan and William Verplank described ten levels at which humans and computers share a task in remotely operated undersea vehicles, from the level at which the human does everything to the one at which the computer acts alone and ignores the human.

A zone helps only if the time available for each case fits it. Otherwise it is just a claim. An accounts payable clerk, also invented, checks incoming invoices; the system reads them and presents them for approval. Zone B. Before the rollout, sixty invoices arrived per day. Since the supplier portal was connected, the invoices of all three sites land with her, 240 a day.

Six hours of focused work are 21,600 seconds. Divided by 240, that leaves ninety seconds per invoice. An automated match can confirm amounts and invoice data. Whether the service was delivered as billed only shows in the substantive check: against the purchase order and delivery note for goods, against proof of performance for services. Whether ninety seconds are enough depends on the checking mandate. Nobody looked into it when the system was introduced.

How much machine a decision can tolerate



Reversible decisions may run at machine pace. Binding ones are made slowly, in writing, by more than one head.

Zones and example after Cheap to Make, Costly to Check (2026). Reversible means: an error is noticed early, and the damage can still be repaired.

A quieter shift adds to this. A system may formally only prepare a decision and still shape it, when its preselection determines what gets attention in the first place. The test question is: how often do people in practice still depart from the system's preselection?

And where a new checkpoint is created in response to all this, it needs a reason, a name, and an expiration date. After twelve months, someone decides whether it stays. Otherwise every precaution becomes permanent.

Article 5 · Voice: Who may stop the line?

Whoever works in an AI-supported process often sees first that something is off: a draft that sounds plausible and is wrong, or a rating that keeps hitting the same people. Whether they say so depends on whether they know in advance what will happen to their objection.

So anyone with well-founded doubts about a single case may stop it without asking first, and learns the decision and its reasons. At first only that one case is paused; the rest of the process continues. If the doubt points to an error that affects many cases, it goes straight to the system's owner. The owner or their deputy decides whether to stop the whole process; that is the crisis stage from Article 9. If serious harm is possible, the process is stopped right away as a precaution; otherwise a deadline that matches the risk applies. Whoever stopped a case may suffer no disadvantage from it, even if the doubt turns out to be unfounded.

An objection to a rule is recorded with the name of the person raising it, and the rule continues to apply until a decision is made.

Use without permission, often called shadow use, is a special signal. Two cases need to be told apart. Whoever puts confidential data without authorization into a tool whose terms do not protect it puts data protection or trade secrets at risk, and the organization has to respond. Whoever uses a different tool because the approved one falls short mainly shows where the rule misses everyday reality. That carries risks too, and the organization needs to know them. Punishing it, however, destroys the very information that could correct the rule. If shadow use grows despite an existing rule, the rule has apparently stopped serving its purpose.

Where employees have formal representation, involve them early enough that their concerns can still change the plan. In Germany, for example, the law requires employers to inform the works council in time about AI being planned into work processes.

And where nothing about a decision can be changed anymore, say so up front and listen anyway. A consultation that leads nowhere and pretends otherwise does more harm than none.

Article 6 · Knowing and Checking: Can the person who should check actually check?

Part of your rules sits inside systems without anyone having decided it. A purchase order without a cost center cannot be submitted; an applicant portal ranks by a criterion that arrived with an update. Such rules are harder than any written procedure, because they cannot be bypassed.

Each of them needs an owner, a named person who stands behind the rule. And each needs a description of what the rule does not detect and what it is not meant for.

This matters most for updates. In *Cheap to Make, Costly to Check*, an IT colleague installs an update to the procurement software on a Tuesday evening, runs two test cases, and drives home. He did nothing wrong. What remains open is why nobody had the job of asking, that day, what the new feature decides.

For the check itself, one principle applies. Whoever is responsible for a result cannot delegate an independent check to the same instance that prepared the result, and a second system does not check independently merely because it is a second one. If both take a convincing-sounding answer

as a sign of correctness, they only confirm each other.

The simplest metric is one almost every organization has and few look at: the share of proposals that are rejected or changed. The idea is not new. In an appendix to its AI Risk Management Framework, the U.S. National Institute of Standards and Technology notes that data on how often and why humans overrule AI output may be useful to collect, though it treats this as an open research question.

If a checkpoint never rejects anything over a long period, that is a reason to look. A rejection rate of zero allows several readings. The system may be exceptionally good, the number of cases small, or changes may not be recorded as rejections at all. It may also be that no one is checking anymore. The number cannot decide between these readings, but it does make sure someone asks.

And finally, the pilot. Many organizations are trying out AI tools right now, and many trials end in a slide deck, because nobody decided in advance who decides afterward and which rule would have to go. A trial clause answers both up front: a time-limited exception with a named owner, a scope, a success criterion, and a stop criterion.

A pilot also needs an answer to what happens if the vendor changes or fails. A dependency with no way back is a binding decision, even if procurement filed it as a license renewal.

Article 7 · Allocating: Where does the time AI frees up go?

When AI frees up time, that time goes where freed-up time always goes: into the next urgent task, or into a position that is not refilled. Either way, nothing is left for whatever the same system changes next. So decide where the time goes before the first saving shows up in the budget.

Protect part of it for learning, for experiments, and for looking outward. A quarter is a good start; that is a rule of thumb from my work, not a measured value. That share needs a way back too: if the reserve goes unused for two quarters, decide again.

If someone says AI is forcing job cuts, run a simple test: what has actually changed in operations? If the answer is a software license and a pilot project, it is a cost decision in an AI costume, and such decisions usually hit the entry level from Article 2 first.

Metrics belong in this article too, because they are rules nobody calls rules. Thrushwood measures the shorter rounds. Nobody measures the director's evening hour. Whoever governs allocation has to notice such shifts before someone burns out, and every metric needs a date on which someone checks whether it still measures what matters.

Then there is the load limit. Every rollout costs attention. Some areas are hit by every single one. In my experience, an area rarely tolerates more than two major changes at once. The right to postpone one of them belongs in your rules.

6 · What Is Still Open? Conflict and Amending

Article 8 · Conflict: Who decides when two good reasons collide?

AI causes a lot of disputes in organizations. The business unit wants a feature because it saves days. IT will not switch it on while it is unclear where the data goes. Both rightly appeal to the same value: responsibility.

One sentence that hardly any mission statement contains helps here: what a value does not mean. For one person, diligence means everything gets checked; for another, it means the important things get checked thoroughly. If you write down in advance what diligence does not mean in your organization, you have settled a good part of the dispute.

For the rest, a conflict needs one of four paths. You can postpone it for a set time, with a mandate saying what will be clarified by when. You can keep the rule with reasons and record the objection, or change it through the path from Article 9, also as a limited exception. You can hand the dispute to a higher or independent body. Or you can explicitly take the question out of the procedure.

Each path needs a name, a date, and the burden that will be accepted until then. A conflict that is sat out comes back, usually at the worst possible moment.

Article 9 · Amending: Is a software update a change to your rules?

The AI company Anthropic has published a constitution for its language model Claude, with a purpose, extensive reasoning, and an order of precedence, more clearly drawn than in most companies. It says the document will change. For now, Anthropic gathers feedback on it from Claude models and intends to develop more formal mechanisms. A fixed procedure for who may propose a change and who decides on it is not described.

The same holds for companies: a constitution without a way to amend it is a declaration of intent.

For an AI rollout, this means first of all that a changed setting in a system is a change to your rules as soon as it alters what gets pre-sorted, rated, or decided. It therefore follows the same path as any other rule change, prepared by the system's owner, who knows which rules are built into it. Security updates and changes the vendor installs without your involvement need a fast lane: the owner checks right away what has changed, and the next meeting confirms or corrects it.

The path itself needs little. Anyone may propose a change, without going through channels. A small group decides, with seats for employees, filled by the works council where there is one, at a fixed date each quarter that takes place even when nothing is on the agenda. A procedure that runs only when someone pushes falls asleep.

For emergencies, add a crisis stage. With AI, that usually means shutting the system down and falling back on the manual process, after a vendor outage, for example, or the discovery of a systematic error. A reduced operation or another system can also step in. A fallback nobody has ever run is no fallback in an emergency, so it has to be practiced. The stage is time-limited, and whoever declares it does not extend it themselves.

7 · Six Anchors for Every Rollout

Nobody knows in advance what is coming. But a set of rules can be built at certain points so that the organization can adapt quickly enough when the time comes. I distinguish six such points and call them anchors, because they are fixed in the rules themselves.

For an AI rollout, they become a test: does the rollout strengthen the anchor, or hollow it out?

Anchor	What it does	The question for the rollout
Reserve	Time and money no one has planned for	Is a protected share of the freed-up time left over?
Way back	Decisions can be withdrawn before they cause harm	Can the consequences of decisions already made be corrected, and is there a practiced fallback?
Failover capacity	A person drops out, and the process keeps running	Does the skill stay with more than one person, or does it migrate into the system?
Signal path	What is coming reaches someone who may decide	Does an observation about AI use reach someone who has to respond?
Learning	Deviations and experiments become knowledge	Is every error and every aborted trial reviewed, without assigning blame?
Load limit	The operation can absorb the change	How many rollouts hit the same area at the same time?

The control reflex is especially quick to appear around precautions. Whoever writes down foreseeable risks wants to regulate them. A year later, the rulebook is thicker than before. The six anchors work differently. They establish who may act, how much slack remains, and how something comes back. They are not free, because reserve, backup, and practice take time. That time, however, can be counted and planned in advance.

8 · In What Order

Sequence matters more than any single step, and the most important rule is quickly stated: **Set the way back before you cut.** Whoever removes checkpoints first and decides afterward who may restore them pushes rules into the unwritten, where whoever has always decided without a mandate keeps deciding.

The second rule: small, reversible steps before large, binding ones. The large building blocks come once the small ones have held.

The first three days require no resolution and no budget of their own, only working time and access to the systems. On day one, list your systems; the project list will not help much. For each piece of software, ask whether it gained a feature in the past year that proposes, pre-sorts, rates, or drafts.

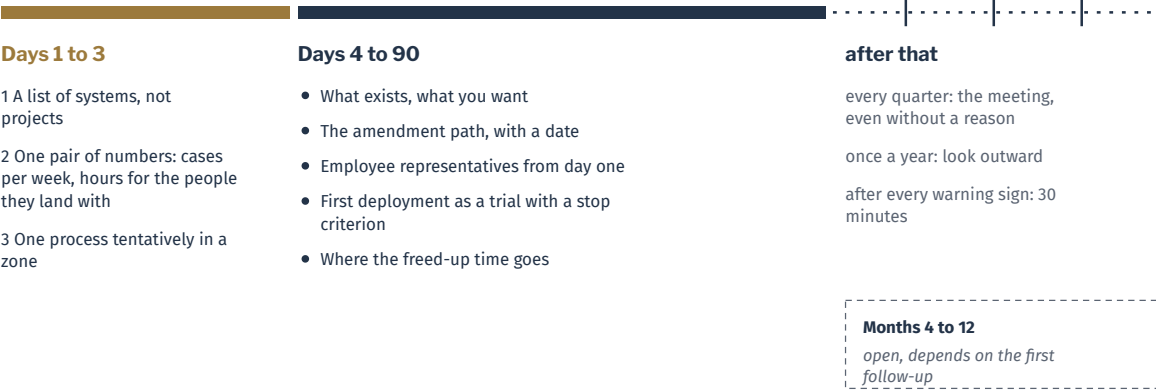
On day two, get a pair of numbers for the AI use that has been running longest: how many cases it produces per week, and how many hours are allocated to the people they land with. On day three, tentatively assign a single process to a zone, specifically the one where the answer is hardest for you.

In the first ninety days, start with the sentences that already steer a rollout, then take the two separate steps from Article 1: what already exists and what you want. Then set the amendment path from Article 9, with a date in the calendar before there is any occasion for it. Employee representatives are involved from the start, long before any agreement is negotiated.

The first deployment runs as a trial with a stop criterion, and for every entry-level role it is settled what newcomers will now practice on. Before any saving is booked, it is settled where the freed-up time goes.

After that, rhythm counts. The quarterly meeting takes place even without an occasion, and once a year you look outward with someone from outside. Every warning sign gets a half-hour review, and after three months comes the first follow-up measurement, with the same numbers as on day two.

In what order



Set the way back before you cut.

Small, reversible steps before large, binding ones.

9 · How You Know It Holds

Whether a rollout keeps its promise can be tested, for the process under review, against one sentence that can be proven wrong: **Lead time goes down, and the share of cases in which an error is found after the fact does not go up.** Compare similar cases, each with a follow-up period of equal

length, and count serious errors separately. Add the total effort across all units involved; otherwise shifted work stays invisible.

Behind that sentence are two bills paid by different people. The waiting time is paid by those who wait for a decision; the late errors are charged to the decision to skip checks, often to someone who never knew the check had been skipped.

Because less checking also finds fewer errors, the sentence needs a counter-check. Once a quarter, pull cases that have already been approved and have someone else check them thoroughly. Twenty cases are a start, but only a search tool: at an error rate of five percent, a sample of that size finds no error at all more than a third of the time. Where much is at stake, take a larger sample. Count only errors the approval should have caught under its mandate. If the follow-up regularly finds such errors, that is a warning sign. Then the checking procedure, the time for it, and the error rate you are willing to accept belong on the table.

A measurement has three possible results: relieved, warning sign, or undecided. Undecided is an honest result, as long as it comes with a date for the next measurement. Settle before the trial which deterioration is still acceptable. If the cases checked cannot rule it out, the result is undecided.

Four indicators show where you stand on checking load, and most of them can be drawn from existing data and documents. Many systems, however, record only how long a case waits, not how long someone works on it. A short work sample closes that gap. Evaluate them per checkpoint, not per person; where there are employee representatives, involve them beforehand. How long does an approval take today compared with eighteen months ago? How many cases arrive at a checkpoint per week? What share of cases is decided one level higher? And is there a document that states what an approval actually vouches for?

If preparation and processing have become faster and approval slower, that points to a shift in load. Then check volume, case mix, and the working time actually spent. If both have become faster, check whether anyone is still checking.

Each article also has a measure. For Articles 4, 5, 6, 8, and 9 it is a number; for the others, evidence that the rule exists. Whether it works, the table does not show. That only shows in cases you look at closely, as in the counter-check above. You can take this table into your next meeting.

Article	How you measure it
Foundation	For every system, a written assumption of what you trust it to do, with a review date
1 · Purpose	For every AI use, one sentence on how it serves the purpose
2 · Membership	For every role with a checking task: authority, access, and time, plus a trained backup; for every entry-level role, a task people still practice on themselves
3 · Limits and Precedence	For every hard limit, a case in which it would have cost something, and the name of the person who pays; in every evaluation, a conversation about the path to the result
4 · Deciding	Seconds per case at every checkpoint in Zones B and C
5 · Voice	Share of objections that receive a reasoned answer by the promised date
6 · Knowing and Checking	Rejection rate per checkpoint; share of system rules with an owner
7 · Allocating	A written decision on where freed-up time goes, before the first saving
8 · Conflict	Share of open conflicts with an assigned path, name, and date
9 · Amending	Share of updates with new features that someone reviewed before installation or through the fast lane

And one conversation starter that costs nothing: ask a randomly chosen person, ideally someone without a management role, why a particular process sits in Zone C and not in Zone B. If they can explain it, that is a good sign the framework has reached the people who do the work. If they cannot, check whether the reasons and their application have reached that level. Otherwise the framework easily ends up followed the way rules from above are followed: exactly as far as someone is watching.

10 · The Legal Picture

As of October 1, 2026. This section ages faster than the rest.

What applies is in the law. Your own rules should say who implements it in your organization and how; they do not need to repeat its wording. You need rules of your own where you want to do more than the law requires. This overview covers the European Union and one international body. For other jurisdictions, see organization-under-ai.com/law.

AI literacy (EU). Article 4 of the EU AI Act has applied since February 2, 2025, and since July 27, 2026 in the version amended by Regulation (EU) 2026/1744. Providers and deployers must take measures that support the development of AI literacy among their staff and others operating AI systems on their behalf. They are expressly not required to guarantee a particular level for each individual. The practical consequence is what Article 2 of this paper describes: whoever is to check needs authority, access, and time, plus practice. This is a conclusion, not a list from the law.

Human oversight (EU). Article 14(4) requires a high-risk AI system to enable the people overseeing it, as appropriate and proportionate, to exercise effective oversight: to understand its capabilities and limits, to remain aware of the tendency to over-rely on it, to interpret its output correctly, to disregard it, and to interrupt operation safely. This is first addressed to how the provider designs the system. Under Article 26(2), deployers must assign oversight to people with the necessary competence, training, and authority, and give them the necessary support. These obligations apply from December 2, 2027 for high-risk systems listed in Annex III, which include systems used in employment, and from August 2, 2028 for systems under Annex I. Transitional rules in Article 111 apply to systems already placed on the market or put into service.

Transparency (EU). Article 50 sets out information, labeling, and disclosure obligations. People interacting directly with an AI system must be told so, unless it is obvious. Providers mark synthetic content in machine-readable form. Deployers disclose emotion recognition, biometric categorization, deepfakes, and certain texts on matters of public interest. These obligations have applied in principle since August 2, 2026; for the labeling obligation under Article 50(2), systems already on the market before that date have until December 2, 2026.

Time (international). In its September 2025 resolution on meaningful human oversight of decisions involving AI systems, the 47th Global Privacy Assembly, the international network of data protection authorities, recommended giving the person exercising oversight “sufficient time to undertake the oversight and a reasonable workload.” The resolution is a recommendation, not a binding requirement. It names no number.

Automated decisions (EU). Under Article 22 GDPR, no one may be subjected to a decision based solely on automated processing that produces legal effects concerning them or similarly significantly affects them; Article 22(2) contains exceptions. On December 7, 2023, the Court of Justice of the European Union held that the automated calculation of a credit score by a credit agency can itself be such a decision if a third party draws strongly on that score in deciding whether to enter into, perform, or terminate a contract (C-634/21, SCHUFA). That a merely formal human confirmation does not rule out an automated decision was already set out by European data protection authorities in their 2018 guidelines. What matters is what the human review actually does.

Employee representation (example: Germany). Where works councils exist, AI usually triggers co-determination. Under German law, a works council co-determines the introduction and use of technical systems objectively suited to monitor employees’ behavior or performance; no intent to monitor is re-

quired (Federal Labor Court, July 16, 2024, 1 ABR 16/23). Since 2021, the German Works Constitution Act explicitly names AI in the provisions on information about planned work processes, on calling in experts, and on selection guidelines. German occupational safety law also requires every employer, regardless of size, to assess psychological strain at work (Section 5(3) no. 6 ArbSchG); a rollout that multiplies the volume at a checkpoint belongs in that assessment.

11 · Working With Me, Online

Read. The mechanism behind Articles 4 and 6, with all the evidence, counter-evidence, and seven worksheets, is in *Cheap to Make, Costly to Check* (2026), the first volume of the series Organization under AI. Available on [Amazon](#) as a paperback and for Kindle. The companion page with sources, the claims and their limits, and the legal overview is at [organization-under-ai.com/cheap-to-make-costly-to-check](#). A workbook on organizational constitutions is in preparation.

Work with me. All formats below run entirely online, by video and on data you already have. No travel, no on-site days, and your data stays in your systems. Details and prices are at [guidobosbach.com/en/work](#).

If you want to ...	the online format is
know how large your checking load is and where it lands (Article 6)	the Checking Load Assessment : two numbers from data your systems already hold, and a list of functions nobody decided on
first see where your management system stands overall	the Management Baseline : an evidence-based picture of your management system, benchmarked against 458 organizations
make one decision well that does not fit the usual categories (Articles 3, 4, and 8)	Decision Sparring , one session or a series
bring the ten questions into a leadership meeting or online event	a Virtual Keynote , 40 to 60 minutes including discussion

To start, book a free 45-minute Clarity Call by video via [guidobosbach.com/en/contact](#). If nothing fits, I will say so.

12 · Forty-Eight Reports, Once More

At Thrushwood, a single question would have been enough, asked before the rollout: who reads the forty-eight reports, and when? It belongs to Articles 6 and 7, and it costs half an hour as long as nothing has been introduced yet. After the first missed finding, it costs considerably more.

Most organizations have rules that were written for twenty reports a day. At twenty reports, nobody asked whether they were enough.

AI did not ask the rules a new question. It made their checking promises more expensive than the old rules can pay for.

About the Author

Guido Bosbach is an organizational architect and coauthor of *UNMANAGED* (LID Publishing, London). His work is forward-looking organizational analysis: finding the point where things will get stuck while the answer is still cheap.

About this paper. The English text was written directly in English with the help of AI, based on the German original, and was read and approved by the author. The thinking, the position, and the experience behind it are his.

License. Free to share, unchanged, with attribution, for noncommercial purposes (CC BY-NC-ND 4.0). Research current as of October 2026 unless noted otherwise.

Sources

- Agrawal, Ajay; Gans, Joshua; Goldfarb, Avi (2018): *Prediction Machines: The Simple Economics of Artificial Intelligence*. Boston: Harvard Business Review Press.
- Bainbridge, Lisanne (1983): Ironies of Automation. *Automatica* 19(6), 775–779. DOI
- Brysbaert, Marc (2019): How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language* 109. DOI
- Court of Justice of the European Union, judgment of December 7, 2023, C-634/21 (SCHUFA Holding).
- European Commission, AI Act Service Desk: FAQ “When does enforcement start?”
- Federal Labor Court of Germany (Bundesarbeitsgericht), decision of July 16, 2024, 1 ABR 16/23.
- Ford: statements by Charles Poon, June 2026, reported by *Bloomberg* (“Ford’s AI Hiccups Lead Carmaker to Rehire ‘Gray Beard’ Engineers,” June 25, 2026), *TechCrunch*, *Fortune*, and *Forbes*. No verbatim transcript is available.
- Ford job cuts: CNN and NPR (May 20, 2019); *TechCrunch* and *The Washington Post* (August 22, 2022); CNBC and Reuters (February 14, 2023).
- Global Privacy Assembly (2025): Resolution on meaningful human oversight of decisions involving AI systems. 47th Closed Session, September 2025, p. 4.
- Hacker, Winfried (2026): *Arbeits- und Gesundheitsschutz bei informationsverarbeitenden Arbeitsprozessen mit Künstlicher Intelligenz*. baa: Fokus, February 2026. DOI
- National Institute of Standards and Technology (2023): *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1, MANAGE 2.1 and Appendix C. DOI
- Niederhoffer, Kate; Rosen Kellerman, Gabriella, et al. (2025): AI-Generated “Workslop” Is Destroying Productivity. *Harvard Business Review*, September 2025. online
- Regulation (EU) 2016/679 (General Data Protection Regulation), Article 22.
- Article 29 Data Protection Working Party (2018): Guidelines on Automated individual decision-making and Profiling, WP251 rev.01, endorsed by the European Data Protection Board.
- Brynjolfsson, Erik; Chandar, Bharat; Chen, Ruyu (2025): Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence. Stanford Digital Economy Lab, Working Paper, August 2025, revised August 2026. online
- Dietvorst, Berkeley J.; Simmons, Joseph P.; Massey, Cade (2015): Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. *Journal of Experimental Psychology: General* 144(1), 114–126. DOI

- Logg, Jennifer M.; Minson, Julia A.; Moore, Don A. (2019): Algorithm Appreciation: People Prefer Algorithmic to Human Judgment. *Organizational Behavior and Human Decision Processes* 151, 90–103. [DOI](#)
- Anthropic (2026): Claude's Constitution. [online](#)
- German Occupational Safety and Health Act (Arbeitsschutzgesetz), Section 5(3) no. 6.
- Regulation (EU) 2024/1689 (AI Act), Articles 4, 14, 26, 50, 111, Annexes I and III.
- Regulation (EU) 2026/1744 amending the AI Act ("Digital Omnibus on AI"), Official Journal of July 24, 2026, in force since July 27, 2026.
- Sheridan, Thomas B.; Verplank, William L. (1978): *Human and Computer Control of Undersea Teleoperators*. MIT Man-Machine Systems Laboratory.
- University of Chicago Law School (2026): Rethinking Legal Education in the AI Era. Statement of July 9, 2026. [online](#)
- Bosbach, Guido (2026): *Cheap to Make, Costly to Check*. Organization under AI, Volume 1.