

Organization under AI

Volume One

An AI ranked forty-one bids.
A human read ten and signed.

Cheap to Make, Costly to Check

How AI shifts the real cost of decisions —
and who in your organization ends up paying it

Guido Bosbach

Bosbach Edition

Cheap to Make, Costly to Check

How AI Shifts the Real Cost of Decisions — and Who in Your Organization Ends Up Paying It

Short version

An excerpt from Cheap to Make, Costly to Check (Organization under AI, Volume One) by Guido Bosbach. The full book is available on Amazon in paperback and for Kindle. What this short version leaves out, and the links, are at the end.

Forty-one bids and two days

A custom machine builder in southern Wisconsin: six hundred employees, third generation, family owned. In the spring it put the maintenance of its production equipment out to bid. Three-year term, roughly \$250,000 a year, three-quarters of a million over the life of the contract.

Four years earlier, the same RFP had drawn nine bids. This time forty-one came in. Writing a bid used to cost the bidder something: forty pages of structure, references, numbers, and a plausible account of how the work would be organized represented three days of somebody's time. By the spring of that year, forty pages cost the sender almost nothing.

Ines Kessler, who runs procurement, had two days on the calendar, as she had four years earlier. Nobody had decided that forty-one bids would get the same two days as nine. The number in the calendar was simply the one that had been there the last time.

She read the top bid in full that first morning, made notes in the margin, called a reference. By the fourth she was skipping the sustainability section. By the eighth she was skipping the concept section too. On the afternoon of the second day, thirty-one bids sat unread on her desk, and she was working down the list: price, rank, next.

The order of that list came from a scoring function inside the purchasing software. It had arrived in March with an update. On a Tuesday evening, during the maintenance window, Tim Whelan from IT installed the new version, created two test transactions, and confirmed that they ran. He checked the item off his list and went home. The description of the new function was in the release notes, under Improvements. Nobody had looked into what it did. Nobody had switched it off, and nobody had switched it on.

Frank Ostrowski's bid ran three pages. His forty-person shop had kept this production floor running for fourteen years, nights included. The bid said what his crew would do and what it would cost. It did not say that they come out at two in the morning, that they know the two oldest machines on the floor, and that they stock parts the original manufacturer stopped making in 2011. None of that had ever needed to be in a bid. The scoring function sorted on what it could see and placed him nineteenth. The award went to the bid ranked second.

The company and the people in this scene are composite figures. The company does not exist and none of its numbers is evidence; the case is a way of seeing the mechanism at work. The dynamics are recurring ones.

The mechanism

Reading those thirty-one bids properly, at the hour and a half a careful read takes, would have taken about eight more working days. The purchasing software produced the ranking of all forty-one in a few seconds.

Put those two numbers next to each other and you have more information about the state of your organization than any explanation of artificial intelligence. One step in the process is now nearly free. The step that checks and corrects it costs what it always did.

Producing got cheap. Checking did not. And the two are paid for by different people.

What is left over — reading, verifying, deciding whether the thing in front of you is any good and whether you will stand behind it — is the **checking load**, and it does not evaporate; it lands somewhere. In this case it amounted to about eight workdays, and it went to Ines Kessler, because she was at the end of the chain and that is where the signature was due.

The underlying economics are older. In *Prediction Machines*, Ajay Agrawal, Joshua Gans, and Avi Goldfarb laid out the mechanism: when one input to a decision gets dramatically cheaper, its complements become more valuable. What arrives now is finished work, in quantity, and that turns a question about price into a question about capacity. Ines Kessler holds the judgment in this story, and she is no better off: what went up is the number of items in front of her, while the number of hours she has stayed where it was.

Checking stayed expensive for three reasons that have nothing to do with one another.

A budget. A person reads nonfiction at roughly 238 words per minute. Marc Brysbaert arrived at that figure in 2019 in a meta-analysis covering 190 studies, and the oldest measurements he compared, from 1898 and 1901, are in the same range. More people, better tools, and more practice narrow the gap. They do not close it anywhere near the rate at which a machine produces.

Knowledge. Checking means holding a result against something that is not stated in it: what went wrong in this department two years ago, who is leaving in the fall, whether this supplier shows up at two in the morning. That **floor knowledge** sits with people rather than in the files. Tools can

shrink the work around it. The judgment step tied to it gets smaller, and it does not scale with production.

Attribution. Whoever is responsible for a result cannot delegate the check to the same system that produced it, and a second system does not become an independent check merely by being second. Where both draw on similar data and the same pattern of plausibility, their errors stay correlated, and both treat an answer that sounds convincing as evidence of correctness.

The argument has a boundary, and it decides whether any of this applies to you. Where correctness can be written down in advance — a test suite, a sum, a format — both costs fall, and the book has little to add. It holds where three things come together: a system produces or sorts faster than anyone can check, correctness cannot be fully specified beforehand, and the consequences land on people, which means somebody has to answer for them.

One more thing about the ranking. Everyone involved would have called it preparation, since a human read the bids and a human signed. Yet its order determined what her attention landed on and what it never reached. That is **a decision in all but name**, and the test for it is what people actually deviate from. At the machine builder, for this award, nobody had counted.

Who gets the bill

The time saved stays with whoever produces, and the time spent falls to whoever checks. The shift does not appear in cost accounting, because it crosses a department boundary, a role, or a level in the hierarchy, and nobody books it there. The supplier who generates a forty-page bid in ninety seconds is not the person who has to read it.

The load lands where responsibility already sits, on individual people with names and roles, and there it looks like a personal problem, like overload or slipping standards. Anyone who reports it appears to be reporting their own inadequacy.

Where a system meets people, three states recur, and each sends the bill to a different address.

Pass-through. The machine's speed transfers to people with nothing to slow it down: more items each week, the same heads, the same hours. The bill lands on the people at the review desk. They pay in time, then in attention, and eventually in the care they can no longer afford. Ines Kessler paid in all three, in that order, over two days.

Detached sign-off. After a first incident, somebody formalizes the process: a review step, a named person, a field for their initials. The system runs in its own lane and a person signs without having taken any part in how the thing in front of them came to be. This state passes every audit that asks who was responsible, and the bill lands in the future. In October one of the two old machines failed at night, and the new contractor sent a technician the next morning. The question of who agreed to that service level has an answer in the system: the person who signed.

Bottleneck. The human pace is made binding everywhere. At a different machinery company, every use of a language model in customer communication had to clear legal review, which took nine days on average. Eleven requests came in the first six months, three in the twelve months after. The bill lands on the benefit the organization never collects, and meanwhile use without approval and without any record grows. That is pass-through again, only invisible.

Every step was a reasonable answer to the pain of the state before it, and each one led into the next. There is ready-made language for all three. What is missing is a name for the fourth option: deciding where the checking load lands.

Two questions

The load cannot be abolished as long as producing stays cheap and checking does not. What is on the table is the assignment: an organization can determine where the load arrives, who answers for it, and how large it is permitted to grow at that position. Two questions come with it, and they get confused constantly.

Who owns the checking load? That is a question of assignment. For the maintenance RFP, the answer was nobody. No document said what a purchasing approval covers, so the answer defaulted to the person holding the pen.

And can whoever owns it actually carry it? That is a question of capacity: time, access, a second source, practice, the standing to say no. For Ines Kessler the answer was no, even if the assignment had been perfectly explicit. Two days, forty-one bids, no rule permitting a declared sample, and the one person who knew what the bids did not contain, the head of maintenance, was shown only the top ten.

Both questions can be answered independently, and that is where the difficulty lies. An assignment without the capacity to carry it produces a named person answerable for something they cannot change. Capacity without an assignment holds until the week somebody is on vacation.

Permission without capability is paper. Capability without ownership is luck.

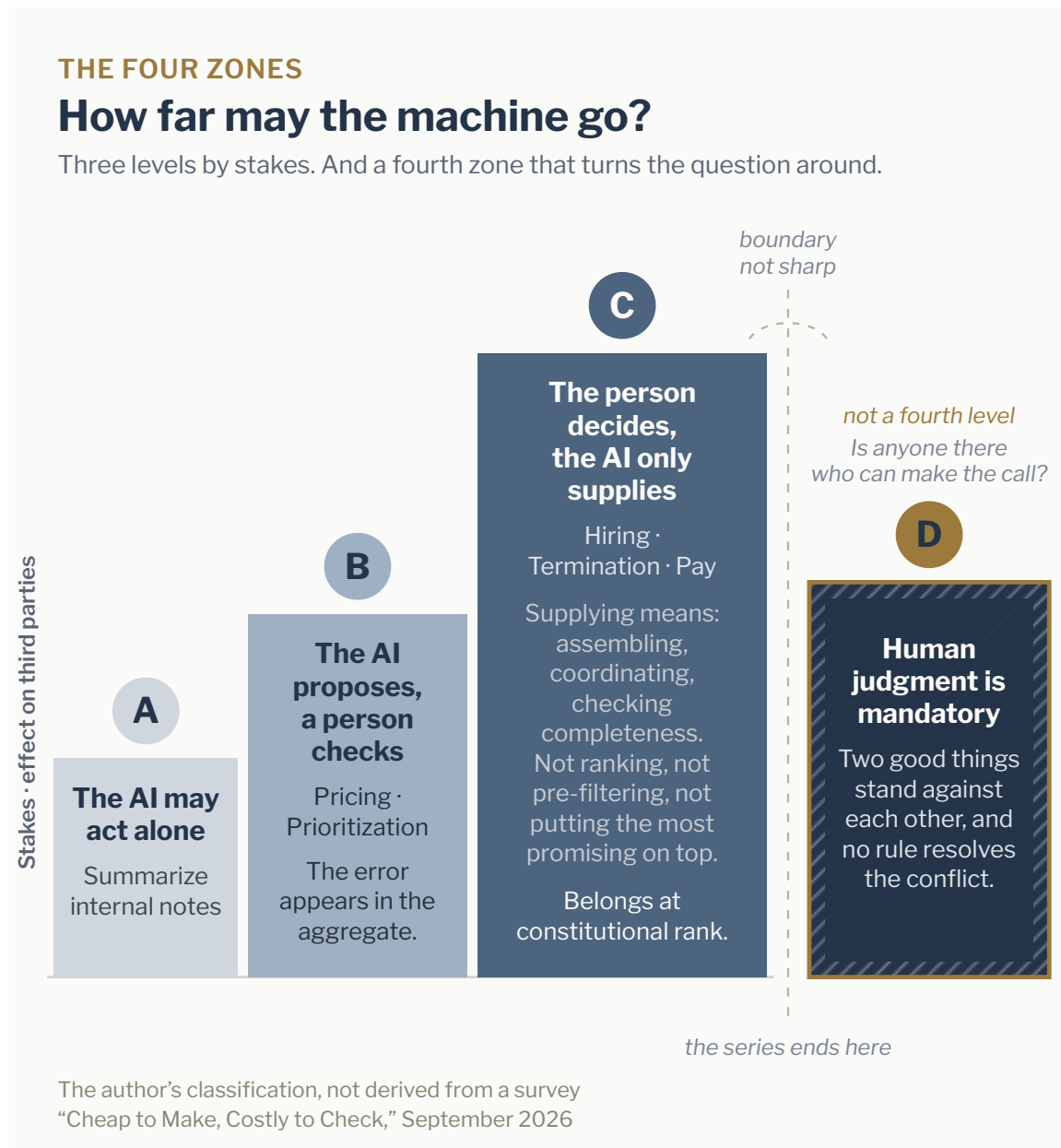
Who owns it: four zones

Before any zone, five triggers settle whether a use needs an explicit rule at all: reversibility, an existing duty, information asymmetry, scale, and spillover. One yes is enough. The scoring function trips four of the five. Nobody asked a single one of them, because a software update gave them no reason to.

Where a rule is needed, four zones say how strictly the decision is handled. **Zone A:** the AI may act alone, and speed is the value. **Zone B:** the AI proposes and a person checks, and the check is worth exactly as much as the real possibility of rejecting the proposal. **Zone C:** the person decides and the AI only supplies. Supplying means assembling documents and checking completeness; it

does not mean ranking or putting the most promising at the top. **Zone D:** human judgment is mandatory and has to be trained, because two good things stand against each other and no rule resolves the conflict.

The maintenance award belongs in Zone C. The approval file records a human decision. What Ines Kessler had in front of her was a ranking. The zones are a working frame; they are set by the author and derived from no statute or study.



Can they carry it: four paces

What most companies have instead of a decision about speed is software that got faster and people trying to keep up. That is a decision by default. There are four paces. At **machine pace** a process runs at system speed with no person in between. At **review pace** it is fast and stops at de-

fined points. At **human pace** the speed follows the person deciding. At **maturing pace** time is the substance of the process itself, as with trust or a succession; anybody invoking it has to be able to say what moving faster would break.

An assignment is worth exactly as much as the pace it is allowed to work at. The failure case that is everywhere is Zone C at review pace. On paper everything is in order: the decision is assigned to a human, a named person carries it, there is a signature. In practice the process runs at the system's speed, because the inflow was never capped. The named person has eighteen minutes per bid. Their decision is formally a Zone C decision and in fact a confirmation.

Where to start

None of this needs a project. Two steps take less than a day.

The list of your systems, not the list of your AI projects. Walk the programs people actually work in and ask of each one: did it gain a function in the last year that suggests, pre-sorts, scores, drafts, or ranks something? No project list can answer that. At the machine builder, the list the CEO received first had two pilots on it. When Tim Whelan went through the systems themselves, he came back with twenty-one entries. For each item, note whether something is being prepared there or decided. Count before you explain what you are looking for; somebody who knows what you hope to find will find it. It takes an afternoon.

One pair of numbers. Pick one checkpoint where a person approves or signs off on things a system prepared. Get two numbers: how many items reach it in a week, and how many minutes a serious check of one item takes, meaning the time the person would need to form a view. The average time it currently takes is the wrong number. Multiply, and compare the result to the hours that person has. Then find the same two numbers from two years ago. What you are looking for is the change in the gap, and whether anyone decided it. Take the numbers averaged across the team and never by name. It takes two hours if the data is in a system.

The book's minimal start adds a third day: provisionally assign a single process to a zone, as a diagnostic with no decision and no binding effect. Choose the one where the answer is hardest for you to give; your most important process can wait.

How you would know this is wrong

The argument rests on a single claim, and it could be refuted: checking has not become cheaper to the same degree as producing. If machine reviewers reliably found errors of judgment, checking costs would fall at the same rate and the situation would be a transitional problem. Where a cheap, independent criterion for correctness exists, the claim is already refuted, and that marks the boundary of the book's scope.

Four measures will tell you where you stand, and all four come from existing data. How long does an approval take today, compared with eighteen months ago? How many approvals does one person issue per week? What share of items is escalated one level up? And is there a document anywhere stating what an approval actually vouches for? If preparation got faster and approval got slower, you are looking at the finding. If both got faster, check whether anything is still being reviewed.

The evidence has a gap, and the book names it. No organization in it built the framework described and then showed, afterward, that something ran differently. Damage that does not happen generates no file. If you build the framework and can describe in two years what changed, you supply the evidence the book does not have.

What is not in this short version

This excerpt carries the mechanism, the case, the two questions, and a first diagnostic. The book adds the three layers of AI use in detail, the five triggers worked through, the core commitments, the pace rule and the coupling points between fast and slow work, strain, configuration power, breadth of view, unassisted practice, the six seams where ownership and capacity cancel each other out, the maintenance award read a second time and Frank Ostrowski's call answered, seven worksheets, a comparison of the two questions across five legal systems, and the sources with their limits.

You may pass this excerpt on freely, unchanged and with the author named. The complete text is available only in the book.

Companion site: organization-under-ai.com/cheap-to-make-costly-to-check

Paperback: amazon.com/dp/B0HJB7DXRS

Kindle: amazon.com/dp/B0HJB3FDG1

The English edition of the book was written by an AI system working from the author's German manuscript, his research, and his decisions on every question of substance. The author reviewed and revised the result and is responsible for it. This short version was condensed from that English text.

Guido Bosbach is an independent management consultant. For more than fifteen years he has worked with mid-sized and owner-run companies on building decision paths that hold under pressure.