One Workforce, Two Kinds of Worker: What Operational Excellence Already Knows About Managing AI Agents

Summary. Most organizations took on a new kind of worker in the last eighteen months. The agents arrived without a requisition, without an onboarding plan, and without anyone accountable for making them better. The industry has answered by arguing about tools, which is precisely the mistake the industrial world made when it copied Toyota's practices and left the system behind. This article makes the case that the disciplines built to develop people, setting a clear standard, making problems visible, going to see the actual work, and growing capability through teaching, are the same disciplines that make digital workers perform. by Nathan Holt

August 2026

The New Hire Nobody Onboarded

Sometime in the last eighteen months, most companies took on a new kind of worker. There was no requisition and no first day. The agents arrived one team at a time, and they are now doing real work inside real processes. Very few organizations have asked about them the question they would ask about any other new worker: how do we know when they are good, and who is responsible for making them better?

I saw the gap clearly in am AI Agentic training class this year. The slide on the screen listed what an agent needs before you turn it loose. Check that it behaves as intended in different situations. Set boundaries so it cannot do harm. Watch it while it works. Let a human step in at the critical moments and escalate what matters.

I have written that plan before. We called it an onboarding and certification plan (Training Within Industry, in the old language), and it was for a new operator on a production line.

We Have Run This Pattern Before

In 1999, Steven Spear and Kent Bowen published a study of Toyota that has aged remarkably well. Their puzzle was that Toyota's practices had spread across the industrial world for two decades while Toyota's results had not. Companies had copied the cards, the boards, the andon lights and the daily huddles, and most of them were still ordinary.

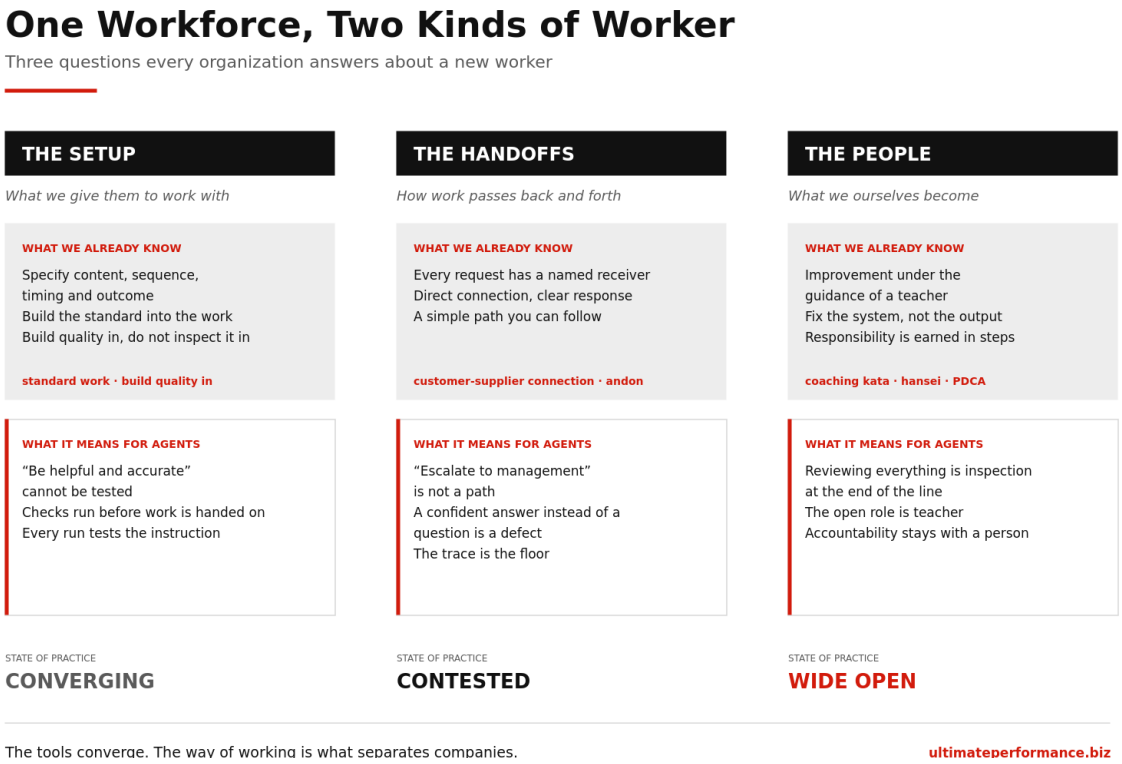
Their conclusion was that observers kept confusing the tools and practices with the system underneath. The visible parts travel easily. The way of working does not. That is exactly where we are with agents. Nearly all of the conversation is about tools. Which model, which framework, which orchestration platform, which set of instructions. Almost none of it is about the way of working that the tools sit inside.

This matters because capability is not performance. A capable model, like a capable hire, produces business results only through the layer built around it: what we expect of it, how we see what it is doing, who is accountable for the outcome, and how it gets better over time. Companies do not underperform on AI because their model is weak. They underperform because nothing was built around it.

Three Questions, in the Order We Usually Get Them Wrong

Edward Donner, in a recent video asking whether humans are now just AI's reviewers, offered a useful frame borrowed from an older one. He recalled a manager of his who held that major transformations come down to platform, process and people. For work with agents, Donner recast those as the setup we give the agent, the way work passes back and forth between the agent and us, and what we ourselves become.

What stayed with me was his read on progress. On the setup, the industry is converging fast and the big questions are close to settled. On the handoffs, there is real argument but the shape of an answer is visible. On the people, it is wide open. Nobody agrees on what the human role even is.



Anyone who has led an operational excellence rollout should feel a chill reading that list, because it is the same order in which we got that wrong. The tools went in first and went in everywhere. The way work passed between people got some attention. The human system, what leaders did every day and how capability was grown, got the least. That is precisely where the value leaked out for thirty years.

We have the chance to run it in the right order this time. Here is what each of the three asks of an executive.

We are copying the cards again, faster and with a much larger budget.

The Setup: Write Down What Good Looks Like Before the Work Starts

Spear and Bowen found that at Toyota, work was specified as to content, sequence, timing and outcome, and that the specification functioned as a hypothesis. Every time the work ran, it tested the claim that doing it this way produces that result. Deviation was not a nuisance. It was the signal.

Translate that into agent terms and you get the most useful sentence I can offer a leadership team. If you cannot say what a good result looks like before the work starts, you cannot tell whether you got one. Most agent instructions in use today fail that test. "Be helpful and accurate" is not a specification. It cannot be checked, so nothing can be detected, so nothing can be improved.

The teams getting real results do something familiar to anyone who has run a quality system. They do not ask the agent to be careful. They build the standard into the work (standard work, owned at the point where the work happens), so the job cannot be handed over until the checks have run and passed. The agent has no more discretion about meeting the standard than a machine has about a fixture that will not accept a part loaded backwards (mistake proofing, or poka-yoke). That is the oldest quality lesson we have. Build it in, do not inspect it in.

The Handoffs: No Request Should Go Unanswered

The second thing Toyota specified was the connection between whoever needs help and whoever provides it (the customer-supplier connection, which exists inside a company as much as between companies). Direct, with an unambiguous way to send a request and receive a response. No grey area about who is asked, how quickly, and how the asker knows the message landed. When something is everyone's problem, it becomes nobody's problem.

Consider that last requirement: escalate what matters. It is the kind of instruction that turns up in almost every agent readiness checklist, and it is the kind that quietly fails. Escalate to whom, exactly? Within what time? And how does the agent know the escalation was received rather than dropped? On a production line, that ambiguity is where quality quietly dies, and it dies the same way here. An undefined escalation path is an escalation path that will not be used.

There is a subtler version of the same failure, and it is the one I would put in front of a board. An agent that returns a confident, plausible answer in a situation where it should have stopped has not made an error of knowledge. It has made an error of connection. It answered a question it was not sure about instead of asking for help. That is the line worker who sees the defect and lets the work move to the next station. Everything we know about stopping the line (jidoka, and the andon cord that triggers it) applies without modification.

The third piece is the path the work travels. Toyota required it to be simple, direct and specified in advance, so that a deviation would be obvious. Much of the current enthusiasm for open-ended autonomous agent swarms runs directly against this. If you cannot say which agent or which tool handled which step, you do not have a path, and without a path there is no deviation to see.

Which leads to the point most leaders need to hear. A dashboard is the report. The record of what the agent actually did, step by step, is the floor. Most executives I speak with are grading agents on averages and have never once watched one do a single piece of work from beginning to end. Go and see the work (Gemba, and it is still the highest-yield hour a leader spends). It has never been easier, and it has never been skipped more often.

The People: The Teacher, Not the Inspector

Here is the asymmetry that is making people anxious. Producing work with an agent is fast and cheap. Checking that work is slow and expensive. Follow that logic and everyone in the organization slowly turns into a full-time reviewer of output they did not produce.

We have a blunt answer to this. If you are inspecting everything, you have moved quality to the end of the line, which is the one position you should never accept. Inspection, while often necessary as capability builds, does not create quality. It discovers the absence of it later and at higher cost.

The way out sits in the last of Spear and Bowen's four rules, and it is the clause everyone skips. Improvement should be made using the scientific method (PDCA, run as a real experiment), at the lowest possible level in the organization, and under the guidance of a teacher.

Agents can already run the first two. A well-built agent can review its own work, learn from what went wrong (hansei, honest reflection after the fact), and improve its own instructions so the same problem does not recur. That is improvement at the lowest possible level, done more reliably than most companies manage with people. The third clause is empty. Nobody has defined the teacher, and that is the open human role the industry has not filled.

A teacher does two things a reviewer does not, and anyone who has run coaching kata will recognize both. First, when something goes wrong, a teacher does not simply correct the output. They improve the system so that particular mistake cannot happen again, which is the only kind of fix that compounds. Second, a teacher grants responsibility as it is earned, on evidence, rather than all at once or not at all.

Four Levels of Earned Autonomy

Nobody hands a new operator the run of the plant on day one. Scope is earned in steps, the way a skills matrix works, and every step keeps the ability to stop.

1. Observed

The agent runs while a person watches, and nothing acts on the output. This is where you learn what the work actually looks like rather than what you assumed it would look like.

2. Checked Before Acting

Every output is reviewed by a person before it takes effect. Slow and expensive, and appropriate exactly as long as the evidence is thin.

3. Bounded, With Spot Checks

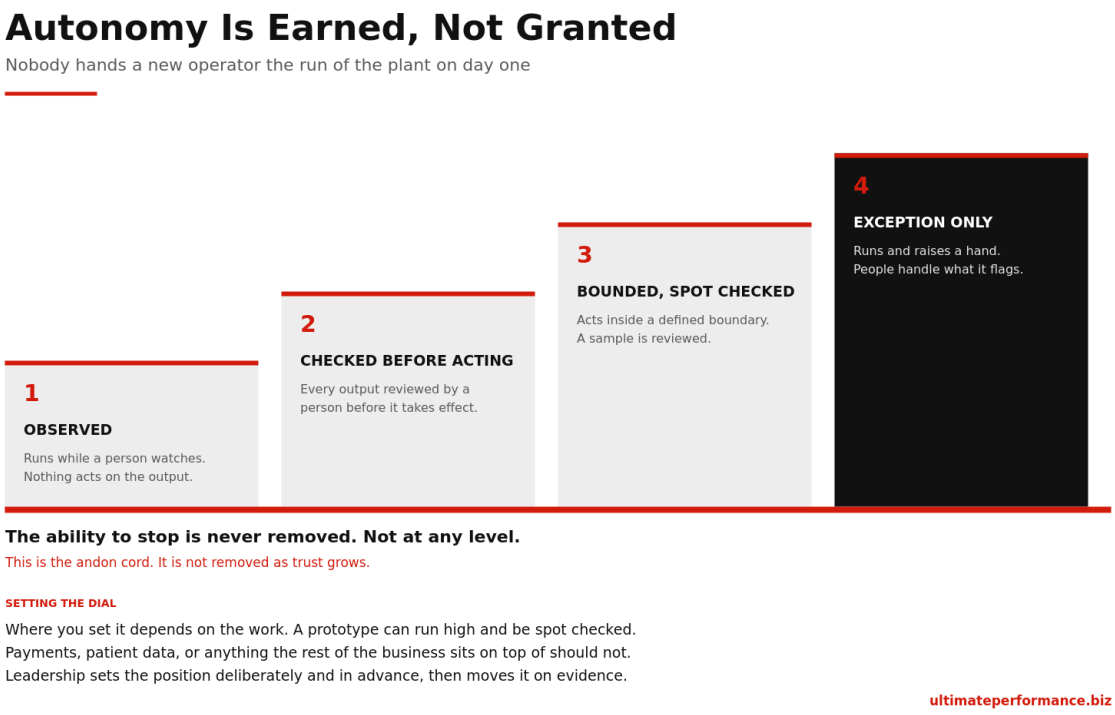
The agent acts inside a defined boundary and a sample of its work is reviewed. This is where most valuable work should sit, and where many organizations never arrive because they jump from level two to level four in a single meeting.

4. Exception Only

The agent runs and raises a hand, and people handle what it flags. Earned, never assumed.

Where you set that dial should depend on the work. A prototype can run with high autonomy and spot checks. Payments, patient data or anything the rest of the business sits on top of should not. What matters is that leadership sets the position deliberately and in advance, revisits it as evidence accumulates, and never lets it drift because a vendor was enthusiastic or a team was busy. Through

all of it, accountability stays where it always was. A person owns the outcome. You cannot delegate that to an agent, and you cannot delegate it to a process.



Where the Comparison Breaks

I would not trust this argument if it did not have limits, so here they are.

- Improvement compounds somewhere different. Teaching a person builds capability inside the person. Teaching an agent builds it inside the standards, instructions and checks that people maintain. Your real student is your own operating system.
- Respect for people does not transfer. I would resist talking about agents as though it does. They are not colleagues. They are capability, and treating them otherwise clouds the judgment of the people who are accountable.
- Variation behaves differently. Human performance drifts slowly and visibly. Model behavior can shift underneath you when a new version ships, which is a supplier change with no notification and no PPAP.
- Failure replicates instantly. One bad judgment made by an agent may repeat thousands of times before lunch. Every argument for making problems visible immediately gets stronger here, not weaker.

Conclusion

You are not choosing a model. You are deciding what kind of working culture your combined workforce is going to have, now that part of that workforce is digital.

The tools will converge. They always do. What will separate companies is whether they already knew how to set a clear standard, make problems visible the moment they occur, go and see the actual work, and grow capability by teaching rather than by inspecting. Those disciplines took a century to learn and they were never really about manufacturing. They were about getting reliable performance out of capable workers inside a system, which is the problem in front of every executive holding an AI budget right now.

If you do not have that for your people, you will not suddenly have it for your agents. And if you do have it, you already own the hardest part of what comes next. The goal has not changed. High performance as a reflex, not a mandate. There is simply more of the workforce to build it into now.

Sources

Steven J. Spear and H. Kent Bowen, "Decoding the DNA of the Toyota Production System," Harvard Business Review, September to October 1999.

Edward Donner, "Are Humans Just AI's Reviewers Now? The New Agentic Development Lifecycle (ADLC)," August 21, 2026. https://www.youtube.com/watch?v=ut-ZwrpPb0U

> "The open human role in the age of agents is not reviewer. It is teacher. A reviewer catches the defect. A teacher makes sure it cannot happen again, and hands out responsibility as it is earned."

About Ultimate Performance

Ultimate Performance helps organizations build the operating systems that turn capability into results — the standards, daily management, problem-solving, and leadership disciplines that make high performance repeatable. Rooted in the Toyota and Danaher operating traditions and the Shingo model, the practice now encompasses the defining challenge of the moment: integrating artificial intelligence and digital labor into how real organizations work, so that AI capability becomes operational performance rather than another stalled initiative.

We work alongside leaders and teams — not from the sidelines — to design and install the operating layer described in this article: shared context across human and digital work, governed autonomy, visual daily management for a hybrid workforce, human development, and continuous problem-solving.

About the Author

Nathan Holt
Founder

Website: www.UltimatePerformance.biz
eMail: Nathan.Holt@UltimatePerformance.biz
Mobile: +1 (770) 329-7307

Nathan Holt is a seasoned expert in company performance transformation, having successfully transformed more than 25 businesses across various industries. He leverages his skills in strategy, culture transformation, continuous improvement, and leadership development to unlock billions in growth and hidden savings for his clients. He started his career at Accenture and Lean Horizons, where he learned how to deliver efficiency and stability for Fortune 500s facing growth and change challenges. He also worked closely with former Toyota and Danaher executives for 15 years, learning from their Lean business systems that are world-renowned for their success.

Nathan's in-house leadership roles at Avery Dennison and Office Depot in the mid-2000s helped restore their fiscal health, boost their workforce resilience, and reverse their declining stock trends. Since 2011, he has focused his expertise on the energy sector, working as a continuous improvement executive at Shell. In 2022, he founded Ultimate Performance, a consultancy that empowers companies to build operational excellence capability and high-performance systems.

Nathan holds an MBA in international business and a bachelor's in industrial engineering. He is also a Shingo organizational excellence examiner, an M&A post-merger integration advisor, and a certified Agentic AI architect.