

BEYOND HUMAN-IN-THE-LOOP

When the Agent Does the Junior Work

How organisations can preserve judgement, professional development and challenge in AI-enabled teams

Giovanni Leonardi

September 2024



When the Agent Does the Junior Work

Transformation · No. 391

Originally written September 2024 | Re-edited and re-rendered for the WhereBeyond Edition,
2025

*"Speed is not the same as direction, and a team that
confuses the two will arrive somewhere fast without
ever choosing where."*

1. The New Teammate

Something has shifted in the last eighteen months that changes the texture of leadership itself. Large language models have moved from being tools we query to being agents we deploy — systems that can be given a goal, break it into steps, call other tools, and return with a finished piece of work while we were still reading our email. The output is not always right. But it is often fast, frequently coherent, and increasingly hard to distinguish from the work of a competent, tireless colleague who never sleeps and never complains about scope.

This essay is not about whether that is good or bad. It is about a narrower and more uncomfortable question that every leader managing a mixed human-AI team is now living with, whether they have named it or not: when the machine produces the plan, drafts the analysis, or executes the task before a human has finished forming an opinion, who is actually leading? And if the answer is "the human, obviously" — what, precisely, are we doing that the AI is not, and are we still doing it once the pressure to move fast takes hold?

I do not think this is a hypothetical for the future. I have watched teams inside technology-forward organisations quietly restructure their working rhythms around agentic tools over the past year — not because a policy told them to, but because the tool was simply faster than waiting for a colleague, and speed has its own gravity.

2. The Comfort of Delegation, and Its Cost

Leaders have always delegated. That is not new, and agentic AI is, on the surface, just a very capable delegate. But traditional delegation carries an implicit contract: the person you delegate to shares your incentives, understands the unstated context of the business, and will flag when something feels wrong even if you did not ask them to. An AI agent, however sophisticated, does none of this by default. It optimises for the objective it was given, not the objective you meant. It has no organisational memory of the client who was burned last year by exactly this kind of recommendation. It does not feel the reputational weight of being wrong in front of a board.

The danger is not that leaders will fail to notice this distinction in the abstract. Most can recite it without difficulty. The danger is behavioural: under time pressure, the fast output becomes the default output, and the discipline of checking it erodes exactly when checking matters most. I have seen this pattern before, in other guises — the over-trusted spreadsheet model, the analyst's confident slide nobody interrogated because the deadline was in an hour. Agentic AI intensifies the pattern because it removes even the modest friction of asking a human colleague to do the work, and friction, it turns out, was doing more governance work than we credited it for.

The risk of agentic AI is rarely a single dramatic failure. It is the slow substitution of speed for judgement, made one reasonable-seeming shortcut at a time.

3. Redefining What Leadership Contributes

If the machine can out-produce the team on raw throughput, then leadership's contribution has to be relocated — not abandoned, relocated — to the things that remain irreducibly human in the loop. I would group these into three categories, each of which looks unglamorous next to the visible drama of an AI agent completing a task in ninety seconds.

- Framing the question. An agent answers what it is asked. It rarely challenges whether the question itself is the right one, or whether it encodes an assumption worth revisiting. A leader's first job is now upstream of the tool: deciding what problem is actually worth solving, and stating it precisely enough that the agent's fast answer is an answer to the right thing.
 - This is harder than it sounds. Vague framing produces confident, fluent, wrong answers at a speed that outpaces the team's instinct to double-check.
- Holding context the model does not have. Every organisation carries tacit knowledge — who was burned by what, which stakeholder reacts badly to which kind of framing, what commitment was made off the record last quarter. None of this lives in a prompt. The leader's role is to inject it deliberately, and to know when its absence has produced an answer that is technically sound and organisationally naive.
- Owning the decision, not just the output. A recommendation produced by an agent is not a decision. Someone has to be willing to be wrong in public for it. That accountability cannot be delegated to a system, however well it performed, and pretending otherwise is where governance quietly fails.

“Speed is not the same as direction, and a team that confuses the two will arrive somewhere fast without ever choosing where.”

4. The Uncomfortable Middle Ground

Most of the current advice on this topic collapses into one of two unhelpful poles. The first insists on "human-in-the-loop" as though the phrase itself were a safeguard — as if inserting a human review step, however rushed or rubber-stamped, discharges the obligation. The second treats AI output with such reflexive suspicion that the tool's genuine advantages in speed and consistency are simply forfeited, and the organisation quietly falls behind competitors who are less anxious.

Neither position survives contact with how teams actually behave under deadline pressure. A review step that exists on paper but is skipped when the calendar is tight is not oversight, it is theatre. And blanket distrust of a capable tool is itself a leadership failure, because it wastes a resource that, used with judgement, makes the team materially better.

The harder, more honest position is that oversight has to be designed for the specific decision, not applied as a uniform ritual. Some outputs deserve five seconds of glance and approval; others deserve the leader personally reconstructing the reasoning from scratch. Knowing which is which — and being willing to slow down on the ones that matter even when the tool has already moved on — is, I would argue, the actual leadership skill this era demands. It is judgement about where judgement is needed, applied unevenly and on purpose.

Signal	What it suggests about oversight needed
Decision is reversible and low-stakes	Light-touch review; speed can dominate
Output touches client relationship or reputation	Leader reconstructs reasoning personally
Model is working from thin or ambiguous context	Inject tacit knowledge before accepting output
Team is under visible deadline pressure	Deliberately slow the highest-stakes step

5. Trust That Is Earned, Not Assumed

There is a temptation, once an agent has performed well several times in succession, to extend it a kind of provisional trust that then quietly becomes permanent. This is a natural human bias — we are built to generalise from a run of good outcomes — and it is precisely the bias that leaders need to name out loud within their teams rather than pretend they are immune to it.

The more sustainable approach treats trust in an AI agent the way a good manager treats trust in a new hire: earned incrementally, calibrated to the type of task, and never assumed to transfer automatically from one domain to an adjacent one. An agent that reliably drafts sound technical summaries has told you nothing about how it will behave when asked to negotiate the framing of a sensitive stakeholder message. Leaders who conflate the two are not being efficient, they are being careless in a way that will eventually cost them.

1. Start with low-stakes, reversible tasks and observe failure modes closely, not just success rates.
2. Expand scope deliberately, one category of decision at a time, rather than in one confident leap.
3. Keep a visible record of where the agent's output was overridden and why — this becomes the team's own calibration data, and often reveals patterns no individual reviewer would catch alone.
4. Revisit the boundary regularly; a tool that is improving month over month deserves a trust boundary that is reviewed, not fixed once and forgotten.

6. What This Means for How Teams Are Led

The deeper change agentic AI forces is not procedural, it is cultural. Teams that get this right are not the ones with the strictest sign-off process; they are the ones where challenging a fast, confident answer — machine-generated or human-generated — is a normal, low-status act rather than a confrontational one. That culture has to be built deliberately, because the natural pull of a fast tool is toward silent acceptance, and silence is cheaper than dissent right up until it is catastrophically expensive.

Leaders also need to be honest with their teams about status anxiety. When an agent produces work faster and more consistently than a mid-level analyst, that analyst notices, and the noticing is not irrational. Pretending the tool is merely an assistant, when it is functionally out-producing junior members of the team, invites quiet resentment and disengagement. The more durable message is that the analyst's value is shifting toward the judgement layer described above — framing, context, and accountability — and that this shift needs to be reflected in how their contribution is evaluated, not just described in a memo nobody believes.

None of this is comfortable, and I do not think it resolves into a tidy formula. What I would resist is the idea that the answer is simply "humans stay in charge" stated as a platitude. Humans stay in charge only where they have deliberately preserved the conditions for judgement to operate — precise framing, injected context, visible accountability, and a culture where fast is not mistaken for right. Where those conditions are allowed to erode under the pressure of a faster tool, the leadership has already been ceded, whatever the org chart still says.