

*October 2026**By Fabrizio Checchi, CEO - Founder at at [Mebius](#) and [Mysteryclient](#)*

Stop Giving Everyone an AI Licence. Start Building Agents.

How to choose which micro-processes to delegate to AI in a market research and mystery shopping context.

As an MSPA partner I end up talking about AI with a lot of colleagues in market research and mystery shopping. The question is almost always the same: "how do I roll AI out in my company?" And I almost always see the same mistake.

Here is one I actually saw happening... A team had AI suggest text corrections on open-ended answers, and the corrected file was handed to the quality control team. As a result, QC now had to read the original, read the correction, and decide what to accept. The AI had added a review step and a new class of potential error. The process got longer, not shorter.

I will come back to how to avoid adding more steps, but first, a lesson in how to think about AI at all, because that is where misunderstanding starts.

AI is a tool, not a superpower

Think of AI as an advanced tool. It can do many things, but to keep this simple, pretend it does exactly one thing, very well. Think of it as a scanner.

A scanner takes a document and files it. Since scanners have existed, almost nobody files originals by hand anymore. That is the point. AI is not there to turn your people into superheroes. It is there to take a piece of work and make it disappear.



October 2026

By Fabrizio Checchi, CEO - Founder at [Mebius](#) and [Mysteryclient](#)

One difference matters from the start though: a scanner is deterministic – there is only one outcome, an AI agent is probabilistic- there may be more than one outcome with varying probabilities. I will come back to this, because it is where most of us trip up!

Two ways to use AI - one matters far more than the other

Before deciding what to automate, separate how.

Copilot (augmentation): the AI sits next to a person and makes them faster. Useful, but the gain is linear and the ceiling is low. The person stays inside the process, the cost grows with headcount, and people usually end up more stressed rather than less, because the intensity goes up while the workload does not go away.

Agent (delegation): the AI takes a micro-process and completes it on its own, passing the output to the next step - human or agent. Here the gain is structural, because a piece of work leaves your people's desks for good.

Put differently: **AI does not increase the speed at which a person does a job. It removes part of that job from the person.** That is a change in nature, not in speed.

How to decide what to delegate

Finding the micro-processes to delegate is our job, not the model's. This is the method I propose:

1. Unbundle each process into micro-processes.
2. Measure frequency - how often each one repeats in your company.
3. Score AI fitness - how well an AI agent handles it today (high means the model is playing at home).
4. Score error risk - what happens if the agent gets it wrong.



October 2026

By Fabrizio Checchi, CEO - Founder at at [Mebius](#) and [Mysteryclient](#)

Frequency multiplied by AI fitness gives you the priority order. Risk is not a fourth term in that multiplication: it is a filter you apply afterwards, I will show why in a moment.

Worked example: quality control

If you do not unbundle quality control, an AI agent is no help to you. Run it through the grid and QC turns out to be five different jobs - one continuous act for a human, five unrelated problems for a machine.

Micro-process	Frequency	AI fitness	Error risk	Verdict
Verbatim correction and client-specific terminology (~20 open texts per questionnaire)	High	High	Low	Delegate first
Consistency between the closed score and the verbatim that justifies it	High	Medium-high	Low	Delegate next
Evidence check - is the photo of what it should be, is the receipt legible and matching	High	Medium-high	Medium	Delegate, with an exception queue
Compliance check - timestamp, geolocation, dwell time, shopper eligibility	High	n/a - deterministic	Low	Do not use an agent. Write a rule.
Overall validation of the visit, edge cases, client disputes	Low	Low	High	Keep it human

Two things fall out of that table immediately. The first is where to start. The second is less obvious and worth more: one of those micro-processes does not need an agent at all. Timestamps, geofences, dwell time and shopper eligibility are rules. If an if-statement gives you a deterministic answer, do not pay for inference and do not accept a probabilistic one instead. Unbundling tells you where AI belongs, and it tells you just as clearly where plain code is the better tool. Half the AI

*October 2026**By Fabrizio Checchi, CEO - Founder at [Mebius](#) and [Mysteryclient](#)*

budget I see wasted on this industry is spent on problems that were never probabilistic to begin with.

The first agent

Verbatim correction: highest frequency (around 20 open texts per questionnaire), best fit, lowest risk. The micro-process:

1. Extract every open text from the questionnaire.
2. Load the client's correction system - every client has its own terminology, starting with what they call their sales assistants.
3. Pass the AI agent the text and the correction system.
4. Return the corrected text into the pipeline - not onto a reviewer's desk.

Step four is the one the team I opened with got wrong. An agent that hands its output to a human for approval is a copilot in an agent's costume: you have added a step, not removed one.

Does this let me skip quality control? No, but it removes phone-typing transcription errors entirely, and it saves roughly two minutes of correction time per questionnaire.

At 5,000 questionnaires a month that is about 167 hours, or roughly €3,000 at fully loaded internal cost. Then we move to the next item by priority...

Risk does not multiply, it filters

Telling an LLM (Large Language Model) "check that this visit was carried out properly" is a request the technology is badly suited to, it is too broad, too open to variability. Breaking QC into micro-checks - "confirm that this photo shows a pizza" - gives you an answer you can actually work with.

*October 2026**By Fabrizio Checchi, CEO - Founder at at [Mebius](#) and [Mysteryclient](#)*

Quality control will probably always sit under human supervision, even if one day it becomes technically easy. In mystery shopping the quality of the data is the product. There the agent supports human judgement; it does not replace it. That is what the risk column is for: it does not reorder your priorities; it tells you which micro-processes need an exception queue and which ones you simply do not delegate.

Delegating a process does not mean forgetting about it

Here is where the scanner-versus-agent difference comes back. A scanner does the same thing forever. An agent does not: it drifts, it is variable, when the client changes the correction system, edge cases arrive. An agent in production has to be measured continuously for accuracy, and you have to design in advance what happens to the share of cases that must inevitably reach a human. The hard part is not making the agent work the first time - It is keeping it reliable the hundredth time.

Our second example is the one that makes money

Everything above is cost compression: quality control, operations, back office. It is the half that is easy to measure, and it is the half most people stop at. For a research company the more interesting move is the opposite one - using the same method to expand what you can sell.

Same grid, different process. Take unsolicited customer reviews, which most of us still treat as a separate product line:

1. Ingest the reviews outlet by outlet.
2. Code them against the client's code frame - structurally the same pattern as the client-specific correction system above.

*October 2026**By Fabrizio Checchi, CEO - Founder at at [Mebius](#) and [Mysteryclient](#)*

3. Use the same taxonomy (classification system) as the mystery shopping questionnaire. This is the step that matters and the one almost everybody skips.
4. Output, per outlet: what customers say, next to what a trained observer recorded.

Two data sets that have always lived in separate decks now sit on one axis. And the deliverable stops being "here is what your customers think". It becomes the gap: the outlets where perception and observed behaviour disagree, and in which direction. An outlet that scores well on the MS checklist and badly in the Customer reviews has a problem the checklist does not measure. An outlet with the opposite profile has a reputation problem, not an execution one. Those two need opposite interventions, and until both sources sit on the same taxonomy you cannot tell them apart - you can only argue about which data set to believe. Believe both, it in the Gap that we find the value,,,

That gap is worth more than either data set on its own, and it is something we can put in front of our own client and charge for.

There is a third step, and I would argue it is the real end of the pipeline: the gap, outlet by outlet, is the training brief for that outlet. Measure, explain, fix - not three projects, one loop. We run that loop at Mebius, with review analysis alongside mystery shopping at one end and mobile retail training at the other. I am not going to turn this article into an advert for my company; if the mechanics are useful to you, ask me. It is my offer to connect, learn and share!

The most common mistake

Which brings me back to where I started. The most common mistake is giving every employee an AI licence so they can answer their emails better. That is buying diffuse augmentation, reducing noise but not increasing efficiency: a cost that grows with headcount but changes no process.



October 2026

By Fabrizio Checchi, CEO - Founder at at [Mebius](#) and [Mysteryclient](#)

Twelve months later you have a larger software bill, the same process map, and no way to prove what any of it did for you, with more stressed colleagues believing they are delivering more for less!

The agentic era has arrived. It does not need licences sprayed across the org chart. It needs a model, a person who unbundles the processes and works out which micro-processes to delegate and how, and AI agents that close them one at a time.

This is the only version in which the cost of AI goes down while the company grows, instead of up.

If you want the scoring grid I use - frequency, AI fitness and risk, with the quality control example filled in - send me a message and I will share it.
