

BEYOND ADOPTION DASHBOARDS

Copilots Everywhere, Business Cases Nowhere

The largest enterprise software purchases of the decade are justified by the weakest benefit cases

Giovanni Leonardi

May 2024



Copilots Everywhere, Business Cases Nowhere

Portfolio · No. 102

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

"The most expensive autocomplete in history might also be the most valuable productivity tool in a generation."

The Procurement Wave Without a Value Case

Somewhere in your organisation right now, a procurement team is finalising an enterprise-wide copilot licence. The business case — if it exists at all — cites a vendor-commissioned productivity study, an adoption target expressed as a percentage of seats activated, and a benefit uplift figure borrowed from a press release. The number of organisations committing seven- and eight-figure annual spends on generative AI tooling in the first half of 2024 is remarkable. What is more remarkable is how few of them have asked the question that actually matters: which specific workflows, performed by which specific roles, will produce value that exceeds the cost of the licence and the overhead of verifying the tool is doing what it claims?

This is not an argument against copilots. The underlying technology is genuinely capable, and in certain workflows the productivity gains are real and measurable. The argument is against the way they are being bought: blanket deployment justified by blanket assumptions, with adoption dashboards standing in for value measurement and per-seat pricing accepted as though every knowledge worker benefits equally.

They do not.

The Adoption Fallacy

The metric most commonly reported to boards and steering groups is adoption — the percentage of purchased licences that have been activated, or the number of employees who have used the tool at least once in the past thirty days. This tells you almost nothing about value. A developer who uses an AI code assistant to generate boilerplate that would have taken four minutes to type is "adopted." A legal associate who uses a document summariser to produce a summary that then requires thirty minutes of manual correction is also "adopted." The metric cannot distinguish between the two.

What makes this worse is that it is self-reinforcing. Vendors design onboarding programmes to maximise activation rates because their renewal conversations depend on

them. Enterprises accept activation as proof of return because measuring actual workflow-level impact is harder. The result is a measurement system optimised to confirm the purchase decision rather than to test it.

The Workflow-Level Value Test

The question copilot buyers should be asking — and largely are not — is narrower, harder, and more useful than "does AI improve productivity?"

It is this: for each role category in the organisation, which specific workflows are candidates for copilot-assisted improvement, and what does "improvement" look like in terms that survive scrutiny?

The components of a proper workflow-level test are not exotic. They are the same components any serious benefit case would demand for a technology investment of this scale:

- Workflow identification — not "knowledge work" as a category, but named processes: drafting client correspondence, reviewing contract clauses, triaging support tickets, writing unit tests. The level of specificity matters because the gains are not uniform.
- Role segmentation — a copilot that saves fifteen minutes per day for a senior analyst may save nothing for an operations coordinator whose work is procedural and already templated. Per-seat pricing obscures this entirely.
- Measurable output criteria — time saved is the most common claim and the least reliable. The real questions are whether the output quality is maintained, whether the verification overhead erodes the time saving, and whether the workflow change persists beyond the novelty period.
- Counterfactual discipline — the productivity study that compares copilot-assisted workers against unassisted workers in a controlled setting is rare. The study that compares them against workers using the tools and shortcuts they had already developed is almost non-existent.

None of this is technically difficult. It is organisationally inconvenient, because it risks producing an answer that challenges the purchase decision already made.

The Renewal Reckoning

The enterprise copilot market in 2024 runs on annual contracts, minimum seat commitments, and the assumption that adoption will vindicate the investment before the renewal conversation arrives. For organisations that signed in late 2023 or early 2024, that conversation is twelve to eighteen months away.

When it arrives, the CFO's question will not be "how many people activated their licence?" It will be "what did we get for this?" And the organisations that cannot answer at the workflow level — that cannot point to specific processes where the tool demonstrably improved output, reduced cycle time, or eliminated rework — will find themselves in the weakest possible negotiating position: committed to a renewal they cannot justify and

unable to reduce scope because they never established which scope was producing value in the first place.

The time to build the workflow-level value case is not at renewal. It is now — while the deployment is live, the usage data is accumulating, and the organisation still has leverage to adjust its commitment.

The Point

This is not a case against enterprise copilots. It is a case against enterprise copilot procurement that substitutes vendor metrics for organisational measurement. The technology may well deliver transformational value — but "may well" is not a business case, and the organisations that will be best positioned in eighteen months are those building the evidence now, workflow by workflow, role by role, with the discipline they would apply to any investment of comparable scale.

The most expensive autocomplete in history might also be the most valuable productivity tool in a generation. We will not know which until someone bothers to measure it properly.