

Measure agent outcomes, not just agent accuracy

Value. A technically accurate agent can create no operational value at all: the recommendation was right, nobody acted on it, or acting on it moved a number nobody was watching. Accuracy is the vendor’s metric. These are yours.

Five measures for the scorecard. Baseline them before the agent goes live.

1	Time-to-decision. From signal to committed action: the measure that shows whether the agent removed work or added a review step.
2	Adoption. How often operators act on the recommendation. Low adoption is a judgment gap, not a training gap.
3	Override quality. Not the override rate: whether overrides came with reasons, and whether the reasons were right.
4	Service, inventory, margin. The operating outcomes the decision exists to move, attributed to the workflow the agent runs in.
5	Realised value. What the business kept, and the honest line between the agent’s contribution and everything else that changed.

A scorecard answers one question:
did the agent change business performance?