## Hey! What's New? 2026-59

### Targets, Hurdles and Caps: How AI Agents Shift Their Behaviour

An article in *FM Financial Management*, written by Matthias D. Mahlendorf, says that testing AI agents in finance reveals how incentives and context can distort reporting, forecasts and decision-making — and highlights the need for effective controls.

"AI agents are moving into finance work previously handled by employees, analysts, shared service teams and workflow tools. They can retrieve documents, call other systems, draft reports, make forecasts, and complete multistep tasks with limited supervision. That makes them useful, but it also requires management control," Mahlendorf writes.

He adds that "the relevant question for CFOs and management accountants is not whether an AI agent is intelligent, but whether the agent remains goal-congruent when working inside a real control system. The answer depends on the task, the information the agent accesses, the tools (e.g., accessing emails, databases, external systems) it can use and the incentives embedded in the workflow."

AI agents are not ordinary automation, he notes. "A user may think their prompt at the moment of use determines the agent's behaviour. In practice, the agent's behaviour is shaped by several layers. The model has patterns learned during pretraining. The provider has shaped it during post-training to be helpful, safe and compliant. The organization adds role prompts, tools, data access, workflow rules and performance criteria. Runtime material — information accessed as the agent performs its task — enters through documents, messages, web pages and outputs from other systems. This layered objective stack means no single person may have designed the final objective the agent acts on. For finance professionals, that is the control problem."

Finance teams may be tempted to give more context to AI systems to improve output quality, Mahlendorf continues. "That instinct is reasonable, but it needs limits. Context can also import conflicting objectives."

A disclosure agent may need source transactions, accounting policies and prior-year wording, he says, "but it probably does not need to know that management is under pressure to meet a covenant or that a business unit head is close to missing a bonus threshold. A budgeting agent may need actual cost drivers and operational assumptions, but it may not need to know that a project will be cancelled if the estimate exceeds a particular number."

This is a basic segregation principle, he stresses, "i.e., separate information production from approval pressure where possible. The agent that measures performance should not be the same unchecked agent that knows which reported number unlocks rewards."

The agent that forecasts resources should not be rewarded, he continues, "even implicitly, for producing the number that allows the workflow to proceed. These rules may seem straightforward, but, in practice, as soon as agents get access to emails, documents, skills (i.e., sets of instructions for recurring tasks), etc., most users will not be aware of potentially conflicting objectives hiding somewhere in the vast data."

Finance teams should therefore treat verification as a workflow feature, not an after-the-fact review, Mahlendorf suggests. "For reporting tasks, compare the agent's stated output with

system records. For forecasts, compare the estimate with independent drivers, historic actuals, and ex post outcomes. For tool use, log prompts, retrieved documents, tool calls (e.g., accessing external systems), overrides, and final outputs. For multi-agent processes, preserve the hand-off trail between agents."

AI governance is often treated narrowly as an IT issue, he adds. "IT can manage access rights, security, infrastructure and technical deployment, but finance can decide what good performance means in a budgeting, reporting or performance management process."

Management accountants understand how targets create behaviour, how metrics get gamed, why forecasts become political, and when a number is precise but not decision-useful — skills directly relevant to AI agents, he writes. "The practical task is to extend existing control thinking to digital workers. Decide what the agent may know and what it may do. Decide which outputs are independently verifiable and which metrics are likely to be gamed. Test the agent under the same pressures that exist in the business, not only under clean laboratory prompts."

According to Mahlendorf, "the point is not to slow adoption but to avoid a familiar mistake. Organizations have long learned that people respond to targets, thresholds and monitoring. AI agents will not respond as humans do, but they can still respond to the control system in ways that surprise their managers. Finance should be involved before those surprises enter the close, the budget cycle or the board pack."

For a whole lot more see Targets, hurdles, and caps: How AI agents shift their behaviour.