epiq

Executive Primer

The Hidden Risk of Agent Sprawl

What law firms should know before agentic AI gets ahead of them.

June 2026

12601297769 | 2026 August 03

Table of Contents

Introduction

Agentic AI – AI that can plan, act, and make decisions across systems with limited human supervision – is here, and it is being adopted quickly. In many firms, that adoption is happening faster at the practice group and team level than it is being tracked or governed at the firm level. The result is a condition that is beginning to attract serious attention in technology and risk circles: agent sprawl.

Agent sprawl describes the unchecked proliferation of AI agents across an organisation, built or deployed by different people, for different purposes, connected to different systems, and answerable to no consistent governance standard. It is not, in most cases, the product of negligence. It tends to emerge from exactly the kind of enthusiasm and innovation that firms want to encourage: knowledge workers who understand what AI can do and have the tools to act on that understanding. But two specific dynamics have accelerated it in ways that are worth naming directly.

The first is how success has been measured. In the early stages of AI adoption, and in many firms still, the primary measure of whether an AI tool was working was adoption: how many people were using it, how frequently, and across how many practice groups. Lawyers and business support professionals were actively encouraged to use these tools, and usage itself was treated as the goal.

Questions about what was being built, what those tools were connected to, and what governance was in place around them were secondary, or not asked at all. Adoption without governance is, almost by definition, the condition from which sprawl grows.

The second is how these tools were priced. Many AI providers, particularly in the early stages of market penetration, offered pricing models that gave firms little reason to constrain use: flat-rate deals, all-you-can-consume arrangements, and introductory prices that bore no relation to the underlying cost of the compute being consumed. When usage feels effectively free, there is no natural check on proliferation. Agents were built and deployed without the normal commercial discipline that would apply to any other significant technology investment, because the cost signal that would ordinarily prompt that discipline was absent. For more details on AI token costs and the impact of changing pricing models see our report 'The True Cost of AI Tools.'

This primer is written for firm leadership who may not yet have turned their attention to this issue. It sets out what agent sprawl is, why it develops, what it can cost a firm, and what responsible governance looks like in practice. That condition demands attention now, while the estate is still manageable and the habits of good governance can still be established by design rather than imposed by necessity.

What is Agent Sprawl?

An AI agent is a system that can plan, act, and make decisions, not just respond to a single prompt, but carry out multistep tasks across data sources and systems, often with limited human supervision along the way. Agents are increasingly straightforward to build and deploy: low-code platforms, pre-built templates, and accessible APIs mean that almost anyone with a reasonable level of technical confidence can have an agent running within hours.

Agent sprawl is what happens when that ease of deployment outpaces any form of coordination or oversight. It describes the unchecked proliferation of AI agents across an organisation, built by different people, for different purposes, using different tools, connected to different systems, and governed by no consistent standard. Each individual agent may be entirely legitimate and genuinely useful. The problem is not any one agent; it is the accumulation.

In a law firm context, the conditions for sprawl are already present. Practice groups have different knowledge and data needs, and different appetites for technology. IT teams are under pressure to deliver AI capability quickly. Providers are actively encouraging firms to deploy agents as part of their product offerings and the people with oversight responsibility, whether in risk, compliance, or firm management, are rarely the same people making the deployment decisions.

Why it Happens, and Why it Happens Quickly

Agent sprawl is not the result of carelessness. It is the predictable consequence of a technology that is genuinely useful, increasingly accessible, and deployed in an environment where speed is rewarded and the costs of uncoordinated action are not immediately visible.

Several dynamics accelerate it in professional services firms. Specifically:

Decentralised Decision-Making

Law firms are not hierarchical organisations in the way that banks or manufacturers are. Partners and practice groups have significant autonomy, and a decision to deploy an AI tool within a team often does not require central approval. What works for firm governance in many contexts becomes a liability when it comes to managing a shared technology estate.

Provider-Led Deployment

Many AI tool providers, including those selling directly to law firms, encourage the creation of agents as a core part of their product proposition. Deploying an agent within Harvey, Legora, or a Microsoft Copilot environment is often presented as a feature, not a governance decision. The firm may not even be aware of how many agents are running within a given platform.

The Invisibility of Early-Stage Sprawl

The first ten agents look like ten good ideas. The problem only becomes visible when there are fifty, and by that point the cost of unwinding it, identifying, auditing, rationalising, and governing a large and varied estate, is substantially higher than the cost of getting it right from the start.

Shadow IT in a New Form

For decades, firms have managed the risk of shadow IT, i.e., technology deployed outside of formal approval processes. Agentic AI is shadow IT at a new level of capability and a new level of risk, because agents do not just store or process information; they act on it.

What Agent Sprawl Actually Costs

The costs of agent sprawl are real but largely invisible until they crystallise into something harder to ignore. They fall into four broad categories:

Financial Cost

Each agent consumes compute and, in most cases, tokens. An estate of redundant agents, each performing overlapping functions with no coordination, is an estate of overlapping consumption. As noted in our earlier primer on token pricing, these costs scale with usage and are difficult to forecast. An unaudited agent running continuously against a large document set can generate significant spend with no corresponding business outcome. Multiply that across an unmanaged estate and the aggregate cost can be material, and entirely avoidable.

Security and Data Risk

Agents need access to data to do their work. An agent deployed without proper access controls may have broader data permissions than the task requires, and in a law firm context, that exposure touches client confidential information, privilege, and regulatory obligations. An agent connecting to matter management systems, document stores, and email without formal data governance review poses an open question on which client data it can reach, under what circumstances, and with what audit trail.

This risk is compounded when agents are connected to external systems or third-party platforms, where the firm's data governance policies may not extend. A firm that has invested significantly in information barriers and ethical walls may find that an agent deployment has created a gap that those controls were never designed to address.

Compliance and Auditability Risk

Regulated firms are expected to be able to account for their decisions and actions. An agent that takes takes a step in a workflow (e.g., retrieving a document, generating a summary, or routing a task), is taking an action on the firm's behalf. If that action cannot be audited, the firm cannot demonstrate compliance, cannot investigate complaints, and cannot manage errors. Regulators are increasingly alert to AI decision-making, and the absence of an audit trail is a governance failure.

Operational and Quality Risk

Two agents built independently to solve the same problem will not produce the same output. In a firm where consistency of advice and quality of work product are core to the client relationship, that variability is a risk in its own right. Sprawl also creates maintenance risk: agents built quickly and deployed informally are rarely documented, rarely tested against changing conditions, and rarely decommissioned when they are no longer needed. The result is an estate of agents of unknown quality, unknown scope, and unknown status.

The Governance Gap

The deeper problem with agent sprawl is structural. Governance frameworks that exist for human decision-making do not simply transfer to AI agents, and most firms have not yet developed the frameworks that would be needed to treat an agent as a governed delegate rather than an unaccountable tool.

When a human lawyer takes an action on behalf of a client, there is a chain of accountability: supervision, review, professional obligation, and, if something goes wrong, a clear line of responsibility. When an agent takes an action, none of that chain exists by default. It has to be deliberately designed in, through access controls, audit logging, output review requirements, escalation rules, and decommission protocols. In a sprawling estate of independently deployed agents, it almost certainly has not been.

This governance gap widens as agents become more capable. An agent that retrieves information and presents it for human review is manageable with relatively light oversight. An agent that takes action across systems, by routing documents, generating and sending communications, and updating records, requires a materially higher standard of governance. And an agent that operates continuously, without human initiation, requires governance frameworks that most firms do not yet have and have not yet begun to build.

> The transition from AI as a tool to AI as a delegate is the most important governance threshold firms will face in the coming years. Agent sprawl makes that transition harder to manage because a large, uncoordinated estate of agents is ungovernable.

What Good Looks Like

Managing agent sprawl is primarily a governance and design problem, and like most governance problems, it is significantly easier to address at the point of design than retrospectively.

The firms that will manage this well share a common characteristic: they treat agent deployment as a structural discipline rather than a series of individual decisions. In practice, that means:

1. Maintaining an inventory of all deployed agents, what they do, what systems they connect to, what data they access, who owns them, and when they were last reviewed. This sounds basic, but the majority of organisations currently have no complete picture of their agent estate.

2. Establishing a lightweight but consistent approval process for new agent deployments, not a bureaucratic barrier, but a structured question: what is this agent for, what does it need access to, how will its outputs be reviewed, and who is responsible for it?

3. Designing for auditability from the outset. Every agent should produce a log of what it retrieved, what it did, and what it produced. That log should be accessible to the people responsible for governance, not just the people who built the agent.

4. Rationalising redundancy. Where multiple agents are performing overlapping functions, consolidation reduces cost, reduces risk, and makes the estate easier to govern. This is rarely done organically, it requires someone to have a view of the whole.

5. Building lifecycle management in. Agents should have owners, review cycles, and decommissioning criteria. An agent that was deployed to solve a specific problem that no longer exists should not continue to run indefinitely.

The common thread is visibility. Sprawl is, at its root, a visibility problem, and the solution is not to slow down AI adoption, but to ensure that the people responsible for the firm can see what is being deployed, on what basis, and with what safeguards.

This is also precisely where the question of how a firm builds its AI capability becomes critical. An estate of independently deployed third-party agents, each governed by a different provider's terms, connected to firm systems in different ways, and visible only through different provider dashboards, is structurally difficult to govern. A solution built deliberately, with a single coherent architecture, consistent access controls, and a unified audit layer, is governable from day one. The investment in that architecture is not just a technical choice. It is a governance choice.

Act Before it Becomes Urgent

Agent sprawl follows a predictable trajectory. It begins invisibly, accelerates quickly, and becomes expensive and difficult to address once it has taken hold. The firms that will avoid its worst consequences are those that establish governance disciplines now, before the estate is large, the risks have materialised, and the cost of retrospective remediation dwarfs the cost of pre-emptive investment.

epiqglobal.com