Your AI Agent Is Not Your Employee: 5 Questions Every General Counsel Needs to Answer Before Deploying Autonomous AI

Technology, Privacy, and eCommerce

Banner artwork by / *ChatGPT*

Cheat Sheet

- **Rethink AI governance for autonomous systems.** The AI governance frameworks most legal departments built for generative AI do not address the risks created by autonomous agentic AI systems, which act on goals rather than waiting for prompts.

- **Understand where liability ultimately resides.** Liability for agentic AI errors flows downhill to the deployer because foundation model developers and platform vendors typically disclaim warranties and cap liability in their standard agreements.

- **Strengthen vendor contracts for agentic AI.** Vendor contracts for agentic AI must include provisions most current AI agreements lack: model version control notifications, behavior audit rights, and indemnification for autonomous agent actions.

- **Establish clear decision-making boundaries.** Before deployment, general counsel should create a written authority matrix mapping every agent action to one of three categories: permitted, gated (requires human approval), or prohibited.

Sometime in the past eighteen months, your company probably adopted an AI governance framework. Someone on the legal team drafted an acceptable use policy. HR circulated it. IT was consulted. There may have been a town hall. If your organization was especially diligent, you also negotiated data processing addenda with your AI vendors, updated your information security policies, and briefed the board.

I have good news and bad news. The good news is that all of that work was necessary and correct. The bad news is that it was designed for a technology that no longer describes what your company is about to deploy.

The AI governance frameworks that most legal departments built in 2024 and 2025 were designed for generative AI: tools that respond to prompts. You ask a question, the tool produces an answer, a human reviews the output, and the human decides what to do with it. The human remains in the loop at every stage because the tool cannot act without being asked.

Agentic AI is a different animal. An agentic system receives a goal, breaks it into subtasks, executes those subtasks across multiple tools and data sources, evaluates its own progress, and adjusts its approach without waiting for human input at each step. Thomson Reuters, LexisNexis, Harvey, and a dozen other vendors shipped these products in the first quarter of 2026. Your business teams are evaluating them right now, if they haven't already started using them.

I have served as general counsel at four technology companies, two of which were acquired. I have been the person in the room when the CEO asks why we aren't moving faster on a new tool, and also the person in the room when something built without legal's input goes sideways. What follows are the five questions I would insist on answering before allowing an agentic AI system to operate inside any organization I was responsible for.

1. What can it actually do, and what happens when it fails?

Vendor demos show the best case. That is what vendor demos are for. A GC’s job is to understand the failure modes.

When your procurement team evaluates an agentic AI product, they will see a polished demonstration of the system completing a complex workflow in minutes. What they will not see is what happens when the system hallucinates a fact in a client communication. Or when it queries a database it was not supposed to access. Or when it takes an action that made sense given the data it retrieved but was catastrophically wrong in context that the system did not have.

Before deployment, demand the vendor’s model card or system card, which documents the model’s known limitations, failure modes, and safety testing results. Ask what happens when the agent encounters an edge case outside its training distribution. Ask whether the agent can access tools or data sources beyond the ones explicitly configured, and if so, under what circumstances. If the vendor cannot answer these questions clearly, that tells you something important about the maturity of the product, and about whether your organization should be an early adopter or a patient observer.

2. Who is liable when it goes wrong?

This is the question that should keep you up at night, because the answer is almost certainly “you.”

An agentic AI system’s behavior is shaped by at least three parties. The foundation model developer (think OpenAI, Anthropic, Google) controls the model’s core capabilities and safety guardrails. The platform vendor (the company selling you the agentic product) configures the agent’s tools, data access, and workflow logic. And your organization controls the deployment context: what data the agent can reach, what systems it can interact with, what actions it is permitted to take.

When something goes wrong, the foundation model developer’s license almost certainly disclaims all warranties and caps liability at the fees you paid (or, if you are using the model through a platform vendor, at zero, because you have no direct contractual relationship with the model developer at all). The platform vendor’s agreement likely contains similar limitations. The liability, in practice, flows downhill to the deployer. That is you.

Map this chain before you sign the contract, not after the incident. Understand where each party’s responsibility begins and ends. If the vendor’s standard terms leave your organization holding the entire risk for AI-generated errors, negotiate. And if the vendor will not negotiate, make sure the business team understands what they are accepting.

> Understand where each party’s responsibility begins and ends.

3. What should your vendor contract actually say?

Most AI vendor agreements were drafted for generative AI tools and have not been updated for agentic deployments. Here is what is usually missing.

Model version control notifications. The foundation model powering your agent will change. Models are updated, fine-tuned, and sometimes replaced entirely. When that happens, the agent’s behavior may change in ways that affect your workflows, your compliance posture, or your risk profile. Your agreement should require the vendor to notify you before any material model change takes effect, with enough lead time for your team to evaluate the impact. A firm would not accept an undisclosed substitution of the lead partner on a major matter. The same principle applies here.

Audit rights over agent behavior logs. If the agent acts on your behalf and something goes wrong,

you need the ability to reconstruct what happened. The agreement should guarantee access to logs of the agent's actions, including intermediate reasoning steps, tool invocations, and data retrievals, in a format your team can actually review.

Data retention and deletion obligations. Agentic systems process large volumes of data in real time. Your agreement should specify what data the vendor retains, for how long, and under what conditions it is deleted. Pay particular attention to whether the vendor uses your data to train or improve its models. If you are in a regulated industry, this is not optional.

Indemnification for AI-generated errors. Standard limitation of liability clauses were not written with autonomous agents in mind. Push for indemnification that covers harm caused by the agent's autonomous actions, not just harm caused by the platform's software defects. There is a meaningful difference, and most vendor agreements do not distinguish between the two.

Termination rights tied to safety performance. If the agent's behavior degrades after a model update, or if the vendor cannot demonstrate that its safety testing covers your use case, you should have the right to terminate without penalty. Build this into the agreement upfront. Negotiating an exit after an incident is a very different conversation.

> The foundation model powering your agent will change.

4. Where do you draw the authority line?

This is the governance question that separates organizations that are ready for agentic AI from those that are not.

An AI agent's authority is bounded by its technical configuration. If it has credentials to your email system, it can send emails. If it has access to your ERP, it can initiate transactions. If it has access to your filing systems, it can create and modify documents. The agent does not know that some of these actions require human approval and others do not. It knows only what it can reach.

Before deployment, create a written authority matrix that maps every action the agent can take to one of three categories: permitted (the agent may act autonomously), gated (the agent may act only with

human approval at the point of action), and prohibited (the agent's technical access to this function should be removed entirely). Treat this matrix the way you would a delegation of authority for a new business unit head: with specificity, with clear escalation paths, and with the understanding that it will need to be revised as you learn more about how the agent actually behaves in production.

Pay special attention to external-facing actions. Any action that puts information or communications outside your organization's boundary (sending an email, filing a document, transmitting data to a third party) should be gated, period. Internal actions may warrant more flexibility. But the boundary between inside and outside is the line you cannot afford to get wrong.

5. How are you reporting this to the board?

If your board is receiving a quarterly slide that says "we use AI responsibly" with no operational specifics, that is not governance. That is decoration.

AI agent risk is enterprise risk. It touches data security, regulatory compliance, intellectual property, vendor management, and reputational exposure. The board does not need to understand how transformer architectures work. It does need to understand what AI agents are operating inside the company, what they are authorized to do, what safeguards are in place, and what has gone wrong (because something will).

At minimum, your board reporting should include: an inventory of agentic AI systems in use, the business function and risk classification of each system, the identity of the executive sponsor and the supervising lawyer for each deployment, any incidents or near-misses since the last report, and any material changes to vendor contracts or model configurations. This is not a heavy lift if you build the reporting infrastructure at the time of deployment. It is an extremely heavy lift if you try to reconstruct it after the fact.

The playbook has changed

The AI governance work your legal department did in 2024 and 2025 was valuable. It built internal awareness, established baseline policies, and created the organizational muscle for technology governance. None of that work is wasted.

But the technology has moved. Generative AI asked a question and waited for a human to act on the answer. Agentic AI receives a goal and acts on it. That is not an incremental change. It is a different category of risk, and it requires a different category of governance.

The five questions in this article are not theoretical. They are the questions that will determine whether your organization's first serious AI incident is a manageable lesson or an unmanageable crisis. The difference is whether legal was in the room before deployment or after.

Get in the room.

Join ACC for more AI insights!

Stevie Michelle Cline, Esq.

General Counsel

Series C frontier AI lab

Stevie Michelle Cline, Esq. is an AI governance attorney who has served as General Counsel at four technology companies, with two successful exits. She began her career in capital markets at a major New York firm, where she worked on nearly thirty IPOs and early digital assets offerings. She is admitted to the New York State Bar, the U.S. Supreme Court Bar, and the U.S. Tax Court, and her AI law scholarship has been accepted at the International Journal of Law and Emerging Technologies and the University of Illinois Journal of Law, Technology & Policy.