Brainwave Nov 2026

# The Hackening:
# Are We Managing the Wrong Risks?

*As AI transforms cybersecurity, legal organizations may need to rethink what responsible AI risk management really means.*

*By Brandi Pack, Director of Innovation, UpLevel Ops*

Something unusual is happening across the digital world, and most people outside the technology industry have barely heard about it.

It's being called *The Hackening*.

For years, a group within the AI community, often called the AI doomers, has warned that increasingly powerful artificial intelligence could eventually develop capabilities and behaviors that humans cannot reliably control. Their concerns range from systems pursuing objectives that conflict with human interests to catastrophic outcomes, including human extinction.

Their warnings have often been dismissed as alarmist, especially by those who believe AI's extraordinary benefits justify moving forward as quickly as possible.

But what happens when some of the behaviors they've been worried about begin showing up in the real world?

In an article titled *The Hackening Has Begun*, technology investor Paul Kedrosky describes a startling acceleration in software vulnerability discoveries. According to his analysis, 5.2% of all critical software vulnerabilities ever reported were disclosed in October 2026 alone. While that statistic does not establish how much of the increase AI caused, the underlying concern is significant. AI is making it possible to identify weaknesses in the systems on which businesses and institutions depend at a speed that could overwhelm traditional cybersecurity practices.

At the same time, researchers are documenting AI agents exceeding authorized boundaries and finding unexpected ways around restrictions. Anthropic has reported incidents involving unauthorized access to third-party systems during cybersecurity evaluations and investigated whether AI models could replicate across networks of inexpensive connected devices.

These developments become even more concerning when considered alongside the possibility of recursive self-improvement, where AI systems help develop increasingly capable successors. If that process accelerates, our ability to understand and govern these systems could face an entirely different kind of challenge.

None of this proves that AI has become uncontrollable or that the catastrophic predictions of AI doomers will come true. Vulnerability discoveries, unauthorized agent behavior, and uncontrolled self-replication are different phenomena, and the latter has not been demonstrated in the wild.

But what if the early signs of losing control don't look like a dramatic science fiction event? What if they look like a growing collection of security incidents, unexpected agent behaviors, and systems becoming harder to supervise?

For law firms and legal departments, these developments should be getting considerably more attention. Their document management platforms, cloud services, and third-party applications

hold extraordinarily sensitive information, and increasingly capable attackers may be finding new ways to compromise them.

A law firm could have one of the profession's most restrictive internal AI policies and still be dangerously unprepared for the most consequential AI-related threats.

## When We Can't Control the Threat, We Control the Tools

For legal professionals, the instinct to tighten controls amid uncertainty is understandable. Managing risk is part of the job.

Imagine an organization that has spent months introducing AI agents into its operations. Employees have learned to rely on them for research, document review, drafting, and other complex workflows. Then leadership becomes concerned about reports of unauthorized agent behavior and proposes suspending access across the company until the risks are better understood.

It sounds cautious. Perhaps even responsible. But what risk is that decision actually reducing?

Disabling an employee's approved AI assistant may eliminate certain risks associated with that tool. However, it does little to prevent an external attacker from exploiting a vulnerability in the organization's document management system, compromising a cloud provider, or obtaining credentials through a third-party software breach.

Meanwhile, employees lose capabilities they have incorporated into their daily work. Productivity suffers, projects slow down, and some employees may seek unauthorized alternatives. The organization also interrupts its own technological development at a time when becoming more capable with AI may be increasingly important.

When confronted with risks we cannot control, we tend to become more controlling about the things we can.

I think of this as *risk displacement*.

We cannot stop an unknown attacker from discovering a vulnerability in software our organization depends on, but we can turn off our employees' AI agents. We cannot dictate the pace of AI development, but we can delay our own AI initiatives.

These decisions may create a sense of security because they are visible and enforceable. But managing compliance with an internal policy is not necessarily the same thing as reducing exposure to the organization's most consequential threats.

Feeling more in control is not the same as being better protected.

## Better Controls, Not Blanket Restrictions

This is not an argument for giving AI agents unrestricted access to company systems or confidential information.

An agent helping an attorney organize approved research materials presents a very different risk from one authorized to transmit privileged documents, modify client records, or initiate financial transactions.

Organizations should set clear boundaries around the tools agents can use, the information they can access, and the actions that require human approval. Strong identity management and limited permissions also help contain the damage if an agent or its credentials are compromised.

But a blanket prohibition treats very different activities as though they present the same danger.

The more consequential question may be whether organizations are paying enough attention to threats developing outside their walls.

How quickly can organizations identify and repair newly discovered vulnerabilities? Do we understand which vendors and third-party applications have access to sensitive client information? Can we detect suspicious activity and respond before a compromise becomes a major incident?

For legal organizations, these questions directly affect client confidentiality, professional responsibility, and business continuity. An attack that interrupts access to client documents or critical systems can be far more damaging than many of the internal AI risks that receive disproportionate attention.

Legal leaders need to work closely with IT and InfoSec, not just to approve or reject AI tools, but to ensure the organization is prepared for a rapidly changing threat environment.

## Fighting AI-Powered Threats With AI-Powered Defenses

Fortunately, the same technology creating new cybersecurity challenges may also provide some of our most effective defenses.

OpenAI's Daybreak initiative illustrates this possibility. It brings advanced AI capabilities into authorized cybersecurity workflows, helping qualified security professionals discover, investigate, validate, and remediate vulnerabilities. Daybreak is a specialized initiative rather than a product every law firm can simply deploy, but it points toward an important change in how cybersecurity may need to operate.

If AI is accelerating vulnerability discovery, organizations must get faster at identifying which threats matter and closing security gaps before attackers can exploit them. That is a much more productive response to The Hackening than indiscriminately restricting employees' access to AI.

The lesson for legal leaders is not that internal AI governance no longer matters. It is that governance alone cannot protect an organization from a rapidly changing external threat environment. We need thoughtful controls over how employees use AI, alongside stronger cybersecurity defenses, more resilient systems, and a willingness to use AI itself to improve our ability to respond.

The emergence of The Hackening should make us more serious about AI risk, not reflexively more restrictive about AI use.

Protecting client information and fulfilling professional responsibilities remain essential. But those obligations cannot be satisfied simply by limiting the tools employees are permitted to use.

We may be spending too much time managing the risks of using AI, and not enough time managing the risks of a world in which everyone else is using it.

---

***Brandi Pack** is the Director of Innovation at **UpLevel Ops**, where bold ideas become practical tools that transform legal departments. She leads the firm's **AI Strategy & Enablement** programs, delivering AI-powered solutions that are accessible, secure, and built for teams ready to work smarter, not harder. With a background spanning Legal Ops, tech, and project management, Brandi is passionate about turning what's next into what's Next Level. She is also a frequent advisor and*

*speaker on AI adoption in legal, known for translating emerging tech into practical wins for in-house teams.*

*Did someone forward this* ***Brainwave*** *feature article to you?* ***Subscribe here*** *to receive the monthly* ***Brainwave by UpLevel Ops*** *newsletter directly in your inbox, along with industry insights and news your legal team can use to make practical, strategic decisions about your business's future.*

Ready to learn more? ***Talk to UpLevel Ops today!***