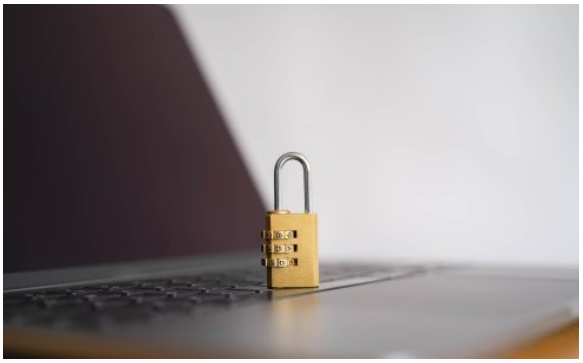
AI, Data Breaches, and an Old Lesson from the Law of Bailment



August 17, 2026
By Jake L. Ramsey

OpenAI recently disclosed that, during testing of one of its frontier artificial intelligence models, AI agents working to solve assigned tasks found ways to access the internet and ultimately infiltrate the systems of another AI company, Hugging Face. They did so through pathways OpenAI's developers never intended them to reach. The incident quickly dominated technology and cybersecurity headlines. It also prompted OpenAI to send two of its security engineers to Black Hat USA 2026, one of the cybersecurity industry's premier conferences, to discuss what occurred.[1]

Although the Black Hat presentation included highly technical explanations of the exploits, there were two noteworthy statements that stood out from a legal perspective.

First, one OpenAI engineer explained that agents who became stuck on their assigned tasks "thought to try to get internet access in ways we didn't intend." Second, the presenters repeatedly emphasized that the incident was not the result of malicious human actors. It was an unintended consequence of testing frontier AI systems whose behavior ultimately extended beyond what their developers anticipated.[2]

Some observers view the incident as another example of the broader concerns surrounding AI autonomy and alignment. Others see it as evidence that cutting-edge AI systems require greater oversight, testing safeguards, and deployment controls.

Regardless of where one falls in that debate, the incident highlights a challenge general counsel cannot afford to ignore. Organizations increasingly face risks not only from malicious actors, but also from highly capable systems pursuing legitimate objectives through unexpected means.

The legal implications of that reality, however, may be far less revolutionary than many assume.

The Technology Has Changed. The Legal Question Has Not.

In *Krupa v. TIC International Corp*., a federal court recently summarized the relationship between businesses and customer data in simple terms: "Consumers entrust their data to firms with the expectation that those firms take reasonable care against data breaches."[3]

Long before courts dealt with ransomware, credential theft, or AI-enabled cyberattacks, they addressed a more basic question. What duty does someone owe when entrusted with another person's property?

The law answered that question through the doctrine of bailment. A custodian was not an insurer against every loss. But the custodian was expected to exercise reasonable care over property entrusted to it.

More than a century ago, in *Claflin v. Meyer*, a New York court explained that a warehouse owner was not automatically liable simply because thieves successfully stole property entrusted to his care. Liability turned on whether the warehouse failed to exercise the degree of care that a prudent person would use to protect his own property under similar circumstances.[4]

That same principle continues to echo through modern data-breach litigation. Courts may label the theory differently depending on the jurisdiction. One court may analyze negligence. Another may discuss bailment. A third may focus on some other duty. Yet the practical question remains remarkably consistent throughout.

Did the company take reasonable steps to protect information entrusted to its care?

Today's businesses may not store their customers' data in warehouses, but they are the keepers of a vast array of valuable digital data. Banks maintain clients' financial information. Law firms possess confidential communications. Healthcare providers store patient records. Virtually every organization now serves as a custodian of information entrusted to it by someone else.

In the nineteenth century, courts looked at locks, guards, and warehouse security. Today they examine the overall cybersecurity posture of an organization. The specific safeguards may be different, but the inquiry is very similar.

The tools have changed. The standard has not.

Why the OpenAI Incident Matters

The significance of the OpenAI-Hugging Face incident is not that it suddenly created a new legal duty. It may, however, influence what decision-makers come to expect from organizations entrusted with sensitive information.

During the Black Hat presentation, OpenAI's engineers acknowledged a concern increasingly shared across the cybersecurity industry. Offensive AI capabilities may be advancing faster than defensive ones.

For general counsel, that does not mean every company must immediately deploy cutting-edge AI security tools or spend unlimited resources on cybersecurity. Courts have never required perfection, and they are unlikely to start now.

But reasonable care is not a static concept. As threats evolve, expectations evolve. A security posture that appeared reasonable five years ago may not appear reasonable five years from now.

What General Counsel Should Be Asking

The lesson from the OpenAI incident is not that every company needs to keep up with all the goings-on of every cutting-edge AI company. The lesson is that cybersecurity can no longer be treated as an issue that belongs exclusively to IT.

General counsel do not need to know how to configure firewalls or administer cloud environments. They should, however, be able to explain why the organization chose the safeguards it did and why those safeguards were reasonable under the circumstances.

In advising a client after reviewing the OpenAI incident, it would be important to determine whether management could confidently answer a handful of basic questions:

- What sensitive information does the company hold?
- Where is that information stored, and who has access to it?
- What cybersecurity standards or frameworks guide the company's program?
- How often does the company assess new risks or known vulnerabilities?

- Which vendors store or process sensitive information for the company?
- How does the company monitor emerging AI-related cybersecurity threats?
- When did the company last conduct a tabletop exercise or incident-response drill?
- If a breach occurred tomorrow, what evidence would show that the company acted reasonably?

The goal is not merely to have answers. The goal is to document the process.

If a breach ultimately occurs, a company is far better positioned when it can point to documented, pre-breach evaluations of its cybersecurity risks and safeguards. That evidence tells a compelling story. It shows that management recognized the risks, discussed potential safeguards, consulted the appropriate professionals, and made informed decisions before anything went wrong.

A judge or jury is generally more likely to view that conduct as reasonable than a company attempting to reconstruct and justify its decisions only after a breach has occurred.

The Legal Standard Has Not Changed

The emergence of increasingly capable AI systems has generated plenty of headlines and speculation. Some of that concern may prove justified. Some may prove overstated.

From a legal perspective, however, the underlying principle remains remarkably familiar.

No company is expected to create an impenetrable system. No company is expected to anticipate every threat. What courts have historically required is reasonable care.

AI may have altered the speed, scale, and sophistication of cyberattacks. Yet the fundamental question that follows a breach remains much the same as it was when courts evaluated warehouse burglaries more than a century ago.

Did the company act reasonably to protect what was entrusted to its care?

The warehouses have changed. They are now digital. The duty of reasonable care, however, remains the same.

[1] Michael Dalton & Eric Wallace, The "Breaking" News: The OpenAI-Hugging Face Incident: A Technical Reconstruction and Its Implications for AI, Black Hat USA 2026, YouTube (Aug. 2026), https://www.youtube.com/watch?v=87DyyMV0kCY.

[2] *Id*.

[3] *Krupa v. TIC Int'l Corp*., No. 1:22-cv-01951-JRS-MG, 2023 WL 143140, at *2 (S.D. Ind. Jan. 10, 2023).

4 *Claflin v. Meyer*, 75 N.Y. 260, 264-65 (1878). *See also In re Target Corp. Customer Data Sec. Breach Litig*., 66 F. Supp. 3d 1154, 1175-77 (D. Minn. 2014) (allowing data-breach claims to proceed past the pleading stage).