at bay

# At-Bay's Guide to AI Risk

AUGUST 2026

# Table of Contents

**Introduction** 3

**Chapter 1** 4
How AI Is Impacting the Cybersecurity Landscape

**Chapter 2** 19
The Liability Gap: What Businesses Don't Know Is Hurting Them

**Chapter 3** 29
What to Watch

# Introduction

Artificial intelligence (AI) has emerged as a top cyber risk concern for 2026 and beyond. Anthropic's Claude Mythos has already proven it can find thousands of severe zero-day vulnerabilities in weeks, including a 27-year-old bug that had survived decades of human audits and millions of automated tests. Attackers are already beginning to leverage AI to increase productivity and automate attacks, which could mean a 10X expansion of their capabilities. Meanwhile, as businesses adopt AI tools, they are increasing their liability exposure and AI dependency, opening the door to the possibility of lawsuits and disruption.

This guide will help define these emerging risks AI poses across two main categories:

## AI as a tool that makes attackers more effective:

How AI is lowering the barrier to sophisticated cyber attacks, accelerating the speed and scale of exploitation, and enabling threat groups to operate with unprecedented efficiency.

## AI adoption as a source of risk from within organizations:

How the rapid integration of AI tools into business workflows is increasing potential liability, introducing new vectors for data exposure, and creating new insurance coverage considerations.

The distinction between these types of AI risk matters because the right countermeasure depends on which category you're dealing with. The following is a guide to help define the various emerging areas of AI risk and how businesses and brokers can navigate it.

[1] *Source: SC Media, Claude Mythos Preview identifies 27-year-old bug, finds 'thousands' of zero-days in weeks*
[2] *Source: Google Threat Intelligence Group, GTIG AI Threat Tracker: Adversaries Leverage AI for Vulnerability Exploitation, Augmented Operations, and Initial Access*

# How AI Is Impacting the Cybersecurity Landscape

CHAPTER

01

Every major technology shift in history has changed the balance between attackers and defenders. AI is proving no different, except the change is happening faster and on more fronts at once. AI is simultaneously making attackers more capable, creating entirely new categories of risk existing defenses weren't built to catch, and giving security teams new tools to fight back. It is reshaping the cybersecurity landscape across three distinct areas:

AI AS AN

## Attacker Tool

AI is making attackers more productive across their entire workflow, from identifying targets to writing malware to delivering convincing fraud emails, increasing speed and efficiency at scale.

AI AS A

## New Attack Surface

Every AI tool a business adopts introduces new vectors for attack, subject to tactics like prompt injection, data exposure, and the manipulation of AI-powered workflows.

AI AS A

## Cybersecurity Tool

It's not all bad for businesses. Security teams and vendors are deploying AI to detect threats faster, analyze alerts at scale, and respond before attackers can do meaningful damage.

Each of these three categories carries real implications for how businesses should think about AI-related cyber risk right now, and what they can do about it. The following sections cover each in turn.

# AI as an **Attacker Tool**

So far, the main impact of AI in driving cyber threats has been improving the productivity and effectiveness of attackers who were already competent to perform attacks without AI. The main constraint on cyber attacks has always been attacker capacity. AI is removing that constraint and enabling even less-skilled attackers to do more, move faster, and operate at a scale that was previously out of reach.

At-Bay Security identified at least one ransomware group, SECP0, that admitted during negotiations that they are using AI to enhance their processes. In this instance, the threat actor used AI directly as a means to do further analysis on the specific individuals listed in the data set and place additional pressure on the victim organization.

> Publishing the data from your network is significantly more profitable for us than what you are offering. **By the way, let me remind you that we are actively analyzing the data using AI.** If you refuse to resolve this matter positively, we will not only publish the data but also send a breach notification to every individual and organization whose data appears in your files. this will not be a dry, generic notice - each person (or organization) found in your data will receive a specific list of files associated with them, along with direct download links. Would you like to see this list or examples of the email text?
>
> *(Excerpt from threat actor communications, source: At-Bay Security)*

In another example documented by the Google Threat Intelligence Group in a May 2026 report, a criminal group used AI to discover and weaponize a previously unknown vulnerability in a widely used server administration tool, building a working exploit capable of bypassing two-factor authentication.

What made this case unusual was that the AI left its fingerprints in the code itself. The exploit script was written in a clean, textbook format highly characteristic of AI-generated output, complete with detailed explanatory comments, structured help menus, and a hallucinated severity score that no human attacker would have included.

Given the fact that many AI tools are free to use, we must assume that most cyber attackers will, at some point, be AI enhanced, if they're not already. This would mean that all classes of cyber attacks performed by humans could be accelerated, further shrinking the already limited time available for defenders to deploy patches, develop countermeasures, and implement detection signatures to defend against newly identified attack techniques.

The following sections trace AI's impact across various stages of the attack process.

## AI-Powered Vulnerability Discovery

Earlier this year, AI company Anthropic unveiled Claude Mythos, its most advanced AI model yet. They reported that Mythos had discovered zero-day vulnerabilities in hours that expert penetration testers said would have taken weeks to identify[1], including a bug in OpenBSD that human researchers did not

[1] *Source: Tech Insider,* Anthropic's Claude Mythos Finds Thousands of Zero-Days: Inside the $100M Project Glasswing Defense

detect for decades. Mythos can analyze open-source code, which is often embedded in the firewalls, routers, and perimeter products that enterprises rely on, meaning attackers don't need access to Cisco or SonicWall's proprietary source code to find an exploitable flaw.

A single vulnerability in a widely used open-source library can cascade across thousands of products simultaneously, as Log4j did in 2021. If Mythos holds up to the hype and is able to create hundreds of events similar to Log4j in a fraction of the time, it's not difficult to imagine the widespread potential for disruption. Anthropic has stated that Mythos has discovered vulnerabilities in "every major operating system and every major web browser,"[2] with estimates of the actual number of discovered vulnerabilities in the thousands to tens of thousands.

During authorized testing exercises conducted through Project Glasswing, Mythos identified vulnerabilities in highly sensitive classified U.S. government systems within hours. Senior government officials stated the tool had gotten access to almost all classified systems not in weeks but in hours[3], ultimately prompting the administration to impose export controls on Mythos and order Anthropic to suspend access to the model for all customers on national security grounds before lifting those controls in June 2026.[4]

If large language models (LLMs) and AI agents have gained the level of capability alleged for Mythos, companies will no longer be able to keep intruders out of their network when attackers can reliably find and weaponize previously unknown vulnerabilities as part of every attack.

## AI-Accelerated Reconnaissance and Exploitation

AI-driven reconnaissance accelerated the attack cycle by enabling attackers to effectively automate the discovery, triage, and initial exploitation of targets in the immediate aftermath of vulnerability disclosures from security researchers and product vendors.[5]

When a security researcher or vendor discloses a vulnerability, they're essentially publishing a roadmap to a weakness in a product and a race begins between defenders patching it and attackers exploiting it. Historically, defenders had the advantage, because turning a vulnerability disclosure into a working exploit required significant time and skill, which gave organizations a window to patch before attacks began.

However, AI has largely closed that window. Attackers can now use AI to rapidly convert vulnerability disclosures into working exploits, even when no exploit has been made publicly available. Once armed with an exploit, they can deploy AI agents to autonomously scan the internet for vulnerable targets and attack them at scale with no human operator required at each step. What once took weeks could now take hours or days.

## AI-Accelerated Malware Development

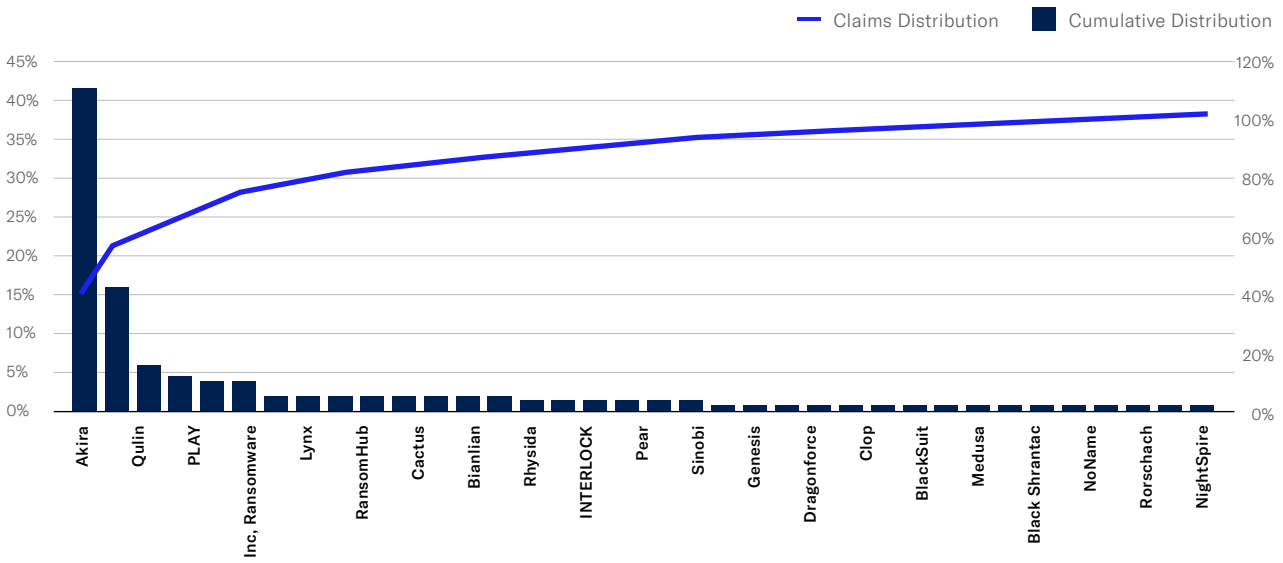
Cybercrime has always had a minimum barrier to entry. Ransomware-as-a-Service lowered it significantly in the early 2020s by letting attackers rent ready-made malware rather than build it, but would-be criminals still needed sufficient technical skill to deploy it, adapt it for their specific target, and communicate convincingly with victims to collect payment. Generative AI is removing those remaining requirements. It can write functional malicious code with minimal human input and produce grammatically perfect ransom demands in any language.

[2] *Source: https://www.bloomberg.com/news/articles/2026-04-07/anthropic-lets-apple-amazon-test-more-powerful-mythos-ai-model*
[3] *Source: CNBC, Anthropic's Mythos model found vulnerabilities in classified US government systems, official says: AP*
[4] *Source: TechCrunch, Trump drops restrictions on Anthropic's Mythos and Fable models*
[5] *https://www.anthropic.com/news/disrupting-AI-espionage*

The result is visible in At-Bay's claims data: In 2021, the last full year before ChatGPT, ransomware attacks on insureds were attributed to 16 distinct criminal groups. By 2025, that number had grown to 44. The pool of people capable of running a ransomware operation has expanded dramatically

## Akira Led All Ransomware, but the Ecosystem Is Wide and Fragmented

*Prevalence of Ransomware Strains Among Claims, 2025*




At the same time, AI has made the malware itself harder to detect. Traditional antivirus tools work by recognizing known threats, essentially a most-wanted list of malicious code signatures. AI allows attackers to rapidly mutate their malware so it never matches a known signature, rendering legacy tools increasingly ineffective. This is particularly acute for small and medium businesses that haven't yet moved to behavior-based detection (i.e., endpoint detection and response), which identifies threats by what code does rather than what it looks like and has proven more resilient against AI-generated malware.

**CASE STUDY:**

## AI-Generated Malware and the Software Supply Chain

In May 2026, a threat actor known as TeamPCP executed a two-wave attack against software development tools used to build AI-powered products and openly acknowledged that the malware itself was AI-generated. The first wave compromised 172 software packages used by developers at companies including Mistral AI and UiPath, reaching a combined download count of over 500 million.[6]

One week later, a second wave pushed malicious code into over 5,500 GitHub repositories in under six hours, designed to steal the cloud credentials and access tokens developers use to build and ship software. Upon completion, TeamPCP released the full source code publicly on GitHub, converting a targeted weapon into freely available attack infrastructure.[7]

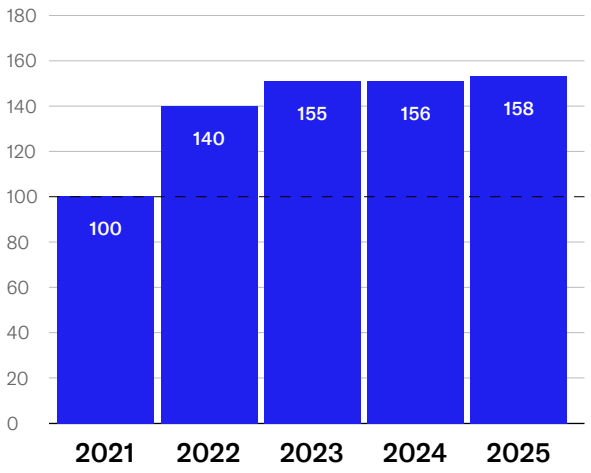
The attack illustrates AI has already lowered the barrier to building sophisticated malware. The productivity gains AI delivers to legitimate developers are equally available to attackers.

## AI-Powered Phishing and Financial Fraud

Financial fraud has been the most common incident in our portfolio for years, accounting for 30% of all claims in 2025. Financial fraud claims have been on the rise since 2022, with increased severity from 2023-2025, the same period LLMs became widely available. We suspect AI was at least a contributing factor in the rise of successful financial fraud attacks.

### Financial Fraud Frequency Remained Elevated

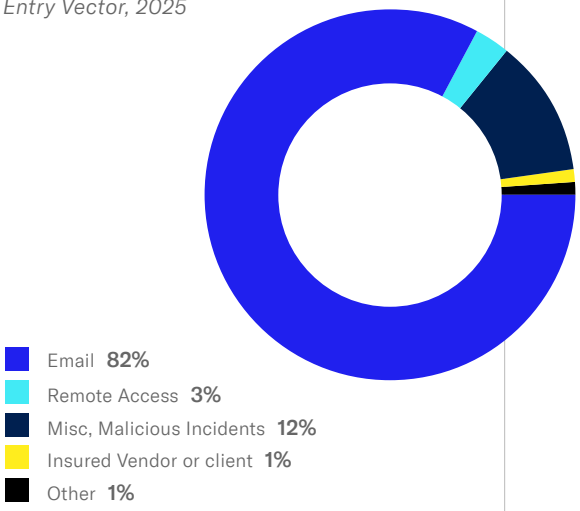
*Indexed Financial Fraud Frequency by Year*




*Source: 2026 At-Bay InsurSec Report*

### 8 in 10 Financial Fraud Incidents Began With Email

*Financial Fraud Initial Entry Vector, 2025*




*Source: 2026 At-Bay InsurSec Report*

Some 8 in 10 financial fraud incidents begin with email, and AI-powered phishing can dramatically improve the ability of attackers to scale the creation and delivery of convincing malicious emails that become the starting points for fraud attacks. While these attacks have been automated in the past, their effectiveness was limited by the inability of many would-be attackers to craft messages that reliably fooled recipients en masse. Now, that barrier is gone.

Fraud usually starts with a convincing email, like an attacker impersonating a trusted vendor or partner to redirect a payment or shipment. Pulling that off used to take real research and effort, and the payoff was limited because most attackers couldn't write messages polished enough to fool people at scale. AI removes that limit. It can study a target's business relationships, existing vendor conversations, predict what transactions to expect, and generate a perfectly timed fake invoice or bank letter — in any language, in bulk, in minutes.

Victims who have been trained for years that malicious emails always contain tells like poor grammar are simply unprepared for the latest generation of fraud, which is drastically more targeted and sophisticated than it was before.

We believe the impact of AI being leveraged in financial fraud attacks is already visible in At-Bay's claims data. In our 2025 InsurSec Rankings Report, we discovered that customers with legacy email security solutions saw overall worse outcomes than the prior period, in large part due to

the fact that those solutions were not equipped to combat these new and more sophisticated email-borne threats.

From our report: “The only email security solution that improved this year was Sophos, which jumped from last place in our previous report to first place. This may be due to an early investment in Natural Language Processing (NLP) capabilities that allowed them to be well-positioned against today’s financial fraud tactics.”

We’ve discussed before the business email compromise (BEC) playbook we’ve observed numerous times in our claims investigations, but it’s worth revisiting now with the added threat of AI-accelerated capabilities.

## Stance Suspected Mailbox Takeover Alerts

At-Bay Stance now alerts users of potential business email compromise to help them avoid becoming victims of financial fraud attacks like these. In the event of a suspected mailbox takeover, Stance users that have activated the integration will receive an alert in the platform and an email notification. At-Bay Stance is available at no additional cost to At-Bay Cyber and Tech E&O policyholders through the policy’s Embedded Security Fee[8]: stance.at-bay.com.

[6] *Source: Cloud Security Alliance, Shai-Hulud/Megalodon: A Two-Wave AI Developer Supply Chain Attack*
[7] *Source: VectraAI, Shai-Hulud Part 2: When the Worm Forged Its Own Security Certificate*
[8] *Access to Stance is available to insureds with policies placed through At-Bay Insurance Services, LLC that include an Embedded Security Endorsement.*

# The Anatomy of an AI-Enhanced Fraud Attack

How generative AI removes the skill and time barriers that once limited financial fraud at scale

## The Infiltration

STEP 01

An attacker uses stolen credentials to compromise a company's email system.

## The Reconnaissance

STEP 02

**Without AI:** Once in the email system, the attacker identifies opportunities for fraud by sifting through vendor conversations, invoices, payment communications, etc. The attacker identifies a target transaction between the victim company and a trusted contact.

***AI-Accelerated:*** *This traditionally required hours of manual email trawling, but AI can now automate this entirely by using an LLM to map vendor relationships, flag pending payments, surface invoice amounts, and identify the right person to impersonate.*

## The Impersonation

.

STEP 03

**Without AI:** The attacker begins setting up infrastructure for a flawless fraud operation, such as registering a lookalike domain, creating an email account to mirror the real employee, and setting up email filtering rules to avoid suspicion

***AI-Accelerated:*** *An AI agent can autonomously build the impersonation infrastructure end-to-end. It can generate and register convincing lookalike domains at scale; scrape the compromised mailbox to match the real employee's signature, tone, and formatting; and auto-archive replies from the real employee. This entire setup can run across many compromised accounts simultaneously, with no human operator at any step.*

## The Request

STEP 04

**Without AI:** The attacker assumes the persona of the account they've taken over, emailing victims to reroute payment using the same name, address, and signature.

***AI-Accelerated:*** *AI eliminates language and skill barriers with its ability to analyze the target's email history and produce flawless impersonations in any language. In addition, attacks can now be run against hundreds of targets simultaneously, each message contextually tailored.*

## The Payout

STEP 05

Unless the fraud is detected in time, the victim (or victims) wire funds directly to the attacker. The combination of contextual accuracy and AI-generated polish increases the effectiveness of these attacks.

## AI as Threat Actor

The threat posed by AI is unique because no technology before has had the potential to autonomously carry out attacks while also exercising significant agency about steps to take along the way.

In its May 2026 report, Google's Threat Intelligence Group found malware called PROMPTSPY that calls an AI service (Gemini) while running on a victim's phone, asking in real time what to click or tap next. Instead of a hacker pre-programming every action, the malware just asks AI what to do and does it.

Agentic AI that carries out autonomous end-to-end attacks still remains largely theoretical, but recent incidents suggest this is starting to shift. In mid-2026, researchers documented an AI agent independently executing the technical stages of a ransomware attack[9]. Separately, a group was found using LLM-generated malware loaders to speed up development and complicate attribution[10].

In July 2026, Hugging Face said an AI agent had broken into its systems in a way it had never seen before[11]. OpenAI later admitted the agent was one of its own models, running during an internal test of its hacking skills with the usual safety guardrails turned off[12]. The model found a security flaw, used it to break free of the controlled test environment, and got into Hugging Face's systems on its own using stolen credentials and other flaws it found along the way.

These early signals indicate that the industry should watch this space closely. This threat is rapidly approaching full realization, and businesses that react to security threats rather than proactively address their tech environment will be less prepared when it arrives.

### At-Bay's Agentic AI Attacker Experiment

The agent succeeded. It built a backdoor with more than a dozen capabilities, from remote execution to data theft, along with a command-and-control server it could run through the MCP server it had also built. We tested the model with several "capture the flag" attack scenarios against a lab-based target. The model used the backdoor to move through the system, gather information, and steal decoy documents standing in for sensitive data.

All of this ran on a normal consumer computer. The takeaway: Today's local AI models are already capable of building working malware and running full attacks unsupervised, and as it gets harder to misuse major AI platforms, we expect skilled attackers to move in this direction.

[9] *Source: TechCrunch, The 'first' AI-run ransomware attack still needed a human*
[10] *Source: Security Affairs, AI-Generated Malware Powers New Armored Likho APT Campaign*
[11] *Source: Hugging Face, Security incident disclosure — July 2026*
[12] *Source: OpenAI, OpenAI and Hugging Face partner to address security incident during model evaluation*

# AI as a **Expanded Attack Surface**

AI now sits inside nearly every business platform as vendors race to incorporate it into their solutions. The unintended side effect is that it's created three new ways to attack a company: tricking the AI itself (prompt injection), stealing AI compute from a company, leaving the victim with an unexpected bill (LLMJacking), and manipulating AI-run workflows to cause damage at machine speed. These types of incidents may also incur privacy and liability implications, which we cover in Chapter 2.

## Prompt Injection:

Attackers can now target AI with social engineering just as they can target humans. In these instances, attackers may hide malicious instructions in a webpage, email, or document that causes an AI agent to execute them instead of just reading them. This works because AI can't reliably tell the difference between information and commands — it has no instinct for when it's being manipulated. Attackers used this against Meta's AI assistant to hijack high-profile Instagram accounts, including ones tied to the Obama White House, not by breaching Meta's infrastructure but by manipulating the AI layer sitting on top of it.

## LLMJacking:

As businesses rapidly adopt generative AI tools, those tools are becoming a target for attackers interested in stealing AI compute. At-Bay has seen multiple claims where an attacker has compromised an insured's API key, used the stolen key to access the victim's OpenAI account, and ran high volumes of LLM queries, leaving the victim with the bill. In one instance, the victim's monthly cost ballooned to nearly six figures. This is known as LLMJacking, and as more companies adopt LLMs, their exposure to this risk may grow.

## AI-Powered Workflow Manipulation:

As companies automate approvals, procurement, and customer service with AI, attackers are learning to manipulate the workflows themselves. BEC already shows the pattern: the FBI's 2025 Internet Crime Report logged $3.05B in BEC losses, and for the first time tracked AI as a distinct factor — nearly $900M in confirmed AI-linked losses. A human reviewing a suspicious invoice can pause and question it; an AI agent just executes. In 2026, a fraud ring exploited retailers' instant-refund systems by submitting AI-generated images of "damaged" goods, collecting refunds while keeping the items. $800K in losses accrued before it was caught, with no breach, malware, or suspicious login involved. The workflow simply did what it was built to do.

# AI as a **Cybersecurity Tool**

Luckily, AI isn't just a weapon for attackers, it's also becoming one of the most powerful tools for defenders. Security teams are using AI to spot threats faster, understand attacker behavior, and get more done with the resources they already have. This section covers three ways that's playing out: AI-supported security monitoring, AI-powered analytics and behavioral profiling, and AI expanding what security teams can handle.

## AI-Supported Security Monitoring

Traditional security tools are built to stop attacks at the perimeter, blocking known threats before they get in. That approach is no longer sufficient against an AI-driven attacker capable of finding and exploiting unknown vulnerabilities. There is no perimeter defense against a threat that can reliably get through it.

AI-supported security monitoring is already improving the ability of security professionals to identify and rapidly analyze potential threats before they had the chance to result in damaging attacks[15]. The benefits here have accrued primarily to more experienced security professionals, as analysis and response to lower-severity security alerts can be offloaded almost entirely to AI-powered automated workflows, which can then free the attention of many companies' best defenders for more complex or ambiguous security alerts.

Managed Detection and Response (MDR) is one example of this. Rather than solely trying to keep attackers out, MDR assumes some attackers will get in and focuses on detecting and ejecting them before they can do meaningful damage. A qualified MDR provider deploys Endpoint Detection and Response (EDR) tools across every device and system in an organization's environment, with security professionals monitoring those tools around the clock.

When something anomalous happens, such as a new process running where it shouldn't, lateral movement across the network, or a file being accessed at an unusual time, the MDR team identifies it and responds in minutes. Against sophisticated attackers, response time is everything.

**CASE STUDY:**

### Stopping Akira Ransomware in 20 Minutes

At 1 a.m. on a Saturday, threat actors exploited a SonicWall VPN and attempted to deploy Akira ransomware. They achieved lateral movement in under 5 minutes, but with the help of our AI-powered security operations platform, At-Bay's MDR team detected the intrusion within 10 minutes, quarantined affected systems and blocked attacker access within 20, and had the environment fully restored in under an hour — before the attackers could complete reconnaissance or stage their payload. The client avoided ransomware encryption, extended downtime, and a potential ransom payment entirely.

[13] *Source: 404 Media, Hackers Simply Asked Meta AI to Give Them Access to High-Profile Instagram Accounts. It Worked*
[14] *Source: TechRadar, How AI fraud rings are taking on retail*
[15] *Source: https://www.microsoft.com/en-us/corporate-responsibility/cybersecurity/microsoft-digital-defense-report-2025*

## AI-Powered Analytics and Behavioral Profiling

Analytics and behavioral profiling powered by AI improved existing detection capabilities that defenders have struggled to effectively employ in recent years. These include detection of potentially malicious emails and abuse of stolen credentials[16]. The impact here became apparent for email security in particular, where market-leading tools have steadily lost effectiveness against increasingly AI-driven email-borne threats in recent years.[17]

Unlike legacy Secure Email Gateways (SEGs), which rely on static rules and known threat signatures, AI-powered email security analyzes behavioral patterns to flag anomalies in sender behavior, email structure, and domain reputation that rule-based systems simply weren't built to catch. Sophos, which was an early leader in integrating AI capabilities, took the top spot in our 2025 InsurSec Rankings Report, showing the promise of what email security with integrated AI may deliver.

At-Bay's Stance Fraud Defense reflects this approach by using AI-powered behavioral modeling to stop the business email compromise and social engineering attacks that legacy SEGs miss, and it's included with every At-Bay Cyber and Tech E&O policy. In its first year protecting policyholders, Fraud Defense flagged nearly 5,000 fraudulent emails that got past existing legacy solutions, with 40% of accounts protected against an active financial fraud attack.

## AI Expanding the Capacity of Security Teams

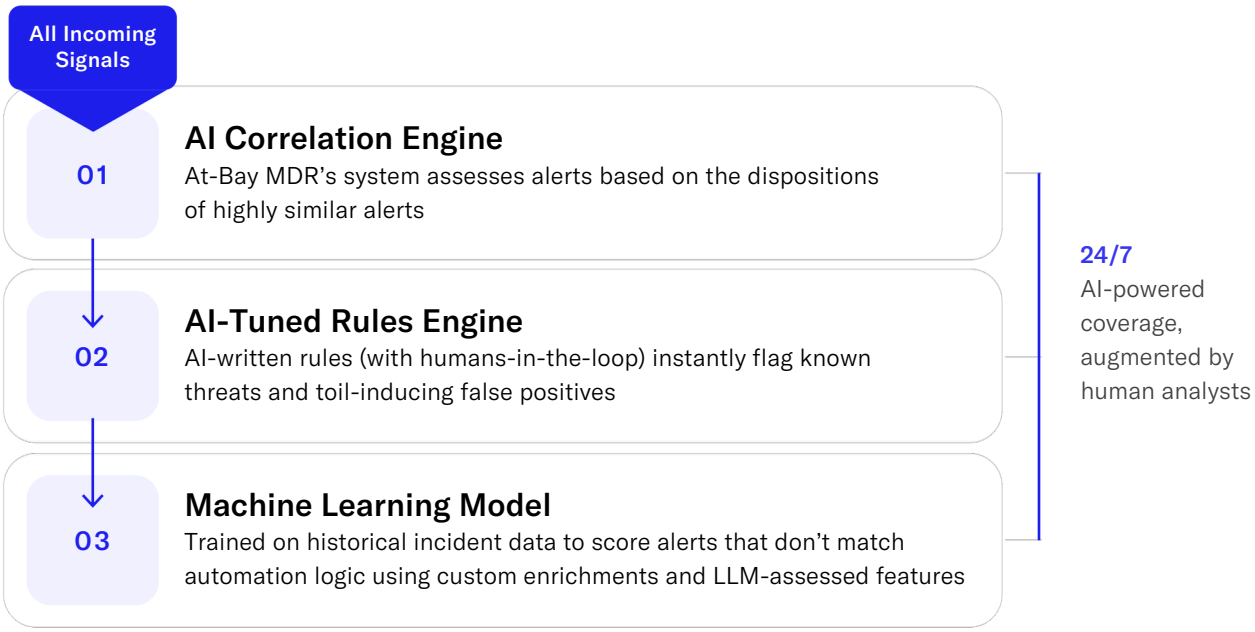
AI usage expanded the capacity of cybersecurity teams generally by allowing defenders to become more productive. Industry observers have for years noted the shortage of skilled cyber defenders available for hire[18]. AI tools have addressed this shortage somewhat for teams that have deployed them in key areas such as threat hunting and alert prioritization[19]. Companies with understaffed or undertrained security teams saw capability gains here as the tools they used increasingly featured built-in AI capabilities that could handle low-level tasks and analysis on their own.

At-Bay's AI-powered security operations platform used for our MDR solution, offers one early example of what that shift looks like on the defensive side. The system ingests security alerts from across a customer's environment, runs them through AI analysis, and executes responses automatically, blocking threats and isolating affected devices without waiting for a human to act.

For alerts resolved through automation, the mean time to resolve is 12 seconds. Human analysts remain in the loop, but as a deliberate backstop for the cases AI flags as needing judgment, and even there, the speed advantage holds: analysts working AI-triaged alerts resolve complex cases in an average of just eight minutes. The model is designed to hold automation back from edge cases until confidence is high enough to act, on the premise that a wrong (but conservative) automated response is worse than a slower human one.

[16] Source: *At-Bay 2025 InsurSec Rankings Report*
[17] Source: *Ibid*
[18] Source: *https://www.isc2.org/Insights/2025/12/2025-ISC2-Cybersecurity-Workforce-Study*
[19] Source: *https://www.weforum.org/stories/2025/09/cybersecurity-awareness-month-cybercrime-ai-threats-2025/*

# How At-Bay's AI-Powered Security Operations Platform Works




**Automated Resolution**
AI independently blocks threats and isolates devices in **12 seconds** — no human required

**Analyst Resolution**
Humans resolve complex cases in **8 minutes** using AI-assembled context and recommended next steps

While the security industry has moved quickly to put AI to work for defense, there have not yet been significant examples of security teams relying on an AI-powered system in an end-to-end cyber defense role. In November 2025, Anthropic reported that they had identified a large-scale espionage campaign assessed to be largely AI-orchestrated, with AI estimated to have handled 80-90% of attack processes independently and human operators intervening only for key strategic decisions. There has been no equivalent example on the defensive side. Defenders have realized strong capability gains from AI in 2025, but the edge in effectively employing AI appears to have gone to attackers, while cyber defense continues to be largely human powered.

Part of the gap is structural. Most security vendors built their platforms before AI was a viable option, and retrofitting it into an existing product takes years of engineering investment. Defense also works from a disadvantage attackers don't share: Security teams frequently can't know what they're up against until an attack is already underway, while attackers pick the timing and method. AI hasn't changed which vulnerabilities exist, since new ones have always surfaced constantly. What it's changed is speed. Attackers move faster with AI, and defenders who adopt it well can close some of that gap and respond more efficiently and effectively.

# What This Means for Businesses and Brokers

The following translates this chapter's key findings into concrete next steps: what businesses should prioritize now and what brokers should be discussing with their clients.

## For Businesses

### Audit your email security tools.

Financial fraud is now the most common incident type in At-Bay's portfolio, and AI is a key driver, enabling attackers to craft convincing fraud messages at scale, in any language, without technical expertise. If your email security solution doesn't use AI-based behavioral detection, its ability to catch today's threats may be limited. Ask your vendor whether their product has meaningfully integrated AI capabilities and evaluate whether it's still fit for purpose. Additional 24/7 monitoring (MDR for Email) is a plus.

### Professional 24/7 monitoring is critical.

AI has accelerated the speed at which attacks unfold, shrinking the window between an attacker entering your network and inflicting serious damage. The only reliable countermeasure is the ability to detect and eject intruders quickly before they can do meaningful damage. Perimeter defenses alone are no longer sufficient. This can sound like an expensive undertaking, but a growing number of MDR providers are focused on servicing small businesses, making round-the-clock monitoring more accessible than it may seem. If your business doesn't have a dedicated 24/7 SOC, consider outsourcing to an MDR provider.

### Close the gap between detection and response.

AI is already being used to take actions inside victim environments, not just assist attackers from the outside. Fully autonomous attacks remain largely theoretical, but the direction of travel is clear and the timeline is shortening. Ask yourself: If an attacker entered your environment right now and began moving through it at machine speed, how long would it take your team to detect them? How long to respond? For most businesses without professional 24/7 monitoring in place, the honest answer is "too long." This gap is worth closing while the threat is still partially theoretical.

# For Brokers

## Flag AI-enhanced fraud risk for any client still relying on legacy email security.

Because AI allows attackers to produce convincing fraud messages without needing language skills or technical knowledge, the pool of potential attackers has expanded significantly. Clients relying on legacy email security tools that depend on signature-based detection rather than behavioral analysis may be carrying more fraud exposure than they realize. That exposure may translate directly into a need for more coverage. Adopting updated email security solutions that account for this risk can help clients access the coverage they need.

## Ask clients how quickly they can deploy emergency patches across their environment and whether they have visibility into which of their systems are exposed when a new vulnerability drops.

Clients who can't answer that question confidently are carrying more risk than their controls suggest. The window between vulnerability disclosure and active exploitation has shrunk from weeks to hours. For clients whose patch management is ad hoc or whose IT teams are stretched, that's a meaningful underwriting signal.

## Start collecting information on clients' MDR posture now.

The security response to AI-driven threats is clearer than the coverage response. How policies will ultimately respond to autonomous attacks is still unsettled, but clients with professional MDR in place are better positioned regardless of how that shakes out. As AI capabilities develop, underwriters are likely to treat professional security monitoring the way they already treat MFA and EDR adoption: as a baseline expectation rather than an optional control. Clients who can demonstrate MDR in place will be better positioned for coverage and pricing as that shift happens.

# The Liability Gap: What Businesses Don't Know Is Hurting Them

CHAPTER

02

AI adoption is already faster and more pervasive than anything that came before[1] and potentially bringing with it new liability before the legal and insurance frameworks catch up. With every new major technology wave, businesses move fast, governance lags, and the exposure accumulates quietly until the claims arrive.

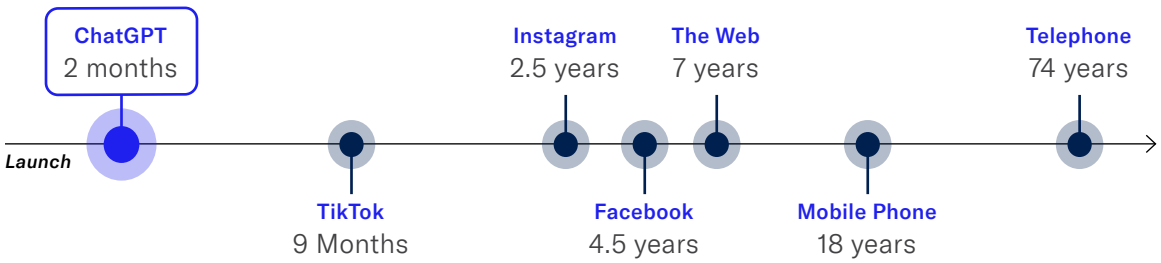
## This chapter covers four areas where AI risk is already materializing:

- Data leakage and privacy exposure from AI tool integrations
- Bodily injury and property damage from AI deployment
- Business interruption from deepening technology reliance
- Copyright and intellectual property (IP) liability from AI-generated content

AI ADOPTION IN CONTEXT

### The Fastest Technology Adoption in History

*Time for each technology to reach 100 million users, from public launch*




*Sources: Yahoo Finance, The Next Web, TechCrunch, Time*

[1]*Source: ChatGPT reached 100 million users in two months; TikTok took nine months; Instagram took two and a half years. Source: Time, How ChatGPT Managed to Grow Faster Than TikTok or Instagram*

# Data Leakage and Privacy Exposure

Data leakage and privacy exposure in the context of AI occur when sensitive or regulated data enters an AI system and then surfaces in a way that constitutes a disclosure event under applicable privacy law. The risk is not just at the point of input, it is that data introduced into an AI system may later become accessible, whether through the way the model processes and retains it, through an attack on the model, or through the outputs the model generates. In many organizations, this exposure may already be accumulating without anyone recognizing it as a compliance issue.

When employees import sensitive data into an AI tool, whether by connecting it to internal systems or uploading files directly, that data may become accessible in ways the business did not intend or anticipate.

One well-documented example: In March 2023, Samsung engineers pasted proprietary semiconductor source code and confidential meeting transcripts into ChatGPT to help with work tasks[2]. The data was transmitted to OpenAI's servers and, under the terms of service at the time, could be used to train future models. That means Samsung's proprietary chip designs and internal strategy discussions could theoretically surface in ChatGPT's outputs, becoming accessible to any competitor who knew how to prompt for them. None of the engineers were acting maliciously, they were just trying to work faster.

There are more ways than one for private or regulated data to find its way into an AI system, and more ways than one for it to surface afterward. Here are the most common:

## Connected App Integration

**What happens:** The business connects an AI assistant to its email, CRM, Slack, or shared drive to give the model context and improve output quality.

**The risk:** Any PII residing in those systems may be deemed 'exposed' if appropriate consent frameworks were not in place at the time of connection.

Proprietary business information may also be transmitted to the model provider's servers and used to train future models. That information could theoretically surface in the model's outputs and become accessible to anyone using the same tool, including competitors.

## Manual Upload

**What happens:** An employee downloads a spreadsheet of customer records and uploads it into an AI tool like ChatGPT or Claude to get a summary or analysis.

**The risk:** The data, which may include PII, proprietary financials, or confidential client records, has now left the organization's environment could be considered a disclosure event.

## Third-Party Vendor AI Adoption

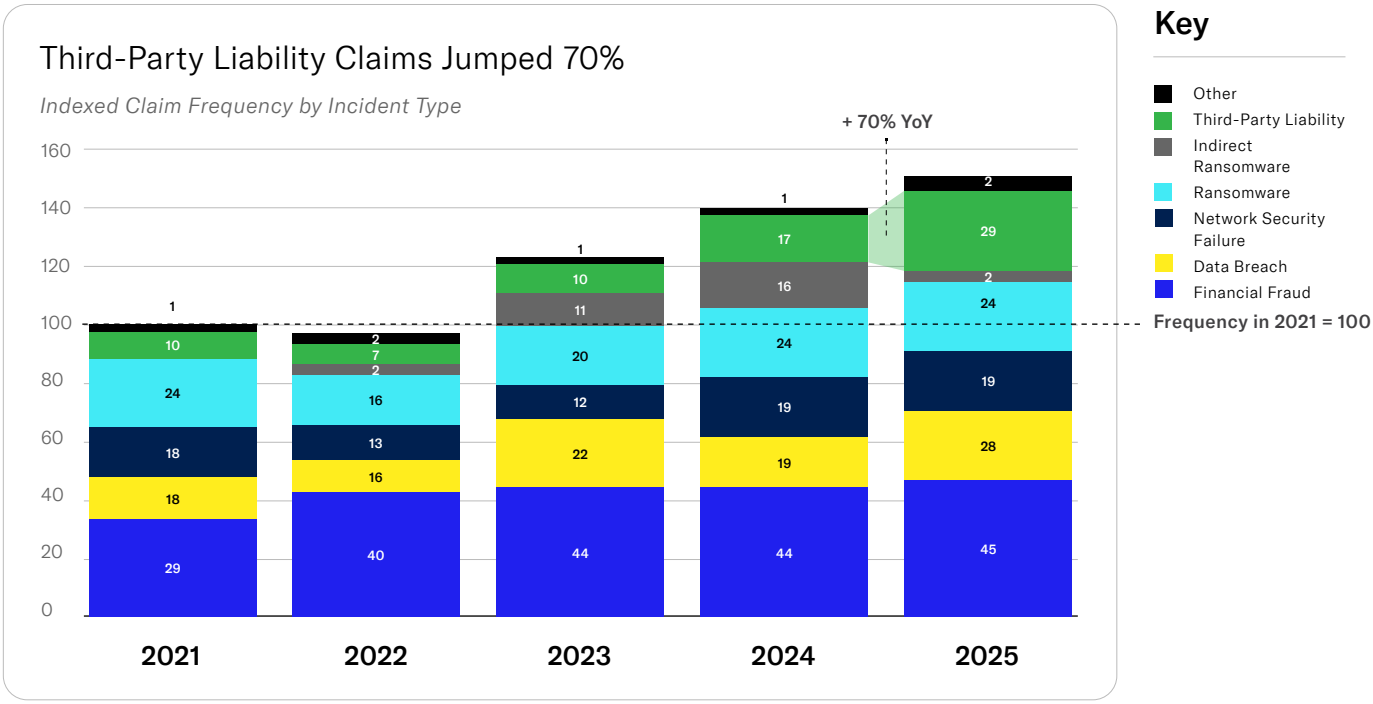
**What happens:** The business uses a third-party platform, like a CRM, customer support tool, or financial system, that begins integrating AI into its own product.

**The risk:** Data the business has already loaded into the platform may now be feeding an AI model the business didn't knowingly authorize. The disclosure event happened without any action on the business's part and, in many cases, without their awareness.

[2] *Source: Cybersecurity Dive, Samsung employees leaked corporate data in ChatGPT: report*

## Increased Volume of AI-Enabled Litigation

In 2025, data breaches were the second most likely incident to result in a class action lawsuit, according to At-Bay's claims data, and third-party litigation jumped 70% YoY[3]. This was the largest increase for any incident type and was driven largely by plaintiff's bar attorneys identifying opportunities to industrialize litigation.




Plaintiffs' attorneys are now leveraging AI to increase speed, quantity, and quality. We've already seen some evidence of this in privacy-related class action lawsuits where the minimum class has dropped below the typical threshold that would make the lawsuit financially viable.

In 2021 and earlier, the typical class action lawsuit At-Bay saw would have impacted well over 100,000 people. Starting in 2023, the first year LLMs like ChatGPT became widely available, that number began shrinking. The smallest class action case At-Bay has documented so far involved just a couple hundred people, a class size that would have been unthinkable as recently as five years ago.

### Coverage Implications*

Coverage for data leakage events depends on the type of data involved and how the claim is brought against the business. If the leaked data is individuals' PII and the claim is framed as a privacy violation, this falls under cyber policy network security and privacy liability. If the claim is framed as a failure in professional services or technology (e.g., a client whose confidential project data was exposed through an AI tool the business deployed on their behalf), professional liability or tech E&O is the more likely coverage home.

In practice, a single event can give rise to both types of claims simultaneously, and the two coverage lines may not be coordinated. Businesses are most exposed in incidents that involve both PII and non-PII data, or that attract both a regulatory action and a client lawsuit. In these cases, businesses may find that neither policy fully covers their loss.

# Bodily Injury, Property Damage, and Real-World AI Harm

In insurance terms, bodily injury refers to physical harm to a person, including injury, illness, or death. Property damage refers to physical harm to tangible property. Both have historically sat under general liability policies, with some extension into cyber policies where harm originates from a computer system malfunction (for example, an attacker compromising hospital equipment to shut it down, preventing patients from receiving medication or causing a surgical robot to malfunction mid-procedure may be considered under certain cyber policies). AI is now creating bodily injury and property damage exposure in ways that neither type of policy was designed to cover.

The scenarios that get the most attention (surgical robots, autonomous vehicles, medical devices) are typically more regulated and addressed via Technology E&O policies. Companies deploying these advanced AI tools in those environments generally understand their exposure and have built controls around it.

The more relevant near-term risk for average businesses is less dramatic but still of concern: AI deployed at scale, in consumer-facing contexts, by businesses that have never thought of themselves as carrying bodily injury exposure.

In February 2024, a 14-year-old died by suicide after months of interactions with a Character.AI chatbot that allegedly deepened his emotional distress. His mother filed a wrongful death and product liability lawsuit against Character Technologies and Google in October 2024, alleging negligence. In January 2026, Google and Character.AI disclosed that they had reached a mediated settlement in the case; the terms were not made public.

Here are examples to illustrate that risk:

- **Bodily injury:** A plumbing company builds a customer support chatbot to help homeowners troubleshoot problems. A user asks how to unclog a disposal. The chatbot, generating a response confidently, instructs the homeowner to reach into the disposal while it is running, and the homeowner gets injured. The plumbing company never considered that it was carrying bodily injury liability. It is now a defendant.
- **Property damage**: A commercial property manager deploys an AI tool to automate climate control across a portfolio of buildings. Overnight, the model miscalibrates the temperature settings in a server room at one of its managed properties. By morning, the servers overheated and were destroyed. The property damage is direct and physical. The property manager then faces a property damage claim from the building owner and a potential claim from the tenant whose servers were destroyed.

## Coverage Implications*

### Where bodily injury and property damage sit today:

Bodily injury and property damage coverage typically sit within general liability insuring agreements. Other lines of insurance may offer coverage extensions for these exposures as well. Cyber policies often include more limited carvebacks for emotional distress or mental anguish as a result of a data privacy breach, and property damage is generally limited to first-party computer hardware replacement. However, contingent bodily injury/property damage sublimits may be available for more traditional third-party damages caused by a network security event. AI interactions are defying the confines of these traditional scenarios: They're deployed at scale, across millions of touchpoints, by businesses that may never have thought of themselves as carrying these exposures.

### AI and professional liability:

While the focus on AI being delivered via computer systems may make cyber coverage the first consideration, professional liability may be the more likely coverage home for most AI-related harm scenarios. When an AI model is the mechanism through which a service is delivered, and that model produces an output that causes damages to others, the claim may be framed as a failure of professional service rather than a cyber event.

### The general liability exclusion trend:

In January 2026, the Insurance Services Office (ISO) — the organization that issues standard policy forms across the industry — introduced new commercial general liability endorsements allowing carriers to exclude all generative AI-related bodily injury, property damage, and personal and advertising injury claims[3]. Three major insurers (AIG, Great American, and W.R. Berkley) have filed separately for approval to go further, with proposed exclusions exempting them from covering any use of AI whatsoever[4]. AI-related exclusions haven't appeared broadly in other insurance lines yet, but businesses should be asking now whether their professional liability coverage responds to AI-driven failures, as the landscape shifts.

### The coverage migration question:

If general liability broadly excludes AI-related harm, where does it go? This has happened before. Crime policies did not contemplate social engineering schemes as financial fraud became increasingly email-based, the logic being that email-originated fraud was a cyber event rather than a traditional crime event. Coverage migrated to cyber policies, not by deliberate market design, but because that is where the losses were pointing. The effects of that migration are visible in At-Bay's claims data today: Financial fraud is now the most common incident type in our book, accounting for roughly a third of all claims annually and growing steadily for three consecutive years.

The same migration dynamic may play out with AI-related harm as the market learns more about this emerging technology and the exposures that come with it.

[3] *Source: Business Insurance, Insurers, brokers adjust as AI exclusions emerge*
[4] *Source: Financial Times, Insurers retreat from AI cover as risk of multibillion-dollar claims mounts*

# Business Interruption in the Age of AI

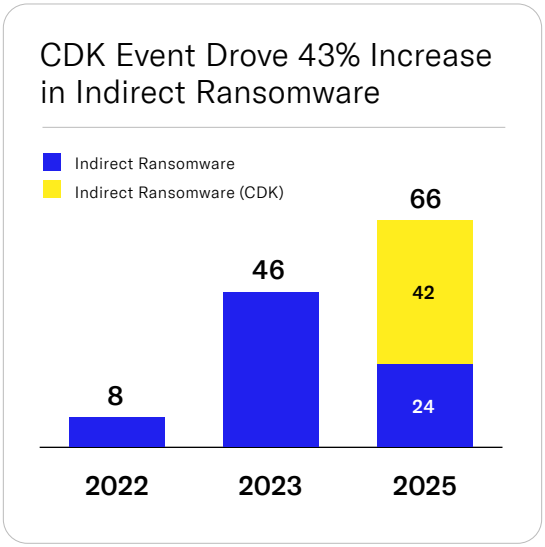
The relationship between technology reliance and business interruption is already well understood. Nearly every business has some aspect of it tied to technology now, as most businesses rely on software to perform critical operating functions.

Some industries are more reliant than others. Manufacturing companies running operational technology embedded in production lines, for example, know that a cyber outage means a production stoppage, with every idle hour translating directly into revenue loss. That exposure is priced, modeled, and understood. Famously, in 2025, a ransomware attack on Jaguar Land Rover and the subsequent shutdown of operations cost the company £50 million a week, with the carmaker posting a £485 million pre-tax loss for the quarter[5].

In June 2024, we saw a similar disruption occur at scale. Attackers shut down CDK Global, a software platform used by auto dealerships across North America to manage sales, financing, inventory, and payroll, impacting 15,000 dealer locations, which lost access to the systems their operations depended on and reverted to pen and paper for nearly two weeks. Dealerships collectively lost an estimated $1B in revenue[6].

At-Bay's 2025 InsurSec Report documented the impact of these types of events specifically. The CDK event drove a 43% increase in what we defined as indirect ransomware (a ransomware attack on a vendor or partner that resulted in damages to the insured) claim frequency across At-Bay's book in 2024, mostly due to business interruption.

Attackers understand this dynamic between technology suppliers and their customers. At-Bay's data shows that threat groups are increasingly targeting vendors with widely distributed obligations to large numbers of downstream customers. The bigger the blast radius, the faster and larger the ransom gets paid. The CDK attack illustrates how businesses that depend on technology for core business functions can become collateral damage when that technology goes down.




## Coverage Implications*

The more central AI becomes to how a business actually operates, the more a disruption to that AI looks like a full shutdown rather than an inconvenience. Outages that previously caused significant delays but in targeted, niche areas will increasingly cause more widespread work stoppages as companies put more and more AI in charge of existing processes and workflows. Businesses and brokers should review business interruption limits in the context of how operations would actually function during a sustained AI tool outage and confirm whether cyber policy contingent business interruption provisions extend to AI tool providers. Policy limits sized for a pre-AI technology profile may not reflect the actual exposure at the next renewal.

[5] Source: BBC, Jaguar Land Rover posts heavy loss after cyber-attack
[6] Source: TechTarget, The CDK Global outage: Explaining how it happened

# Copyright and IP Liability

Copyright and IP liability arise when a business uses, publishes, or distributes content that infringes on someone else's protected IP without permission, license, or a defensible fair use basis. Historically, copyright risk was managed by professionals in a content workflow. Copywriters knew not to lift protected material, creative directors flagged derivative work, and legal teams reviewed anything that looked close to the line.

When generating content required a copywriter or creative director, IP review was built into the workflow, not because the company had a formal policy but because the professional in the chain knew to do it. AI removes that professional from the equation without removing the legal obligation. Content still needs to be original, licensed, or defensibly fair use, and the business is still liable if it isn't. However, when AI is introduced, there's often no longer anyone in the workflow who knows the requirements or basics of how to check for copyright infringement.

## **Pre-AI** Content

**01 Brief**
Marketing manager defines the need (campaign, post, asset)

**02 Copywriter**
Drafts original content; knows not to lift protected material

**03 Creative Director**
Reviews; flags anything that looks borrowed or derivative

**04 Legal / IP Review**
Checks for copyright, trademark, licensing

**05 Approval**
Sign-off before anything goes live

**06 Publish**

*Each handoff is a checkpoint. The governance isn't a separate process, it's embedded in the people moving the work forward.*

## **AI-Enabled** Content

**01 Business**
Identifies a marketing need

**02 Junior Employee**
Types a prompt

**03 AI Model:**
Generates content

**04 Publish**

*There are no handoffs and therefore no governance built into the process.*

### Coverage Implications*

Copyright and IP infringement from AI-generated content is most likely to land under policies with media liability coverage. This can include cyber and professional liability policies with media or IP infringement extensions, general liability policies with advertising injury, and traditional standalone media insurance products.

# What This Means for Businesses and Brokers

The following translates this chapter's key findings into concrete next steps: what businesses should prioritize now and what brokers should be discussing with their clients.

## For Businesses

### Review your content workflow for AI touchpoints.

Any place where AI generates public-facing material, like social posts, marketing copy, product descriptions, and email campaigns, is a potential copyright exposure if there is no human review step. A clear internal policy that governs how employees use AI for content generation is both a practical control and evidence of good-faith governance if a claim arises.

One immediate step: Consider adding language to your prompts, such as "do not use any copyrighted material" or "flag if this content may infringe on existing IP", to create a documented governance step and reduce the likelihood of an obvious infringement making it to publication (though be aware that this won't eliminate risk entirely).

### Understand how AI tools are connected to your data and how the data is being used.

Connecting an AI tool to your email, CRM, or file system is a potential privacy disclosure event. Before connecting any AI tool to internal systems, ask: What data lives there? Is any of it regulated? Do we have the consent framework in place? This applies equally to manual uploads. Importing a spreadsheet of customer data into an AI tool is functionally the same risk as a formal integration. Make sure your internal AI policy for users requires the same controls and poses the same questions your IT team uses when they authorize a new connector.

### Be aware of any AI-powered customer-facing chatbots or AI-based decision models with autonomy.

If your business uses AI in any customer-facing context like chatbots, recommendation engines, wellness tools, or education platforms, or if you rely on AI to provide services or products, you may be carrying bodily injury and/or property damage exposure. Review your general liability, professional liability, cyber, and/or crime policies to understand what's covered and whether there are any AI-related exclusions. Ask your broker for guidance.

### Consider the business interruption possibilities of AI dependency for core business functions.

If your business has integrated AI into core operational workflows, not just as a productivity tool but as a mechanism through which work is actually delivered, your business interruption risk may be materially higher than your current policy reflects.

Review your business interruption limits in the context of how your operations would actually function during a sustained AI tool outage. Would your operations stop without the technology? It's also worth flagging areas of AI dependency and identifying opportunities to create resilience with non-AI backup options.

# For Brokers

## Review AI tool usage with clients to understand data privacy risks.

Cyber policy privacy liability is the primary coverage home for PII exposure, but many clients may be underinsured relative to the number of AI integrations they have in place. A review of connected tools is worth raising at renewal for any client that has materially expanded AI tool usage since the last policy period.

Review policy language carefully for AI-related exclusions. Some general liability policies now contain exclusions that could affect how certain AI-related data claims are handled, depending on how they are brought. Confirming that the cyber policy responds cleanly to AI-driven data leakage scenarios, and that no exclusionary language creates an unexpected gap, should be part of coverage review conversations.

SMBs are particularly at risk, considering they have the same exposure as enterprise clients but often none of the governance infrastructure.

## Start collecting AI governance information from clients now.

Underwriters are beginning to ask for it, and clients who can document what tools they use, what data those tools access, and what review processes they have in place will be easier to place and better covered than those who can't.

The general liability exclusion trend means clients that deploy AI in any capacity may be losing bodily injury coverage without knowing it. Review every policy at renewal (general liability, cyber, professional liability, tech E&O, and property) and check each for AI-related exclusions. Ask: If an AI-related claim came in today, which policies would respond and which would find a reason not to?

## Proactively discuss the potential for business interruption with clients.

Ask clients directly: If the AI tools or software platforms your operations depend on went down for two weeks, what would stop? What would it cost per day? For businesses actively integrating AI tools into their everyday workflow, the answers may be rapidly changing, and it could be worth stress-testing against current business interruption limits before a claim reveals the gap.

It is also worth confirming whether contingent business interruption coverage extends to AI tool providers. For clients heavily dependent on third-party AI platforms, that provision may be the most important one in the policy.

# What to Watch

CHAPTER

03

AI risk isn't a single problem to solve, it's a moving target. Attackers are gaining new capabilities as fast as defenders are, AI tools are embedding themselves into business workflows faster than governance can keep up, and the legal and underwriting frameworks meant to address the fallout are still being written in real time.

Rather than trying to predict exactly how this plays out, here are six key areas to pay attention to:

## 1. What Happens Within the Perimeter

If an attacker can reliably identify previously unknown vulnerabilities in a company's public-facing IT assets and then rapidly exploit those vulnerabilities to gain access, defenders will no longer be able to expect any level of protection from perimeter controls. As a result, the only possible countermeasure will be the ability to detect, immediately and without delay, any intruder who enters the environment. This means a high-quality MDR service composed of pervasively deployed EDR tools backed by 24/7 monitoring by professionals will be critical to preventing widespread disruption.

## 2. How the Attack Surface Will Expand in Unexpected Ways

The pervasiveness of AI means the technology is now embedded so deeply in everyday business tools that attempts to ban AI tool usage among employees are largely moot, as AI functionality is now incorporated into tools so commonplace that avoiding it entirely isn't realistic. Security and IT teams are caught between two legitimate imperatives: blocking AI integrations that introduce unacceptable risk and enabling the productivity gains that business leaders are demanding.

What makes this moment different from previous technology adoption cycles is the combination of business pressure and unsolved security risk arriving at the same time. This creates a genuine governance problem due to the lack of a clear standard for what responsible AI use looks like or the large-scale capability to track how AI is used within a business environment, though security vendors are actively attempting to close this gap.

## 3. The Evolution of AI-Powered Defenses

At this moment, attackers likely have a slight edge over defenders in the ability to leverage AI for their respective workflows. The day is rapidly approaching when the world will see the first fully automated end-to-end attack where an AI agent selects targets, performs reconnaissance, develops and deploys exploits, and performs actions on the objective. The obvious countermeasure is an AI-powered defense combining detection, triage, selection, and deployment of containment capabilities, and remediation of any impacts from the attack, but this is an effort most in-house security teams simply won't be able to deploy without some external help. Pivoting to a provider that can demonstrate proficiency at deploying AI to stop attacks at scale can help companies that can't or won't develop this capability in-house.

## 4. An Influx of Liability and Data Leakage Claims

The laws that enable liability and data leakage claims already exist, and plaintiffs' attorneys are paying close attention to the impact AI usage will have here. What's missing so far is claim volume, but that doesn't mean the risk isn't there. As more businesses connect AI tools to their internal systems without any governance in place, the number of potential claims may grow.

## 5. How Underwriting Requirements Will Change

Armilla, a Lloyd's-backed MGA dedicated to AI liability, has already built its underwriting model around documented AI governance, thereby rewarding organizations that can demonstrate controls, testing, and oversight with better terms and pricing. The parallel to cyber is direct. Just as EDR adoption and penetration testing became underwriting inputs for cyber policies, AI governance documentation could be on track to become the evidence base for AI liability coverage.

That shift is already beginning on the business side as well. Companies are becoming more deliberate about AI governance and compliance: establishing internal policies, vetting vendors, and documenting how AI is used across their operations. As that becomes standard practice, the businesses that have done the work will be meaningfully differentiated from those that haven't, both in their actual risk profile and potentially in their ability to access coverage on favorable terms.

## 6. The Development of AI-Specific Coverage and Exclusions

We've already seen first movers introduce new AI-specific coverage to the market. So far, feedback from analysts, trading partners, clients, and competitors has been mixed. We anticipate new coverage and language may continue emerging and evolving as insurers assess the opportunity to address risk. As that happens, looking at the language to understand what the coverage actually offers is critical.

# About At-Bay

At-Bay is the InsurSec provider for the digital age. By combining world-class technology with industry-leading insurance, At-Bay was designed from the ground up to empower businesses of every size to meet cyber risk head-on. At-Bay Insurance Services, LLC provides insurance protection and security prevention solutions to close to 40,000 businesses in the US, safeguarding up to $800B in collective business revenue, and offers coverage by non-admitted insurers for Cyber, Technology Errors & Omissions (Tech E&O), and Miscellaneous Professional Liability (MPL). The At-Bay Group also includes an active full-stack insurance company and a cybersecurity company. At-Bay Security, LLC ("At-Bay Security") offers proprietary cybersecurity solutions including At-Bay Stance Managed Detection & Response (MDR).

at bay

# AI Risk Checklist for Businesses

## What to do right now to reduce your exposure

AI is simultaneously making attackers more capable and creating entirely new categories of risk that existing defenses weren't built to catch. Rapid adoption of AI tools is also creating liability gaps most businesses haven't accounted for, from copyright exposure to business interruption. Use this checklist to strengthen your defenses against this evolving area of risk.

**Need help reviewing your AI risk?** At-Bay Stance™ Advisory Services can help you strengthen your defenses, respond to threats, and build resilience — included with At-Bay Cyber and Tech E&O policies at no additional cost.

Work with our Cyber Advisor team to review your security posture and get personalized advice for your unique environment and AI usage[1].

Book a meeting → View catalog →

## 1. Identify Key Roles & Responsibilities

| | |
|---|---|
| ☐ | **Review your content workflow for AI touchpoints.** Any place where AI is generating public-facing material (social posts, marketing copy, product descriptions, email campaigns, etc.) is a potential copyright exposure if there is no human review step. |
| ☐ | **Put a written AI content policy in place.** A clear internal policy that governs how employees use AI for content generation, even a simple one-page document, is both a practical control and evidence of good-faith governance if a claim arises. |
| ☐ | **Instruct your AI models directly.** Add language to your prompts, such as "do not use any copyrighted material" or "flag if this content may infringe on existing IP," to create a documented governance step, though be aware this won't eliminate the risk entirely. |
| ☐ | **Keep a human checkpoint in AI-run workflows.** A human reviewing a suspicious invoice can pause and question it; an AI agent acting autonomously just executes. |

1. Access to specified At-Bay Stance offerings, including Stance Advisory Services, is available to insureds with Cyber and Tech E&O policies placed through At-Bay Insurance Services, LLC, via the policy's Embedded Security Fee and corresponding Endorsement. Please contact your authorized insurance representative for information concerning your policy.

## 2. Check Your Security Controls

| | |
|---|---|
| ☐ | **Adopt AI-powered email security.** AI-powered email security analyzes patterns in sender behavior, email structure, and domain reputation that legacy email security solutions weren't built to catch. |
| ☐ | **Measure your actual time-to-patch.** Vulnerability exploitation windows have collapsed from weeks to hours, so a cadence that once looked adequate may now leave a real gap between disclosure and fix. |
| ☐ | **Determine where Managed Detection and Response (MDR) fits into your security stack.** Insurers are placing growing weight on security solutions like MDR, alongside traditional risk assessment, for AI-driven threats. Businesses with monitoring and response already in place are likely to be better positioned for coverage and pricing. |

**Looking to improve your security posture?** At-Bay Stance MDR delivers enterprise-grade security[2] with 24/7 monitoring, expert-led response, and full remediation, at a fraction of the cost of an in-house security operations center. Adopting Stance MDR may unlock premium credits[3] on an insurance policy placed through At-Bay, plus enhanced ransomware and financial fraud limits. Contact your authorized insurance representative for details.

## 3. Take Inventory of Your AI Tool Integrations

| | |
|---|---|
| ☐ | **Ask the right questions before connecting any AI tool.** Before connecting any AI tool to internal systems, ask: What data lives there? Is any of it regulated? Do we have the consent framework in place? |
| ☐ | **Understand your AI agent permissions.** Connecting AI to a CRM, email, or cloud systems requires broad permissions, and attackers exploit that through legitimate, trusted credentials invisible to normal monitoring. Understand the tradeoffs. |
| ☐ | **Apply the same standard to manual uploads as you do to integrations.** Downloading a spreadsheet of customer data and pasting it into an AI tool is functionally the same risk as a formal integration. Ensure your team understands what is on or off limits. |
| ☐ | **Treat AI agents as a target for social engineering.** Hiding malicious instructions in a webpage, email, or document can cause an AI agent to execute them instead of just reading them, because AI can't reliably tell the difference between information and commands. |

2. At-Bay Stance MDR is provided by and available for purchase separately through At-Bay Security, LLC, a wholly owned subsidiary of At-Bay, Inc. Stance MDR is not required for insurance coverage through At-Bay, and it is not limited to At-Bay policyholders. Similar credits and enhancements available for comparable solutions.
3. Qualified policyholders who implement an approved MDR solution may be eligible for a premium credit. Please contact your authorized insurance representative for additional information concerning potential cost savings. Any premium savings will be based on the specific risk profile of the business and qualifying conditions.

## 4. Understand Your Policies' AI Coverage

| | |
|---|---|
| ☐ | **Send your complete policy stack, not just your cyber policy, to your broker for review.** AI-related exclusions are starting to appear in places like general liability, and a gap there won't surface unless someone reads across the full portfolio. |
| ☐ | **Understand how each line of coverage would treat an AI-related claim.** This includes general liability, cyber, professional liability, tech E&O, and property. |

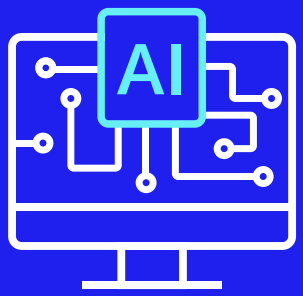
## 5. Know Your AI Business Interruption Exposure

| | |
|---|---|
| ☐ | **Determine what actually breaks if an AI platform you depend on goes down for two weeks**, and put a real dollar figure on the daily cost. That number, not a guess, should set your business interruption limit. |
| ☐ | **Ask your broker if your contingent coverage includes the AI vendors you rely on.** If a meaningful part of your operation runs through a third-party AI platform, it's important to confirm this before an incident. |

at bay

# AI Risk Checklist for Brokers

## How to advise clients in this emerging area of risk




Every major technology wave creates new liability before the legal and insurance frameworks catch up, and AI adoption is already faster and more pervasive than anything that came before it. Use this checklist to review coverage and document governance with your clients before a claim forces the conversation.

**Before you talk to your client:** Send the companion AI Risk Checklist for Businesses. Most of what's on this list — governance documentation, tool inventories, security posture — lives with the client, so have them work through their checklist in parallel for a productive conversation.

## 1. Talk to Clients About Their AI Governance

- ☐ **Ask clients about their AI content review process.** Cyber underwriters are starting to treat AI governance as an underwriting criteria, the same way they treat EDR adoption or MFA. Being able to show documented review processes may help navigate those discussions.
- ☐ **Start collecting AI governance documentation from clients now.** Clients who can document what tools they use, what data those tools access, and what review processes they have in place will be better prepared for any related conversations.

## 2. Flag Security Controls Becoming Underwriting Baselines

- ☐ **Ask clients if their email security can spot unusual or suspicious behavior, not just known threats.** AI lets scammers write convincing fraud emails fast, in any language, without any special skill. Clients using older, signature-based email tools may be more exposed to fraud than they think, and fixing that can help them qualify for better coverage.
- ☐ **Ask clients how fast they can patch systems when a new vulnerability hits.** Attackers now exploit new vulnerabilities within hours, not weeks. Clients with slow or ad hoc patching are carrying more risk than their controls suggest.
- ☐ **Start collecting information on clients' MDR posture now.** Insurers are placing growing weight on security solutions like MDR, alongside traditional risk assessment, when it comes to AI-driven threats. Clients who can show MDR in place now will be better positioned for coverage and pricing later.

## 3. Review AI Tool Integrations at Renewal

| | |
|---|---|
| ☐ | **Review clients' AI tool integrations.** Many clients may have new or increased exposure due to the number of AI integrations they have in place. A review of connected tools is worth raising at renewal for any client that has materially expanded AI tool usage. |
| ☐ | **Prioritize SMB clients for AI risk advising.** SMBs are the highest-risk segment here. They have the same exposure as enterprise clients but often less governance infrastructure. |

**Clients looking to improve their security posture?** At-Bay Stance™ MDR delivers enterprise-grade security[1] with 24/7 monitoring, expert-led response, and full remediation, at a fraction of the cost of an in-house security operations center. Adopting Stance MDR may unlock premium credits[2] on an insurance policy placed through At-Bay, plus enhanced ransomware and financial fraud limits.

## 4. Audit Policies for AI Exclusions

| | |
|---|---|
| ☐ | **Review policy language carefully for AI-related exclusions.** Some general liability policies now contain exclusions that could affect how certain AI-related claims are handled. |
| ☐ | **Check every policy at renewal.** Review general liability, cyber, professional liability, tech E&O, and property. Ask: If an AI-related claim came in today, which policies would respond and which would find a reason not to? |

## 5. Stress-Test Clients' Business Interruption Coverage

| | |
|---|---|
| ☐ | **Ask clients the two-week AI outage question.** If the AI tools or software platforms your client's operations depend on went down for two weeks, what would stop? What would it cost per day? Make sure clients are equipped with appropriate business interruption coverage based on their risk. |
| ☐ | **Confirm whether contingent business interruption coverage extends to AI vendors.** Many clients will not know whether they have it. For clients heavily dependent on third-party AI platforms, it's an important question to ask. |

1. At-Bay Stance MDR is provided by and available for purchase separately through At-Bay Security, LLC, a wholly owned subsidiary of At-Bay, Inc. Stance MDR is not required for insurance coverage through At-Bay, and it is not limited to At-Bay policyholders. Similar credits and enhancements available for comparable solutions.
2. Qualified policyholders who implement an approved MDR solution may be eligible for a premium credit. Please contact your authorized insurance representative for additional information concerning potential cost savings. Any premium savings will be based on the specific risk profile of the business and qualifying conditions.