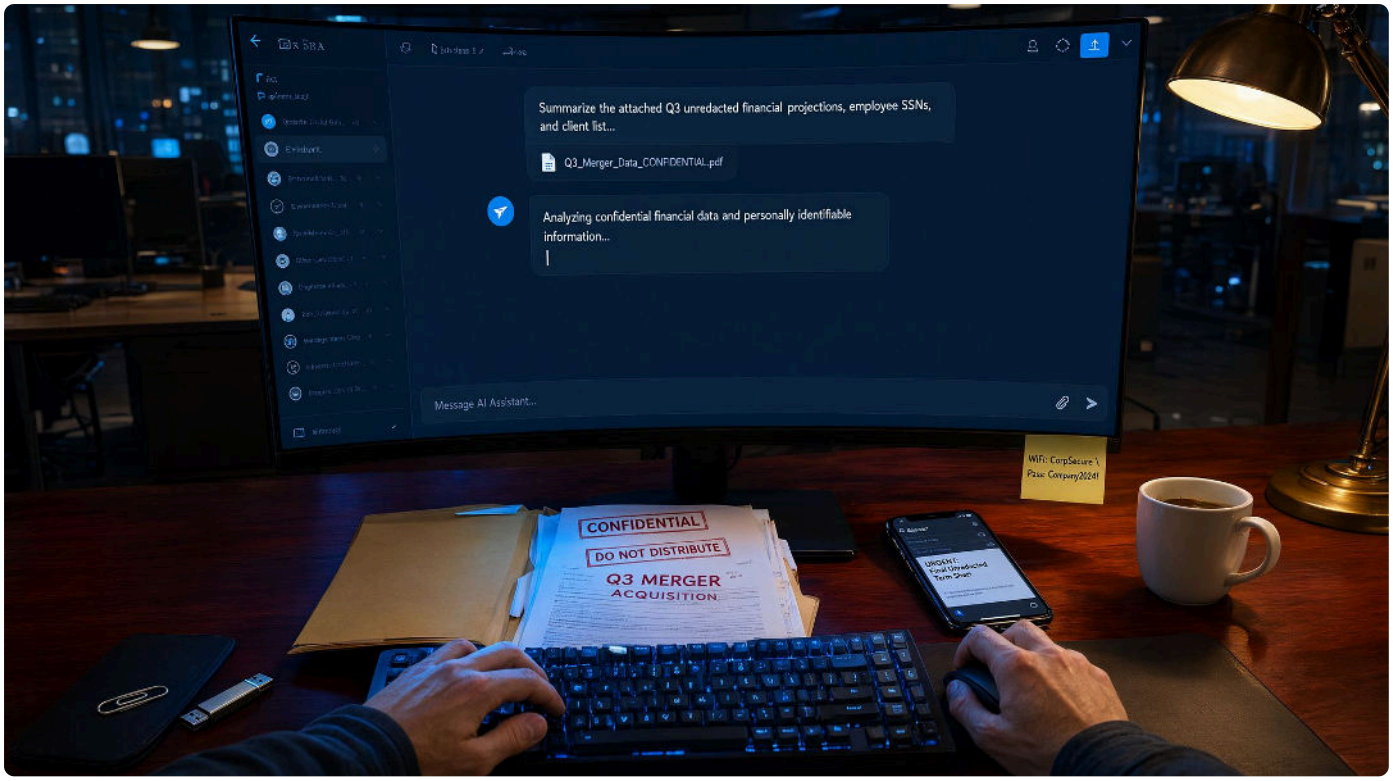
An Employee Put Confidential Data Into ChatGPT, Claude, Gemini, or Copilot. Now What?

Tom Dunlap is a founding partner at Dunlap Bennett & Ludwig, a veteran-owned law firm with over 75 attorneys nationwide. As a former prosecutor and Army officer with extensive litigation experience, Tom has secured notable victories, including a $31 million patent jury verdict and multiple Federal Circuit appeals. His expertise spans government contracts, trademarks, patents, and general intellectual property law, currently serving as an outside counsel for US HHS and the Commonwealth of Virginia. A true renaissance professional, Tom holds a JD, MBA, and an MS in Biotechnology and is an inventor with multiple patents, including technology used in COVID testing. Recognized in the WTR 1000 as one of the world's leading trademark lawyers and a Virginia Super Lawyer for over a decade, Tom combines practical experience as a former CEO and founder of Ceres Nanosciences with academic contributions as an adjunct professor at Washington & Lee University and Lafayette College, where he teaches intellectual property law. Outside of his legal practice, Tom is a PADI Master Diver, instrument pilot, saber fencer, and former Burmese National Kickboxing Silver Medalist.

Tom Dunlap, Partner





An employee pasted something into a chatbot that should not have gone there. The call usually arrives with the word “breach” already in it. Most of the time it is not one. Whether it turns out to be a breach, a contract problem, an HR conversation, or nothing at all depends on five facts, and none of them is the name of the product.

Before you get to the five facts, do three things in the first hour. Capture the whole conversation, the output, and any uploaded files, and do not let anyone delete it until that record exists. If a shareable link was created, turn it off in the tool’s settings, because public links get indexed by search engines. And if a password, API key, token, or private key went in, rotate it now. Containment comes before analysis when a live credential is involved, and that is the only part of this process where speed beats accuracy.

Then work the five facts: what went in, which product and which account, what the service could do with it, what rules already attached to it, and where the output went. The rest of this piece walks through them, then through the four questions every client asks next. Does “no training” mean no problem? Did it go public? Do we have to report it? Can we delete it?

Is This Actually a Problem?

Sometimes. Here is how we tell.

What went in

“Confidential” covers everything from a meeting agenda to a customer database to the formula the company was built on. Those are not the same event. The categories that change the analysis are personal information about customers or employees (health, financial, biometric, and



children's data most of all), working credentials, trade secrets and source code, attorney-client communications, anything received under an NDA, anything tied to a pending deal or dispute, and export-controlled or government-restricted information.

The label on the document matters less than the contents. A file stamped CONFIDENTIAL may hold nothing sensitive. An unlabeled spreadsheet may hold 40,000 customer records. Read the file, not the header.

Two categories get their own rules. Credentials are a security event first and a confidentiality event second; revoking the key matters more than debating whether the prompt will train a model. Privileged material carries a waiver risk the others do not, because disclosure to a third party outside the privilege can waive it. Whether a vendor that contractually promises confidentiality counts as a third party for waiver purposes has not been decided by any court I know of, and I would not want a client's privilege to be the test case.

Government information, meaning CUI, export-controlled technical data, and anything a federal contract restricts, runs on its own reporting clocks. Our companion piece for federal contractors, **Can Federal Contractors Use ChatGPT and Other Generative AI Tools?**, covers that.

Which product, and which account

ChatGPT, Claude, Gemini, and Copilot are each several products with different terms. The account, plan, workspace, admin settings, and enabled features decide most of the analysis. A company email address proves nothing. An employee can open a consumer plan with a work email in ninety seconds. The reverse is also true: the same employee might have been inside a company-managed workspace with negotiated terms, retention controls, and training turned off. Find out which before anyone drafts a notice.

What the service could do with it

Processing a prompt to produce an answer is not publication. The question is the actual data path under the configuration that was used. Was the conversation retained, used to improve the model, read by a human reviewer, submitted as feedback, shared by link, sent to a connected app, stored in the assistant's memory, or copied somewhere else? Human review is real. Google's consumer **Gemini privacy notice** says a subset of chats may be read by human reviewers, including trained service providers, and that reviewed data is disconnected from the account and kept for three years. Whether that could have happened to your employee's chat depends entirely on the product and the settings.

What rules already attached to it

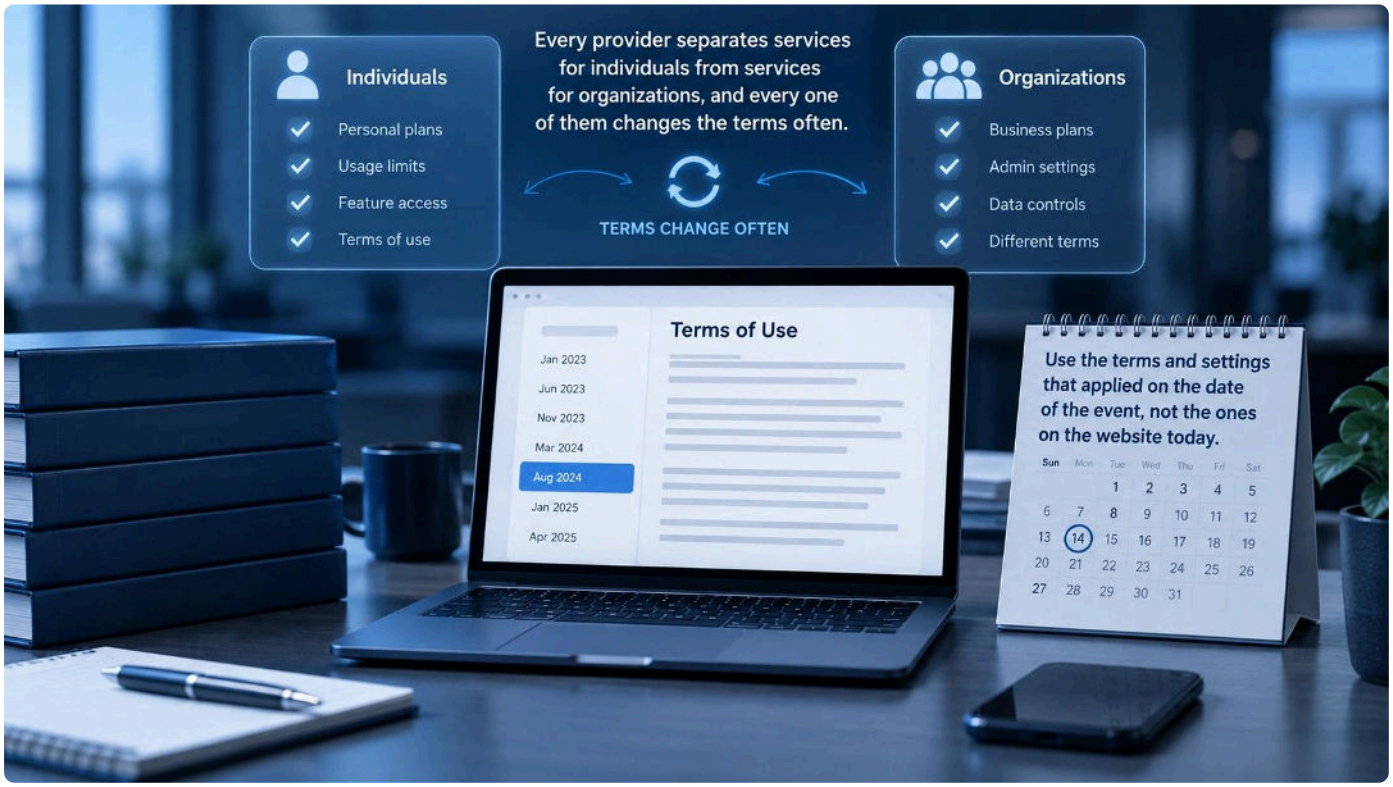
An interaction can be technically contained and still breach a contract. A customer agreement may name approved systems or subprocessors. An NDA may bar third-party access. A privacy statute may govern that class of personal data. Trade-secret status under the Defend Trade Secrets Act, 18 U.S.C. § 1839(3), depends in part on the company taking reasonable measures to keep the information secret, and a habit of pasting it into consumer chatbots is exactly the fact opposing counsel goes looking for.

This fourth question is the one that most often turns a policy violation into an external obligation, or confirms there is none.

What came out, and where it went

This is the question most companies skip, and it produces the most expensive mistakes. Output that stayed a draft and got checked is one thing. Output that went into a customer proposal, a regulatory filing, a code repository, or a hiring decision without a human reading it is another. An invented citation, a wrong number, or a stretch of someone else's copyrighted text in a deliverable is the company's problem, not the vendor's. Trace where the output landed before you close the file.

Why the Specific AI Service Matters



Every provider separates services for individuals from services for organizations, and every one of them changes the terms often. Use the terms and settings that applied on the date of the event, not the ones on the website today.

ChatGPT

OpenAI may use content from its individual services (Free, Go, Plus, Pro) to train its models unless the user turns off "Improve the model for everyone" under Data Controls. Temporary Chats are not used for training and are deleted within 30 days. ChatGPT Business (the plan formerly called Team), Enterprise, Edu, and the API are not used for training by default. See OpenAI's explanations of **individual and business data use** and **business data protections**. The plan is one variable. Custom GPTs, apps, connectors, shared links, memory, and agent features each add a data flow the plan does not describe.

Claude

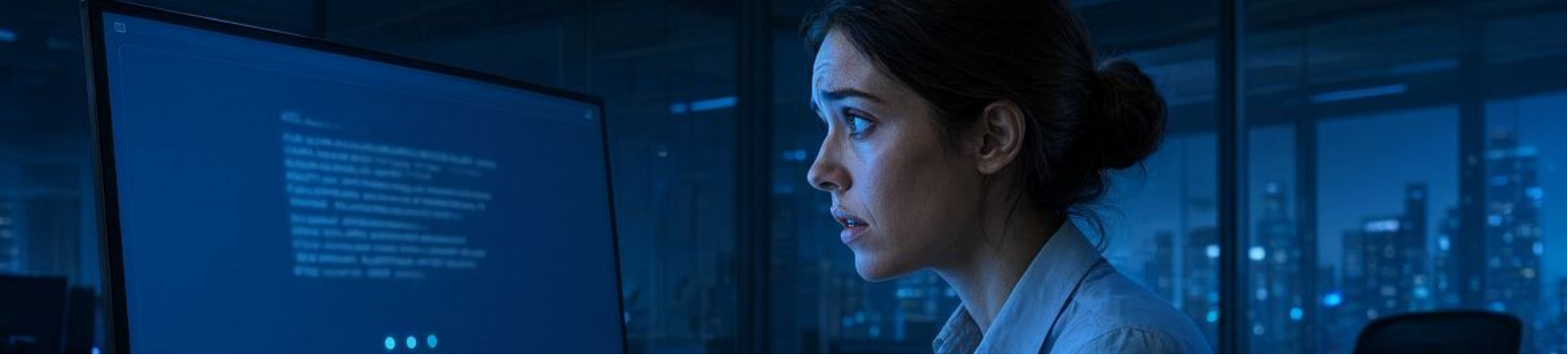
Anthropic runs two regimes. Consumer plans (Free, Pro, Max) have used chats and coding sessions to train models by default since a terms change announced August 28, 2025, with existing users given until October 8, 2025 to choose. Retention is up to five years for users who allow training and 30 days for those who opt out. Commercial products (Team, Enterprise, the API, and access through cloud platforms) are not used for training by default and carry a 30-day standard retention. Feedback submitted from any account is stored separately and longer. See Anthropic's pages for **commercial accounts** and **consumer accounts**.

Gemini

Gemini reaches employees three ways: a personal Google account, an "additional service" on some work accounts, and a core service under qualifying Workspace editions with enterprise-grade data protections. For that last group, Google states that prompts, uploaded files, and responses are not human-reviewed or used to train generative AI models outside the customer's domain, and administrators set retention anywhere from 90 days to indefinite. For personal accounts, the "Keep Activity" setting (formerly Gemini Apps Activity) controls whether chats are saved, used for training, and sampled for human review. The work email address tells you nothing; the license does. Google's **Workspace account comparison** and **Gemini Apps Privacy Hub** are the places to check, and connected Google and third-party services process what they receive under their own policies.

Copilot

"Copilot" is three products. Consumer Microsoft Copilot has its own history, memory, and training controls. Microsoft 365 Copilot and Copilot Chat under enterprise data protection do not use prompts, responses, or Microsoft Graph data to train foundation models, but they log every prompt and response as interaction data that administrators can search, retain, and produce through Purview. GitHub Copilot has separate terms for prompts, suggestions, code snippets, and telemetry. If source code went in, find out which Copilot before doing anything else. Microsoft's guidance on **data, privacy, and security for organizational Copilot** is the starting point.



Does “No Model Training” Settle It?

No, though it removes the loudest worry. Enterprise providers still retain prompts and outputs for history, audit, and admin search, which means the information now exists in a system your customer may never have approved. A connector or agent may have handed it to a service with different terms. And the restrictions that matter most usually sit outside the provider contract altogether: the customer who was promised its data would stay in named systems, in a named region, with a named list of vendors. “No training” answers one question out of five.

Did the Information Become Public?

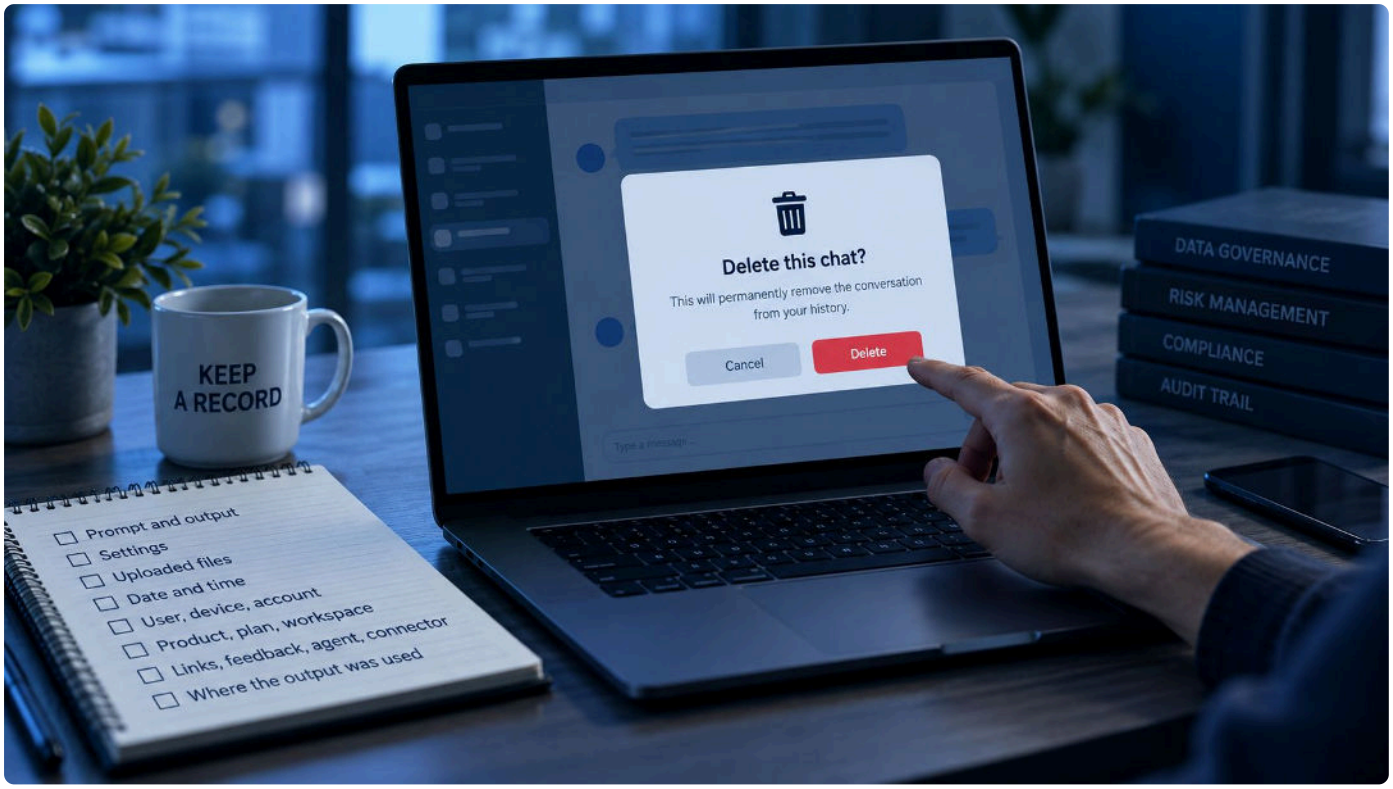
Almost never, by itself. A prompt does not post to the internet or become visible to other users. Public exposure becomes plausible when the employee creates a shared link, publishes an AI-built app, routes the prompt through a third-party agent, or pastes the output somewhere public, or when there is evidence someone actually accessed it. Absent those facts, the accurate statement is that a provider or a third party may have had access, and that is the statement to investigate. Telling a customer the data “went public” when it did not creates a second problem on top of the first.

Is It a Reportable Data Breach?

Not automatically, and the definition you are working under decides it. State breach statutes generally key on unauthorized acquisition of defined categories of personal information, many carry a risk-of-harm exception, and several run 30-, 45-, or 60-day notice clocks. A customer contract may define “security incident” far more broadly than any statute, with a 24- or 72-hour notice term attached. A cyber insurance policy may require notice as a condition of coverage. So the analysis runs through the type and volume of data, whose data it is, the governing jurisdictions and sector rules, the contractual definitions, the service and settings, whether anyone outside could reach it, and what the company did after it found out. Notice recipients can include individuals, customers, regulators, insurers, and contracting parties, and the clock starts at discovery, not at the end of the investigation.

Foreign law can apply even without public exposure. Under the **GDPR**, a vendor processing personal data on your behalf needs Article 28 processor terms, usually a data processing agreement, and a transfer outside the EEA needs a Chapter V mechanism. A U.S. provider is not automatically unlawful, and a DPA alone does not settle the transfer question. California’s **CCPA**

runs on a different logic that turns on whether the provider qualifies as a service provider under a compliant contract and what purposes were disclosed to consumers. The EU AI Act is a separate statute again; its Article 4 **AI literacy** duty has applied since February 2, 2025.



Should You Delete the Chat?

Eventually, probably. First, no. The prompt and the output are the best evidence of what went in, which settings applied, whether the response reproduced the data, and whether it went anywhere else. Document the complete exchange, any uploaded or connected files, the date, time, user, device, and account, the product, plan, workspace, and settings, any link, feedback, agent, connector, or plug-in involved, and every place the output was later used.

Then decide what deletion means. Deleting the visible chat may leave files, memory, feedback copies, connected services, shared links, and administrative retention untouched, and it never reaches the provider's back end on your schedule. OpenAI spent May through September 2025 under a court order in the New York Times copyright litigation preserving every consumer ChatGPT log, deleted chats included, over its own objection. The order was lifted going forward on September 26, 2025, but the logs preserved under it stayed available to the plaintiffs, who were later allowed to sample 20 million conversations. Your delete button does not reach a litigation hold on somebody else's docket.

Your own litigation hold matters too. If a dispute is pending or reasonably anticipated, the conversation and related logs are electronically stored information under Rule 37(e), and deleting them can cost more than the disclosure did.

What Happens Next

Three tiers cover nearly every case I have seen, and the table is the short version of the three paragraphs that follow it.

Tier	What you are looking at	What to do
Closes in a day	Approved enterprise account with training off. Ordinary internal material. No link, connector, or agent. Output stayed a draft or was checked before use.	Document the event, confirm the settings, remind the employee, close the file.
Takes a week	Personal or unapproved account, or nobody can say what the settings were. Proprietary code or strategy material. A contract that names approved systems or subprocessors. Output used without review.	Trace the actual data path. Read the contracts that govern the information. Decide whether anyone outside the company gets a call.
Counsel the same day	Live credentials or keys. Health, financial, or children's data. A public link. A central trade secret. Evidence of outside access or forwarding. A contract or regulator with a short clock. A consequential decision already made on output that looks wrong.	Rotate the credential and kill the link now. Preserve the record. Assess statutory, contractual, and insurance notice duties. Involve counsel. Keep the team narrow: security and the system owner, or legal and the account lead.

Most of these close in a day. Approved enterprise account, training off, ordinary internal material, no link or connector. Document it, confirm the settings, remind the employee, close the file. A generic memo in an approved tool should not produce the same response as an exposed private key, and a program that treats them the same trains employees to report nothing.

The middle cases take a week. A personal account, or one where nobody can say what the settings were. Proprietary code or strategy material. A contract that names approved systems. These call for tracing the actual data path, reading the contracts that govern the information, and a decision about whether anyone outside the company gets a call.

A small number need counsel the same day. Live credentials or keys. Health, financial, or children's data. A public link. A central trade secret. Evidence that someone outside accessed or forwarded the material. A contract or regulator with a short clock. A consequential decision already made on output that looks wrong. Even then the response stays narrow. A credential problem needs the security team and the system owner. A customer-contract problem needs legal and the account lead. The rest of the company does not need to be in the room.

Fix the Program

If it took your company three days to figure out which Copilot the employee used, that is the finding. The event was the symptom.

Employees use personal AI accounts because the company has not given them an approved tool that does the work. That is the pattern in nearly every one of these I have handled. A prohibition without a provided alternative gets violated within a month, and the fix costs money: an enterprise plan, **data loss prevention** rules that catch sensitive text and file uploads to unapproved services, browser controls, and somebody whose job includes reading provider terms when they change, which is often.

The program that holds up is short. An inventory of approved products, plans, agents, and connectors. A clear line between consumer and enterprise services. Data-classification rules that say which classes may go into which tools. Controls on connectors, sharing, memory, feedback, and web search. Training on the scenarios that actually happen. A human-approval rule for consequential outputs. A no-blame way to report a mistake within the hour rather than the week. And a triage process that separates the memo-in-an-approved-tool case from the private-key case, so the second one gets the attention. The **NIST AI Risk Management Framework** describes govern, map, measure, and manage as continuing functions rather than a document, and that is the right model. A policy tells people what you expect. A process tells you in an hour whether this one matters.

Where This Leaves You

The product name does not answer the question. Five facts do: what went in, which account, what the service could do with it, what rules attached, and where the output went. Most of the time the answer is document it and move on. Some of the time it is rotate the key, kill the link, call the customer, and check the insurance policy, today. The point of having a process is knowing which one you are looking at before lunch.

If you want help building that process, or you are looking at one of the same-day cases right now, contact us.

This article is provided for general informational purposes only and does not constitute legal advice. Reading it does not create an attorney-client relationship. Laws, product features, and provider terms change, and readers should evaluate current requirements and their own circumstances.



Dunlap Bennett & Ludwig, a veteran-founded national firm, advises organizations on AI governance, privacy, intellectual property, commercial agreements, internal investigations, and disputes involving emerging technologies. The significance of an AI-related disclosure depends on the specific facts, governing law, and applicable contracts.

To learn more about Dunlap Bennett & Ludwig, call **888-306-4030** or email **clientservices@dbllawyers.com**.

Attorney Advertising. Prior results do not guarantee a similar outcome.