LangProtect

# Stop Choosing Between AI Productivity and Data Security

How Langprotect’s secure sensitive data in every employee AI prompt without interrupts the user experience

# THE PROBLEM EVERY SECURITY TEAM IS GETTING WRONG

The question most security teams are asking is the wrong one.

"Should we allow employees to use AI tools?" That decision is already made. Your employees are using ChatGPT, Claude, Gemini, and a dozen tools IT has never reviewed, not because they are careless, but because the tools work. They write faster, think clearer, and close tasks that used to take hours in minutes. That productivity is real, and no acceptable use policy is going to undo it.

The right question is this: when your employee presses Enter on a prompt that contains a patient record, a client contract, or a set of financial figures, where does that data go? Who has it? Can you prove it never left?

For most organizations today, the honest answer is no.

The gap is not in awareness. Security teams know AI tools are in use. The gap is in infrastructure. Knowing something is happening and having the technical controls to govern it are two different problems, and most organizations have solved only the first one.

> 83% of organizations have no automated controls over what employees send to public AI tools. They rely on training sessions, email warnings, or acceptable use policies to govern interactions that happen thousands of times a day, in real time, in a browser window no existing security tool can see.
>
> Source: Kiteworks AI Data Security and Compliance Risk Study, 2025

The result is a specific and growing exposure problem. Sensitive data (patient records, client files, source code, financial figures) is flowing into public LLM endpoints continuously, at the speed of a keyboard, with no technical control in place to inspect it, stop it, or record that it happened.

This is not a prediction. It is the current state of enterprise AI adoption. And the organizations that respond by reaching for their existing security stack, DLP, CASB, endpoint controls, will find that none of those tools were built for this surface.

The prompt is where the data leaves. And until now, the prompt has been invisible.

## Why the Two Most Common Responses to AI Security Both Fail

Security teams responding to ungoverned employee AI usage almost always reach for one of two responses:

- Enforcing a blanket block on AI platforms across the organization
- Or surfacing a warning notification when sensitive data is detected in a prompt and asking the employee to decide what happens next.

Both approaches are well-intentioned and both are insufficient, not because the intent behind them is wrong, but because neither one addresses the problem at the layer where it actually exists.

### Enforcing a Blanket Block on AI Platforms

When an AI platform is identified as a data risk, IT teams typically enforce access restrictions across the organization, blocking tools like ChatGPT, Gemini, Perplexity, and other consumer AI platforms at the network level and communicating the restriction to employees through a company-wide policy update or email notification.

### What Actually Happens After the Block

When approved tools become too restrictive for daily work, employees don't stop using AI, they switch to personal accounts and consumer tools that bypass all security measures entirely.

Employees who were previously accessing AI platforms on managed work devices begin using the same tools through personal mobile devices, home networks, and browser profiles that operate completely outside the organization's monitoring infrastructure.

More than 80% of workers, including nearly 90% of security professionals, use unapproved AI tools in their jobs.

### The Real Impact on the Organization

- **Productivity loss is immediate**: employees lose access to tools that were accelerating their daily work, with no sanctioned alternative in place
- **Shadow AI usage increases**: the same behavior moves to ungoverned devices and platforms that carry significantly higher data risk
- **Visibility drops**: the security team ends up with less insight into what is happening than before the block was implemented

> Enforcing a blanket block on AI platforms does not reduce data exposure; it reduces the organization's visibility into data exposure that continues regardless. The behavior persists. The oversight disappears.

## Warning Employees When Sensitive Data Is Detected

The second approach is more targeted. Organizations deploy controls that monitor employee prompts in real time and surface a warning notification when sensitive content is identified; presenting the employee with the option to cancel the prompt, revise it, or proceed with the submission anyway. In principle, this creates a meaningful intervention point between the employee and the risk. In practice, it transfers the compliance decision to the person least equipped to make it, at the moment they are least inclined to stop and think carefully about it.

### Why Employees Click Continue Anyway

The employee is in the middle of a task and working against a deadline when the popup appears. The warning asks them to assess whether their prompt violates a data policy in a moment where stopping to evaluate that question carefully is simply not something their workflow allows.

The path of least resistance is to click continue, and that is precisely what most employees do, not because they are careless, but because the system placed the compliance burden on the wrong person at the wrong moment.

### What the Security Team Is Left With

- **A WARN CONTINUED log entry**: a timestamped record of sensitive data reaching an external LLM after the only control in place was bypassed
- **No technical prevention**: the data transmitted regardless of the warning that appeared
- **A compounding problem**: warning fatigue sets in within days, the notification becomes invisible, and the click becomes automatic across the entire employee population

> Every **WARN CONTINUED** entry in an audit log is not evidence of a governance process working. It is a timestamped record of sensitive data reaching an external LLM after the only control in place was explicitly bypassed.

## Both Approaches Share the Same Fundamental Flaw

Enforcing a blanket block breaks the user experience without eliminating the risk; it relocates the behavior outside the organization's visibility.

Relying on employee warnings interrupts the workflow without protecting the data, it delegates the compliance decision to the person least positioned to make it correctly, repeatedly, under time pressure.

Organizations cannot depend on individuals to make correct security decisions every time; safe behavior needs to become the easiest behavior, built into the system rather than bolted on top of it.

Neither outcome is acceptable, and both are entirely avoidable if the sensitive data is handled automatically at the moment of submission, before the employee is ever asked to make a decision.

## What Smart Redact Is and How It Works

Smart Redact is LangProtect's enforcement mode, and the only enforcement approach that protects sensitive data at the prompt level without creating any interruption in the employee's workflow. When Smart Redact is active across the organization, employees continue using ChatGPT, Gemini, Copilot, and any other AI tool exactly as they normally would.

The protection happens entirely between the moment they press Enter and the moment the response appears on their screen, in a process that takes milliseconds and produces no visible indication that anything was intercepted at all.

### Enforcing a Blanket Block on AI Platforms

01

#### The Employee Writes a Prompt

An employee opens ChatGPT and writes a prompt that includes sensitive information: a patient name, a medical record number, a date of birth, a financial balance, a client address. They are not thinking about data classification. They are doing their job, and the AI tool is helping them do it faster. They press Enter.

02

#### LangProtect Intercepts Before Anything Leaves

Before the prompt is transmitted to any external system, LangProtect's Smart Redact intercepts it at the browser layer. Every sensitive entity in the prompt is identified by the PHI, PII, and PCI scanners, covering names, medical record numbers, phone numbers, addresses, insurance identifiers, financial figures, and every other classified data type, and replaced automatically with typed token placeholders. The employee's original prompt is never transmitted anywhere in its original form.

**What the employee typed:**
"Draft a follow-up for patient Aaron Abbott,
DOB 1984-03-12,
MRN 7765102, outstanding balance $4,665.20, contact number 987-654-3210."

**What the LLM received:**
"Draft a follow-up for patient [PERSON_1], DOB [DATE_1], MRN [MRN_1], outstanding balance [FINANCIAL_1], contact number [PHONE_1]."

**PHI transmitted to LLM:** None.
**Employee workflow interrupted:** Zero times.

## 03 Two Things Happen Simultaneously

At the same moment the masked prompt is sent to the LLM, the real sensitive values are routed directly to a secure, encrypted database, bypassing the LLM entirely. In standard deployments this database handles exact data retrieval through SQL queries, and in RAG-based deployments through Vector search, returning precise records about the patient, client, or account in question.

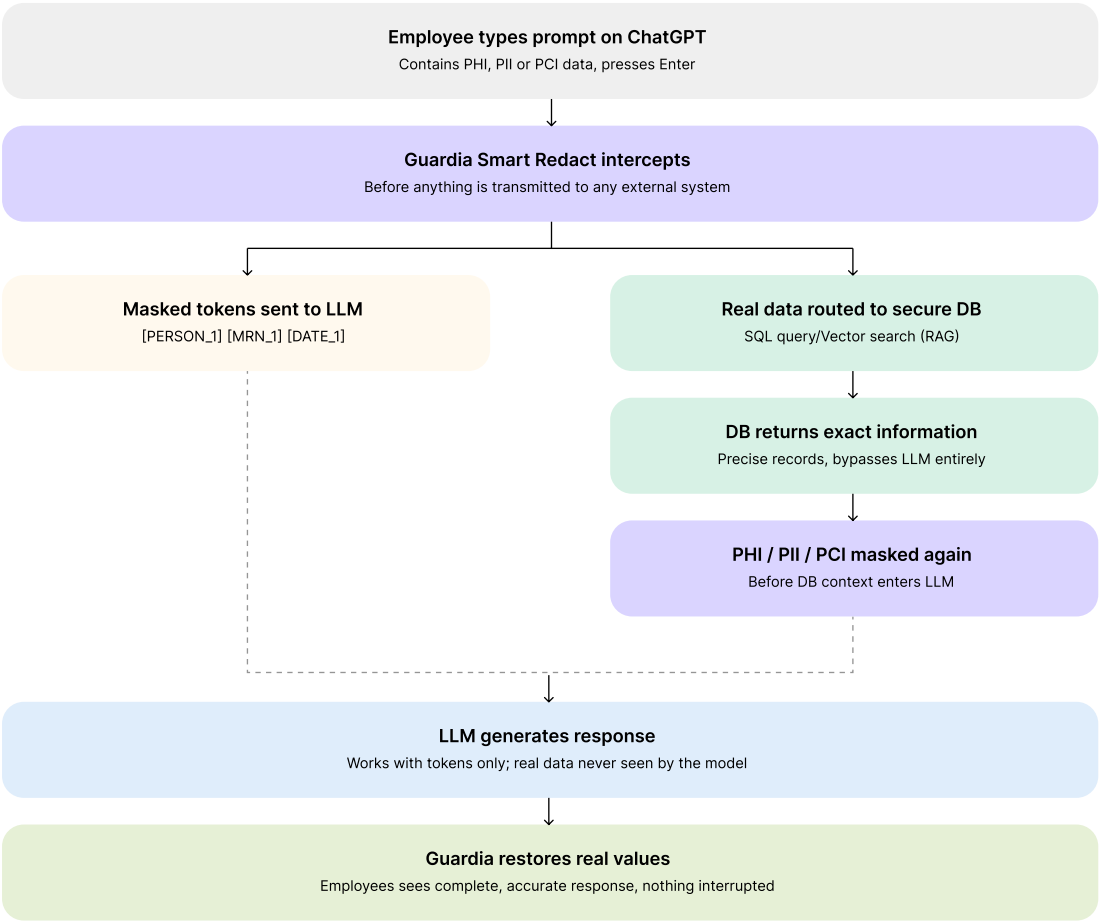
This is the step that makes Smart Redact fundamentally different from basic token masking; the AI response is not degraded or incomplete because the accurate data context is retrieved through a secure channel that the LLM never has access to.

When the database returns the exact information, LangProtect's PHI, PII, and PCI scanners are applied again to that response before it re-enters the LLM's context.

Any sensitive fields in the returned records; names, MRNs, phone numbers, addresses, are masked a second time. The LLM receives everything it needs to generate a complete, contextually accurate response without being exposed to the real sensitive values at any point.

## 04 The Response Is Restored for the Employee

When the LLM returns its response containing token placeholders, LangProtect intercepts it before it is rendered in the browser and restores every token to its original real value, placing each piece of information back in the exact position it appeared in the response. The employee sees a complete, fully accurate answer on their screen. The AI tool behaves exactly as they expect it to. Nothing was blocked, nothing was degraded, and the sensitive data never left the organization's control at any point in the interaction.



## What Smart Redact Protects

Smart Redact enforces protection across every major sensitive data classification that enterprise organizations are required to govern. The PHI scanner covers all 18 HIPAA Safe Harbor identifiers including patient names, dates of birth, medical record numbers, and insurance information. The PII scanner covers personal identifiers including full names, email addresses, phone numbers, physical addresses, and government identification numbers. The PCI scanner covers payment card data, account numbers, and associated financial identifiers.

- **PHI**: Patient names, date of birth, MRN, insurance plan numbers, clinical identifiers
- **PII**: Full names, email addresses, phone numbers, physical addresses, national ID numbers
- **PCI**: Payment card numbers, account data, financial balances, transaction identifiers

# The Semantic Gap

Legacy security architectures suffer from a critical "SaaS Security Gap." Traditional Data Loss Prevention (DLP), Cloud Access Security Brokers (CASB), and network firewalls were designed for a world of structured data fields and static file movement. They are structurally incapable of navigating the fluid, conversational, and often adversarial nature of generative AI.

**Healthcare billing team member: ChatGPT**

Drafting a patient follow-up with account and clinical details

### SCENARIO B

Without Smart Redact

WHAT THE EMPLOYEE TYPES

"Draft a follow-up for patient Aaron Abbott, DOB 1984-03-12, MRN 7765102, balance $4,665.20, Contact 987-654-3210"

WHAT HAPPENS

01 Employee presses Enter. **No interception**. Prompt transmits immediately.

02 ChatGPT receives the full prompt. **Real patient name, MRN, DOB, balance and phone number** are now on external LLM with no HIPAA BAA in place.

03 LLM generates a response using the real PHI. Employee receives the draft.

04 **No record exists.** The interaction happened in a browser session no security tool captured.

OUTCOME

| | |
|---|---|
| PHI transmitted | 5 identifiers exposed |
| HIPAA BAA in place | None |
| Audit record | Does not exist |
| Employee experience | Does not exist |
| Compliance posture | Violation risk |

### SCENARIO B

With Smart Redact

WHAT THE EMPLOYEE TYPES

"Draft a follow-up for patient Aaron Abbott, DOB 1984-03-12, MRN 7765102, balance $4,665.20, Contact 987-654-3210"

WHAT HAPPENS

01 Employee presses Enter. **Guardia intercepts at the browser layer** before anything is transmitted.

02 PHI scanner fires at 99% confidence. All 5 sensitive entities are auto-masked. ChatGPT receives. "**patient [PERSON_1], DOB [DATE_1], MRN [MRN_1], balance [FINANCIAL_1], contact [PHONE_1]**"

03 Real data is routed to the secure DB. Exact records are retrieved and masked again before re-entering context. **LLM never sees real values**

04 Guardia restores real values in the response. Employee sees the complete, accurate draft — with real name, and figures, exactly as expected.

OUTCOME

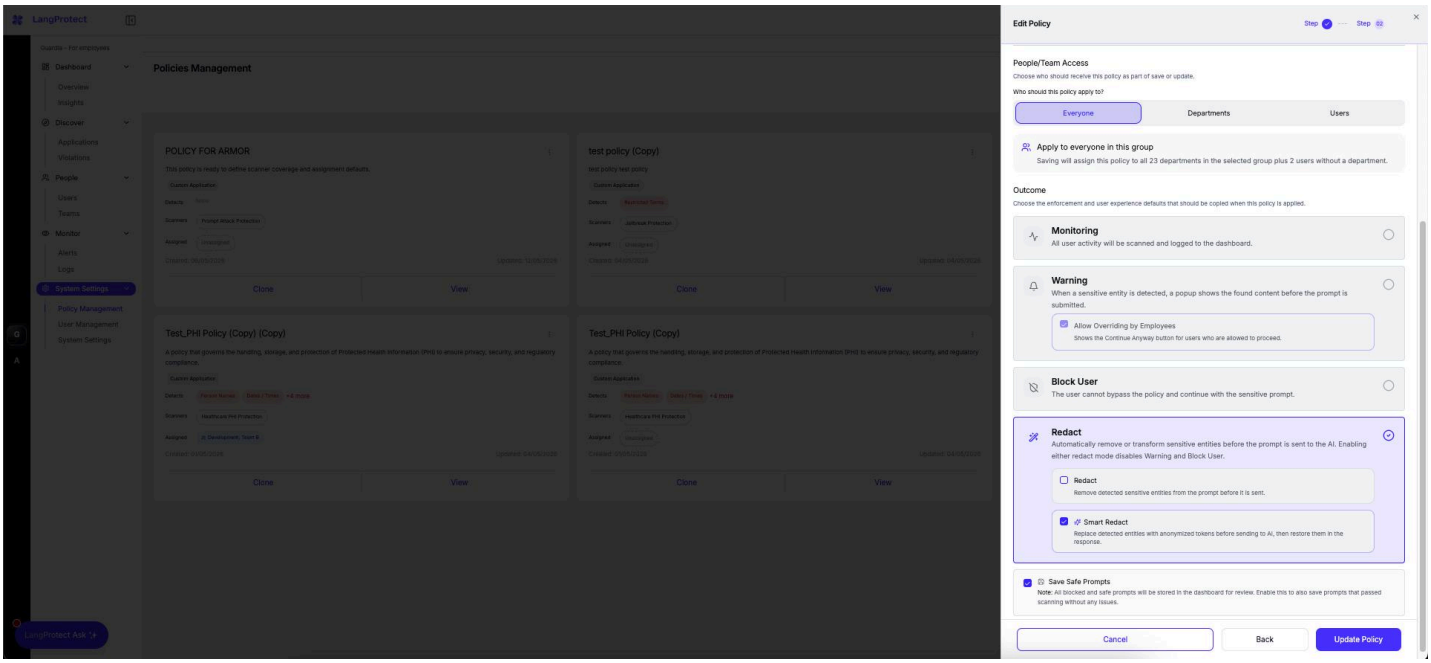
| | |
|---|---|
| PHI transmitted | None |
| HIPAA BAA in place | Not required |
| Audit record | Complete, exportable |
| Employee experience | Uninterrupted |
| Compliance posture | Protected |

## What Security Teams See When Smart Redact Is Running

The enforcement capability is only half the picture. The other half is what the security team gains on their side of the interaction: the visibility, the configuration control, and the compliance evidence that Smart Redact produces automatically every time it fires. For a CISO, this is the layer that transforms an invisible enforcement action into a defensible governance record.

### Setting Up Smart Redact Across the Organization

Deploying Smart Redact does not require network reconfiguration, endpoint agent rollout, or changes to the AI tools employees are already using. The entire configuration happens inside the LangProtect admin dashboard in a single policy edit. The admin selects the Redact outcome, chooses Smart Redact as the enforcement mode, and applies the policy to the entire organization, to specific departments, or to individual users. The policy goes live across all active browser extensions within 60 seconds of being saved, no employee action required, no session restart, no disruption to any workflow in progress.




The same interface controls who the policy applies to and at what scope. A security team running a phased rollout can apply Smart Redact to the highest-risk departments first, healthcare, finance, legal, and expand coverage incrementally as the baseline data confirms the policy is performing as expected. Policy changes take effect in real time. There is no deployment cycle and no change management window required.

**What Each Log Entry Contains**

| Fields | Example Value |
|---|---|
| User Identity | john.doe@organisation.com |
| AI Tool Accessed | chat.openai.com |
| Timestamp | 2026-04-08 — 10:14:12 |
| Scanner Triggered | Healthcare PHI Protection |
| Confidence Score | 99 |
| Action Taken | SMART REDACTED |
| Tokenized Prompt | Patient [PERSON_1], DOB [DATE_1], MRN [MRN_1] |
| PHI Transmitted to LLM | None |
| Risk Classification | None |

Every field is searchable. Every log is exportable. Real PHI never appears in the audit record, only the token identifiers.

The audit log captures the tokenized prompt, not the original, meaning the security team has a complete record of what the scanner detected and what action was taken, without the audit infrastructure itself becoming a repository of unprotected sensitive data. Real PHI is confined to the encrypted session vault for the duration of the interaction and does not appear anywhere in the log.

## Compliance Evidence Generated in Under Two Minutes

For organizations subject to HIPAA, SOC2, GDPR, or PCI-DSS, the audit trail that Smart Redact generates is not a reporting feature; it is the compliance infrastructure. When a regulator asks an organization to demonstrate that patient health information was never transmitted to a public AI platform without appropriate controls in place, the answer is no longer a manual reconstruction exercise that takes days to produce and still cannot be guaranteed to be complete.

The security or compliance officer opens the LangProtect Armor/Guardia audit log, filters by the Healthcare PHI Protection scanner and the relevant date range, and exports the result in PDF or CSV format. Every Smart Redact event in that period, the user, the tool, the timestamp, the scanner confidence score, the action taken, and confirmation that no PHI was transmitted, is in the export. The entire process takes under two minutes from the moment the question is asked.

The difference between an organization that can answer a regulatory question in two minutes and one that cannot is not the quality of their written policy. It is whether the underlying infrastructure was capturing evidence at the moment every interaction occurred. Smart Redact captures it automatically, for every event, from the first day the policy is active.

# The Business Impact of Removing the Trade-off

For most organizations, AI governance has meant choosing between two outcomes: either employees are productive and data is at risk, or data is protected and employees are frustrated. Smart Redact removes that choice entirely. The business impact is not just a security improvement. It is a structural change in how the organization can approach AI adoption at scale.

## For Employees/Users: AI Works the Way It Should

Smart Redact is the only enforcement mode that delivers full data protection without asking anything of the employee and users. No popups, no blocks, no prompts to edit or reconsider. Employees interact with their AI tools exactly as they did before the policy was active.

- No training required on what data to exclude from prompts
- No workflow interruptions during high-volume, deadline-driven tasks
- No degraded AI output from stripped or incomplete prompts
- No reason to seek workarounds or shadow AI alternatives

The result is an organization where AI adoption accelerates because the governance layer is invisible, not one where adoption stalls because the security layer creates friction.

## For Security Teams: Control Without Conflict

Smart Redact gives security teams the enforcement coverage they need without putting them in conflict with the business units and employees they serve. The policy is set once and enforced continuously, across every AI tool, every department, every interaction.

- Complete audit trail from day one, no manual logging, no reconstruction after the fact
- Real-time detection across PHI, PII, and PCI classifications simultaneously
- Policy scoped at org-wide, department, or individual user level
- Compliance exports generated in under two minutes for HIPAA, SOC2, GDPR, and PCI-DSS

Security teams stop being the team that blocks productivity. They became the team that made unrestricted AI use safe.

## For the Organization: AI Governance Built Into the Infrastructure

The organizations that treat AI governance as an afterthought will spend the next two years responding to breaches, regulatory questions, and internal incidents that a single enforcement layer would have prevented. The ones that build it into the infrastructure now will have the audit trail to prove nothing ever happened.

- **Average cost of a healthcare data breach:** $10.9M, IBM Cost of a Data Breach Report, 2024
- **HIPAA civil penalties:** up to $1.9M per violation category per year
- **Reputational risk:** HIPAA breach notifications are public record, organizational names, not just patient counts

Smart Redact is the infrastructure layer that makes AI adoption a governable, auditable, and defensible business decision.

## See Smart Redact in a Live Workflow

Every organization using AI tools today has employees sending sensitive data to external models. Most do not know how often it is happening, which tools it is happening on, or what data has already left. Smart Redact answers all three questions, and stops the exposure, from the first day it is active.

**In a 30-minute session, the LangProtect team will walk through:**

- A live Smart Redact enforcement demonstration using a real AI tool workflow
- Your organization's likely exposure profile based on industry and team structure
- Policy configuration for your specific compliance requirements, HIPAA, SOC2, GDPR, or PCI-DSS
- What a full LangProtect deployment looks like from day one through to org-wide enforcement

WWW.LANGPROTECT.COM/CONTACT-US ›