AI Privacy 101: What Happens to Your Data When You Use AI Tools?

A 60-minute deep dive into AI data practices, your rights, and how to protect yourself

Presented by:

Kyle Smith
Security Advisor Services Team Supervisor
CISA, CISSP, QSA

Allison Zwaschka
Sr. Customer Success Manager
PCIP

Our Presenters

Kyle Smith

Security Advisor Services Team Supervisor
CISA, CISSP, QSA

Allison Zwaschka

Sr. Customer Success Manager
PCIP

Today's Agenda

1 **The AI Data Landscape**

What AI tools exist and why data matters

2 **How AI Collects Your Data**

Input data, metadata, and training pipelines

3 **Where Your Data Goes**

Storage, third parties, and model training

4 **Your Rights & the Law**

GDPR, CCPA, FERPA, and opt-out options

5 **Practical Protection**

Safe habits and organizational policies

6 **Q&A**

Open discussion

The AI Data Landscape

Understanding what's out there and why data is the fuel that powers it all.

AI Tools You Use Every Day

Chatbots & LLMs
ChatGPT, Claude, Gemini, Copilot

AI Writing Tools
Grammarly, Jasper, Notion AI

AI Search
Perplexity, Bing AI, Google AI Overviews

AI Image/Video
DALL-E, Midjourney, Runway

Voice Assistants
Siri, Alexa, Google Assistant

AI in Apps
Smart Compose, autocomplete, filters

Why Data Is the Fuel

AI models don't just use data — they ARE data.

Training:
LLMs are trained on billions of text examples. Your inputs can become part of future training data.

Improvement:
Companies use interaction data to fine-tune and improve their models over time.

Personalization:
Your history makes the AI more useful to you but also builds a profile about you.

Monetization:
Data is sold, licensed, or used to train proprietary models sold to enterprises.

How AI Collects Your Data

Three categories of collection that most users never think about.

Category 1:

What You Type (Input Data)

Everything you enter into an AI tool is captured and deletion isn't always immediate or guaranteed.

- Prompts and questions, including sensitive business or personal details
- Documents you paste in or upload for summarization
- Code snippets that may contain API keys, passwords, or proprietary logic
- Medical, legal, or financial questions asked in plain language
- Personal names, emails, and account details mentioned in context
- System prompts if you're a developer using the API

⚠ **Assume everything you type is stored and potentially readable by company employees.**

Category 2:
Metadata

The data ABOUT your data — often more revealing than the content itself.

IP Address & Location

Every request logs your IP, revealing your rough location and network.

Device & Browser

OS, browser version, screen size, time zone creates a fingerprint.

Usage Patterns

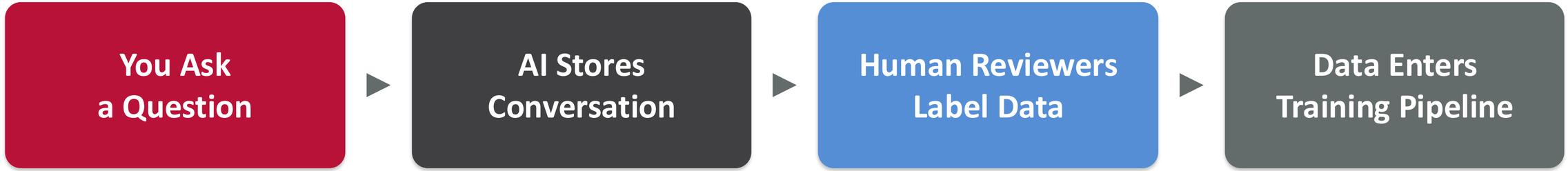
Time of day, session length, feature usage, click patterns.

Session History

How topics flow, what you iterate on, what you avoid.

Category 3:

Model Training Data



Default in many tools: Unless you opt out, your conversations are used for training.

Human annotation: Real employees may read your conversations to label them for quality.

Opt-out isn't retroactive: Turning off training after the fact may not remove past data.

Real AI Usage Study

27%

of corporate data fed into AI tools is sensitive (up from 11% a year earlier)

11%

of all content pasted into ChatGPT is confidential company data

0.9%

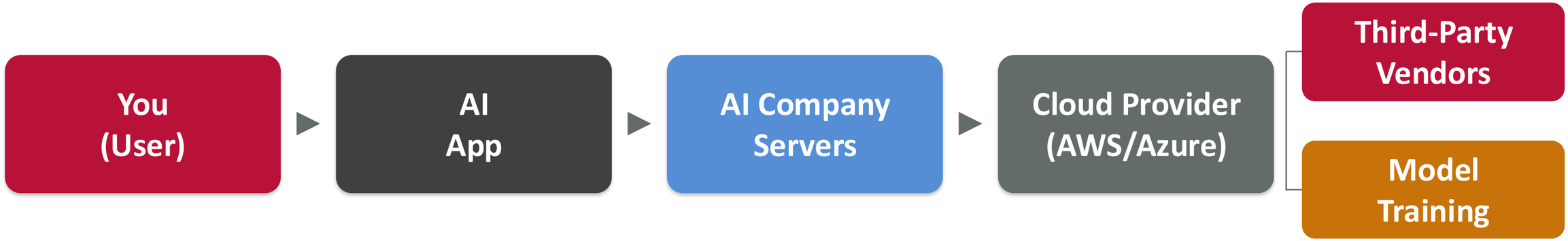
of employees cause 80% of all sensitive-data leaks to AI tools

Source: Cyberhaven Labs AI Adoption & Risk Report (analysis of 1.6M workers), 2023–2024

Where Your Data Goes

Storage, third-party access, and the model training supply chain.

The Data Journey: From You to…



Retention

Most providers keep data 30 days to indefinitely depending on settings.

Subprocessors

AI providers rely on dozens of subprocessors. OpenAI publishes its list publicly.

Gov Requests

Companies can receive legal orders to disclose user data to governments.

What the Privacy Policies Actually Say

OpenAI / ChatGPT	Consumer chats may train models unless you opt out. Deleted chats removed in ~30 days. Business/API data not trained on by default.
Google Gemini	Human reviewers may read chats. “Keep Activity” is ON by default and feeds training; can be turned off. Default auto-delete: 18 months.
Microsoft Copilot	Enterprise (Entra ID) accounts excluded from training. Consumer Copilot trains on chats by default; opt-out in Privacy settings. EEA excluded.
Anthropic/Claude	Since late 2025, trains on consumer chats unless you opt out. Opt-out = 30-day retention; opt-in = 5 years. API & Enterprise/Education excluded.

📌 **Policies change. Always check the current version at each provider's website.**

The Training Data Problem

Once your data trains a model, it's nearly impossible to fully remove.

The problem: AI models don't store data like a database. They “bake” patterns from training data into billions of parameters.

Can you be forgotten? Technically yes under GDPR, but in practice, model retraining is expensive and rarely retroactive.

Regurgitation risk: Models can sometimes reproduce training data verbatim, including personal information.

The “copyright” parallel: Courts are still deciding whether training on personal data requires consent.

Your Rights and the Law

GDPR, CCPA, FERPA, and what they actually protect you from.

Key Privacy Laws and AI

GDPR	EU / EEA + global companies serving EU	Access, correction, deletion, portability, opt-out of automated decisions	Doesn't fully address AI model training; still evolving
CCPA	California residents	Know what's collected, opt out of sale, delete personal info	Limited to California; excludes some small companies
FERPA	US students (K-12 and higher ed)	Controls on educational records; limits sharing with AI tools	Schools often unaware AI tools = FERPA-covered data processors
HIPAA	US healthcare providers	Strict limits on PHI sharing; AI tools must sign BAAs	Consumer AI tools (ChatGPT, etc.) are NOT HIPAA compliant

AI & Regulated Data: Where Liability Lives

Some data triggers legal obligations the instant it touches an external AI tool.

PCI DSS

Cardholder data: PAN, CVV, expiry. Pasting a card number into a public AI tool can pull it into PCI scope and breach the standard, even inside a "draft this email" prompt.

HIPAA

Protected Health Information. Consumer AI tools are not HIPAA-compliant and won't sign a BAA, so typing patient details is a reportable disclosure.

FERPA

Student education records: grades, advising and disciplinary notes. Sending these to external AI is unauthorized disclosure absent a school-official agreement.

GLBA

Customer financial info. Higher-ed financial aid and bursar data falls under the GLBA Safeguards Rule; external AI processing needs documented vendor controls.

The EU Act: What’s New

The world's first comprehensive AI regulation, now in force.

- **Risk-Based Framework:** Four tiers: unacceptable, high, limited, minimal; each with different obligations. Fines up to €35M or 7% of global turnover.
- **High-Risk AI:** Hiring, education, credit scoring, biometrics. Education is an explicitly designated high-risk domain.
- **Transparency Rights:** Users must be told when they're interacting with AI and when AI makes decisions about them.
- **GPAI Already Live:** Rules for general-purpose models (GPT, Claude, Gemini) have applied since Aug. 2, 2025; disclosure and copyright obligations.
- **Timeline:** In force since Aug. 2024. High-risk duties deferred: Annex III → Dec 2027; embedded (Annex I) → Aug. 2028. Extraterritorial scope applies.

How to Opt Out of Training Data Use

ChatGPT: Settings → Data Controls → Improve the model for everyone → Toggle OFF

Google Gemini: myactivity.google.com/product/gemini → turn off “Keep Activity”

Microsoft Copilot: Profile → Privacy → toggle OFF “Model training” (text and voice)

Anthropic / Claude: Settings → Privacy → toggle “Help improve Claude” OFF (reverts to 30-day retention)

Enterprise Plans: Most enterprise/API plans opt out by default; confirm in your contract

⚠ **Opting out stops FUTURE use. It does not delete data already used in training.**

Your Core Privacy Rights

Right to Know

What data is collected about you and how it's used.

Right to Access

Request a copy of your personal data.

Right to Delete

Ask for your data to be erased (GDPR, CCPA).

Right to Consent

Opt in/out of certain processing activities.

Right to Portability

Receive your data in a machine-readable format.

Right to Object

Challenge automated decisions made about you.

Practical Protection

The Golden Rule of AI Privacy

Never type into an AI tool anything you wouldn't be comfortable seeing on the front page of a newspaper.

If in doubt, anonymize, paraphrase, or don't use AI for that task.

For Individuals: Do’s and Don’ts

✓ DO

- ✓ Opt out of training data use in settings
- ✓ Use anonymized or fictional examples when testing AI
- ✓ Read the privacy policy of any new AI tool (skim for “training” and “share”)
- ✓ Use enterprise/paid tiers for work-sensitive tasks
- ✓ Regularly delete your conversation history
- ✓ Use incognito/private mode for sensitive queries

✗ DON'T

- ✗ Paste confidential docs, contracts, or client data
- ✗ Type real patient, student, or employee info
- ✗ Include passwords, API keys, or account credentials
- ✗ Trust AI outputs with PII without verification
- ✗ Use consumer AI tools for tasks subject to HIPAA/FERPA
- ✗ Assume “ephemeral” or “temporary” means not stored

For Organizations: Building an AI Data Policy

1. Inventory

Catalog every AI tool in use, sanctioned and shadow IT. You can't govern what you can't see.

2. Classify

Assign data sensitivity levels. Identify which types can/cannot be used with external AI tools.

3. Negotiate

For business-critical AI tools, negotiate Data Processing Agreements (DPAs). Require training opt-outs.

4. Train

Run mandatory awareness training (like this one!). Make the Golden Rule part of your culture.

5. Monitor

Use DLP (Data Loss Prevention) tools to detect sensitive data leaving for AI endpoints.

6. Review

Reassess quarterly. AI privacy policies change. Your posture needs to keep up.

Safer AI Options for Sensitive Work

Not all AI is created equal from a privacy perspective.

Best Privacy **Local / On-Premise AI**	Run open-source models locally or on your own infrastructure. Data never leaves your environment.	*e.g. Ollama, LM Studio, Azure OpenAI (private deployment)*
Strong Privacy **Enterprise / Private Tiers**	Dedicated instances with contractual data isolation, no training on your data, and audit logs.	*e.g. ChatGPT Enterprise, Azure OpenAI, Claude for Enterprise, Vertex AI*
Moderate Privacy **Paid Consumer Tiers**	Paid ≠ private: paid consumer tiers often share free-tier training defaults.	*e.g. ChatGPT Plus/Team, Claude Pro, Gemini Advanced*
Use with Caution **Free Consumer Tools**	Default to using your data for training and improvement. Limited retention controls.	*e.g. Free ChatGPT, free Gemini, free Copilot*

Your AI Privacy Checklist

Do these 10 things and you'll be ahead of 90% of users.

- ✓ Opt out of training data use on ChatGPT, Gemini, Claude, and Copilot
- ✓ Enable conversation auto-delete (30 days or less)
- ✓ Review all AI tools you use; check their privacy policies
- ✓ Never paste real customer, patient, or student data into consumer AI
- ✓ Use enterprise tier for any AI that touches confidential work data
- ✓ Set up a password manager so you don't accidentally include credentials
- ✓ Tell your IT/security team what AI tools you're using
- ✓ Use incognito mode for personal AI queries on work devices
- ✓ Ask your vendor: “Do you use our data for training?” Get it in writing
- ✓ Revisit this checklist every quarter as policies evolve

What’s Coming: AI Privacy Trends

Federated Learning

AI that trains on your device without sending raw data to the cloud, better privacy by design.

On-Device AI

Apple Intelligence, Gemini Nano; processing stays local. Growing trend in mobile.

Tighter Regulation

More countries passing AI laws. US states are moving first (e.g., Colorado AI Act); a federal framework remains unsettled.

Consent Management

AI “nutrition labels” and consent dashboards; know what you're agreeing to.

Real-World Privacy Incidents

These actually happened. Learn from them.

Samsung (2023)	Engineers pasted chip source code and internal meeting notes into ChatGPT for help.	Proprietary code potentially exposed. Samsung banned AI tools company-wide.
Legal Sector (2023)	A law firm submitted AI-hallucinated fake case citations to a federal judge.	Lawyers and firm sanctioned, fined $5,000 (Mata v. Avianca).
Healthcare (pattern)	Staff at multiple hospitals using consumer ChatGPT to draft clinical notes including PHI.	HIPAA violation risk. No BAA in place. Hospitals issuing AI use policies.
K-12 & Higher Ed	Educators using AI to grade or summarize; student work sent to external AI servers.	FERPA exposure absent a school-official agreement. Districts issuing AI policies.

Key Takeaways

Do we know where our data goes?

Which of our data should never touch consumer AI?

When do we check again?

Resources & Further Reading

Official Privacy Controls

- OpenAI: help.openai.com/data-controls
- Google: myactivity.google.com
- Anthropic: support.anthropic.com/privacy

Regulatory Guidance

- EU AI Act: artificialintelligenceact.eu
- FTC AI Guidance: ftc.gov/ai
- EDPB Guidelines: edpb.europa.eu

Deep Dives

- 'Atlas of AI' — Kate Crawford
- IAPP AI Privacy Resource Center
- Future of Privacy Forum: fpf.org

Tools

- EFF Surveillance Self-Defense (ssd.eff.org)
- ToS;DR (Terms of Service; Didn't Read)
- Mozilla Privacy Not Included

CAMPUSGUARD®

Contact us for more information!

campusguard.com

info@campusguard.com

Questions?

“Privacy is not something that I'm merely entitled to — it's an absolute prerequisite.” — Marlon Brando