

Shhh, It's A Secret: Preventing Sensitive Data Exposure in Logs

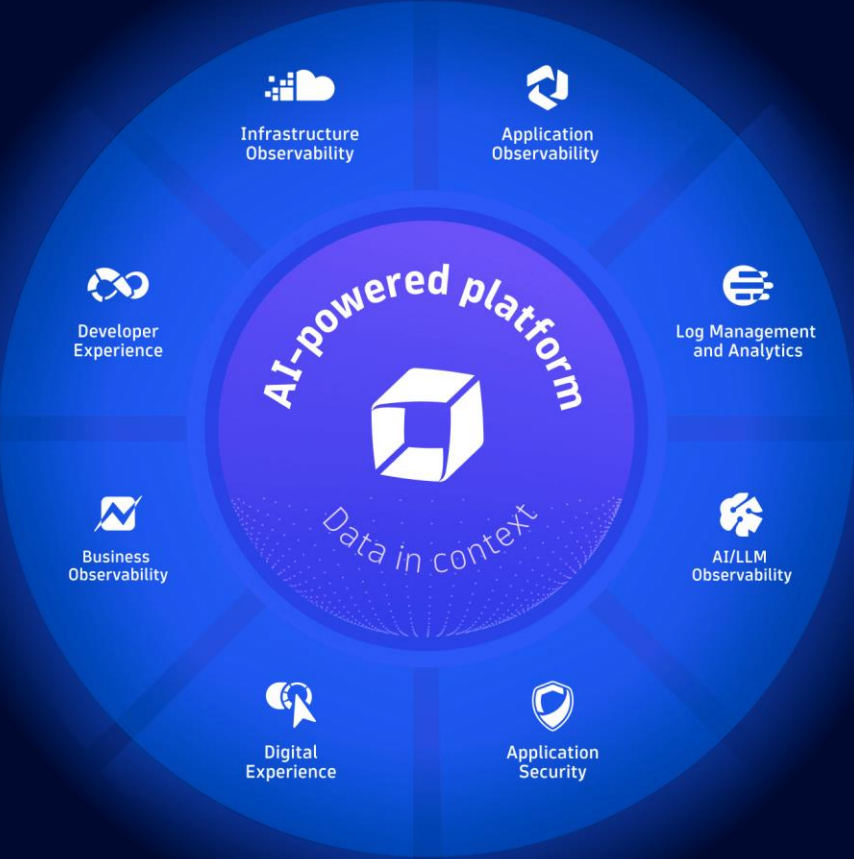
Hagenberg Security Forum 2026



PRESENTER

Cat Easdon
Privacy Engineering

My day-to-day perspective on sensitive data



Dynatrace Intelligence | Fueling the AI-powered observability platform for autonomous operations



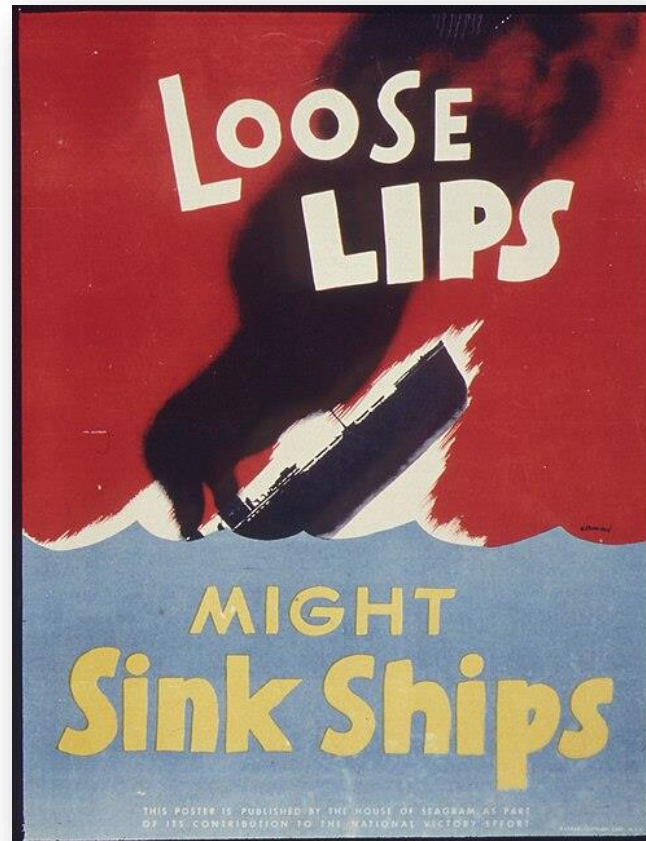
Outline

- **Why?**
- **What?**
 - Defining “sensitive” data
- **How?**
 - Context matters: a cheat sheet
 - A toolkit of technical controls
- **Conclusion**

Motivation

Recap: last year's talk

"In our age there is no such thing as 'keeping out of politics.' All issues are political issues, and politics itself is a mass of lies..." ~ George Orwell, 1941



Recap: last year's talk

*For customers, there is "no need to do anything, since **no sensitive data** like passwords or payment information was affected."*

Volkswagen's statement to [Der Spiegel](#)

“No sensitive data was affected..”



“No sensitive data was affected...”

Unique ID: h [REDACTED]

Employment

- Decision maker for Government - National Security

Most frequent locations

- Rue de la Loi 200, 1049 Bruxelles
- Av. du Bourget 44, 1130 Bruxelles
- Oncology Centre, Universitair Ziekenhuis, Laarbeeklaan 101 1090 Brussels

[More locations >](#)

Home

- [REDACTED] ixelles, Brussels

[Most frequent driving routes >](#)

Financial interest tags

- Personal Debt
- Bankruptcy
- Gambling (high spender)

Psychology affinity

- Dutiful: low affinity
- Materialistic: high affinity

Lifestyle & health tags

- Frequent alcohol purchaser
- Menopause
- Panic / anxiety disorder

Most frequent devices

- Galaxy S22 Ultra 5G: SM-S9080
- Windows 11 PC / Chrome
- Samsung TV Tizen 6.0

Frequent web visit

- www.avocat-vdb.be/en/
- www.uzbrussel.be/en/web/oncologisch-centrum/zorgverleners-maagtumor

Frequent app use

- Tinder
- Uber

Handwritten Annotations:

- MONEY** (next to Financial interest tags)
- Ego** (next to Psychology affinity)
- Compromise** (next to Frequent app use)
- DIVORCE LAWYER'S WEBSITE** (with arrow pointing to www.avocat-vdb.be/en/)
- CANCER TREATMENT CENTRE** (with arrow pointing to Oncology Centre location)

Profile Picture: A woman's face is obscured by a black box. The background features the European Union flag. Text below the image reads "Commission euro" and "European Commis".

[Next page >](#)

Reciprocation

What's happened since?

A Universe of Tokens

The Algorithmic Bridge



Meta employees used 60.2 trillion AI tokens over 30 days. Token counts include the amount of text in all inputs and outputs, as well as the model's thought processes, which often far exceed the visible output in length.

Here's how that compares to other works.

2.66T tokens — Every book in the Library of Congress



15T tokens — Tokens used to train Meta's Llama 3



~20T tokens — All books ever published (est.)



60.2T tokens — Meta's total token usage over 30 days



Source: [The Information](#)

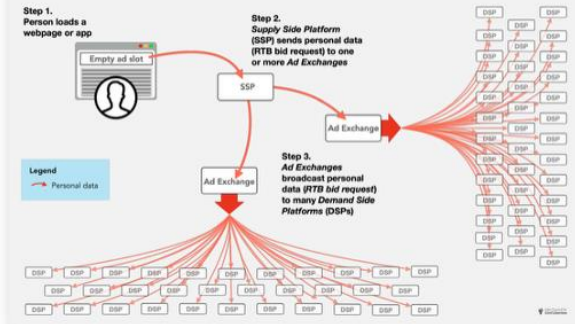
It's not just Meta, though. Nvidia CEO Jensen Huang said in the All-In Podcast that if a \$500k engineer spent less than \$250k a year in tokens, he'd be "deeply alarmed."



We're processing more sensitive data than ever before...

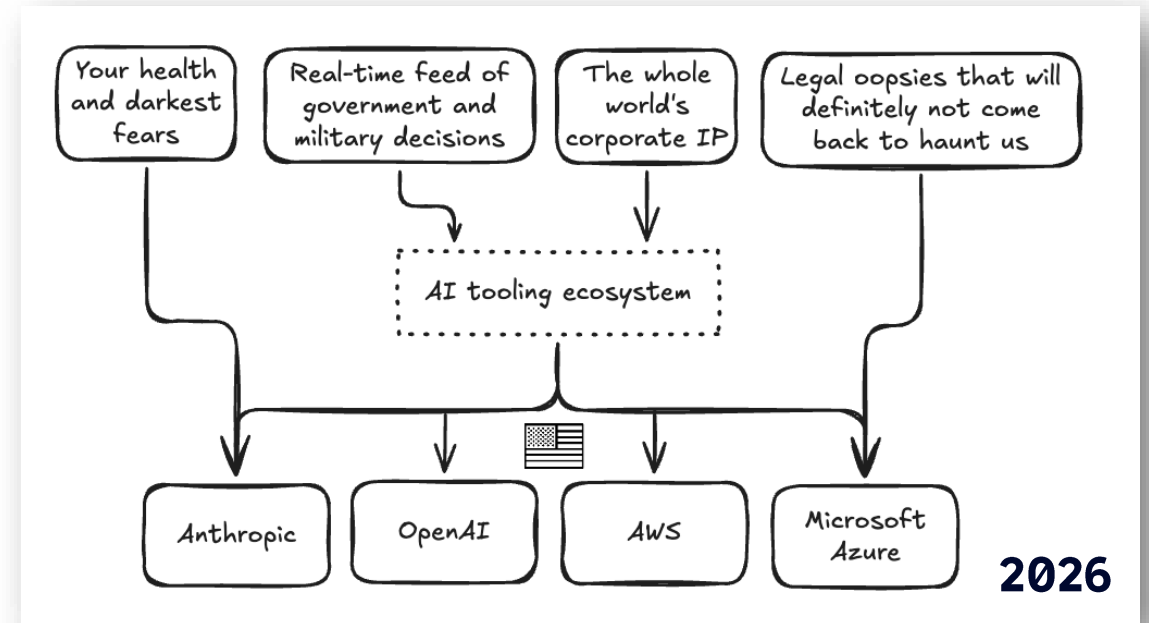
Displaying ads broadcasts your users' data to the world

The sale of a single ad slot often involves an auction of auctions, with several ad exchanges running competing auctions that are coordinated by a Supply Side Platform (SSP). This increases the number of DSPs that receive the broadcasted data.



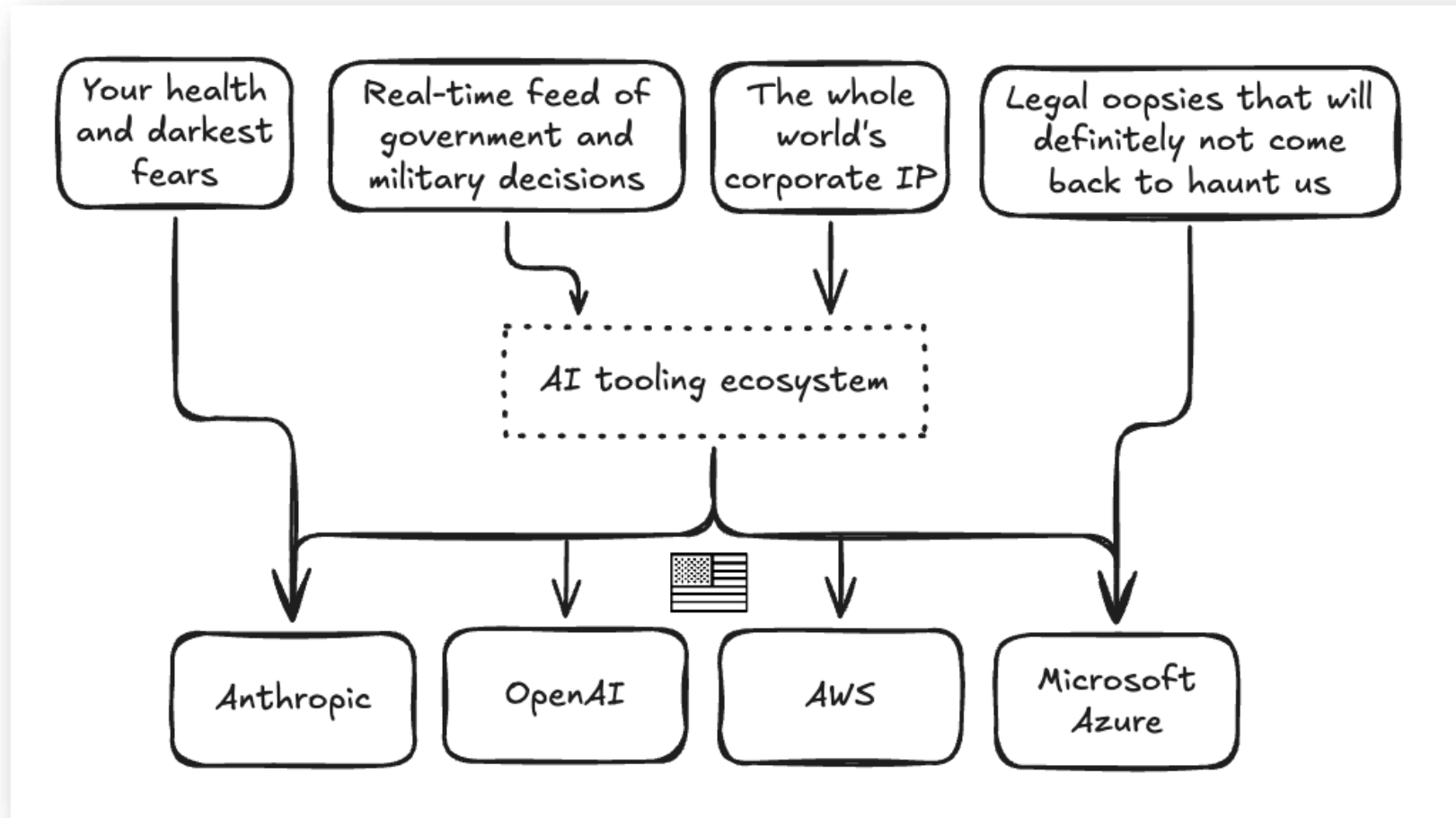
2025

33



2026

We're processing more sensitive data than ever before...



We're processing more sensitive data than ever before...


...certain written exchanges that defendant Bradley Heppner had with a generative artificial intelligence (“AI”) platform were not protected from government inspection...Heppner could have had **no “reasonable expectation of confidentiality in his communications” with Claude.**

~ [United States vs. Heppner](#) (2026)

...in a highly unstable geopolitical context

Credit: [Michael Staudacher](#)

SOVRÄN-WÄSH - HYPERSCÅLER (Model EU-US-702)

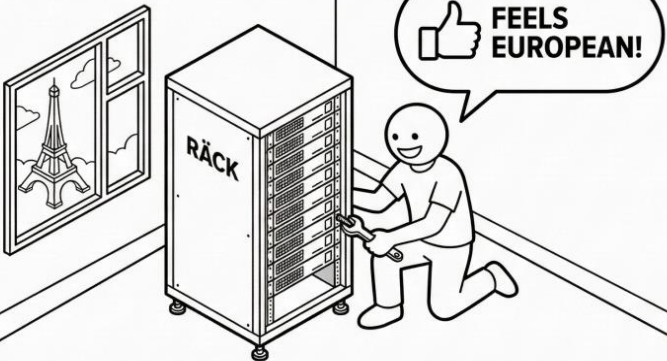


PARTS INCLUDED

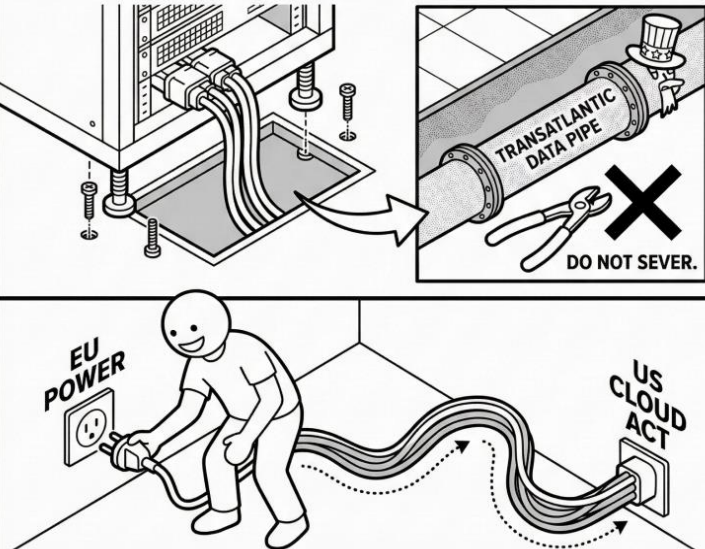
- RÄCK (Made in Delaware) - x1
- FISA-CABLES (Non-removable) - x1
- SOVEREIGN LABELS (Localised adhesive) - x1
- ENCRYPTION KEY (Spare held off-site) - x1

WARNING: JURISDICTIONAL IMMUNITY NOT INCLUDED. Product may contain traces of foreign intelligence.

STEP 1: LOCAL ASSEMBLY



STEP 2: 'SECURE' CONNECTION



The problem

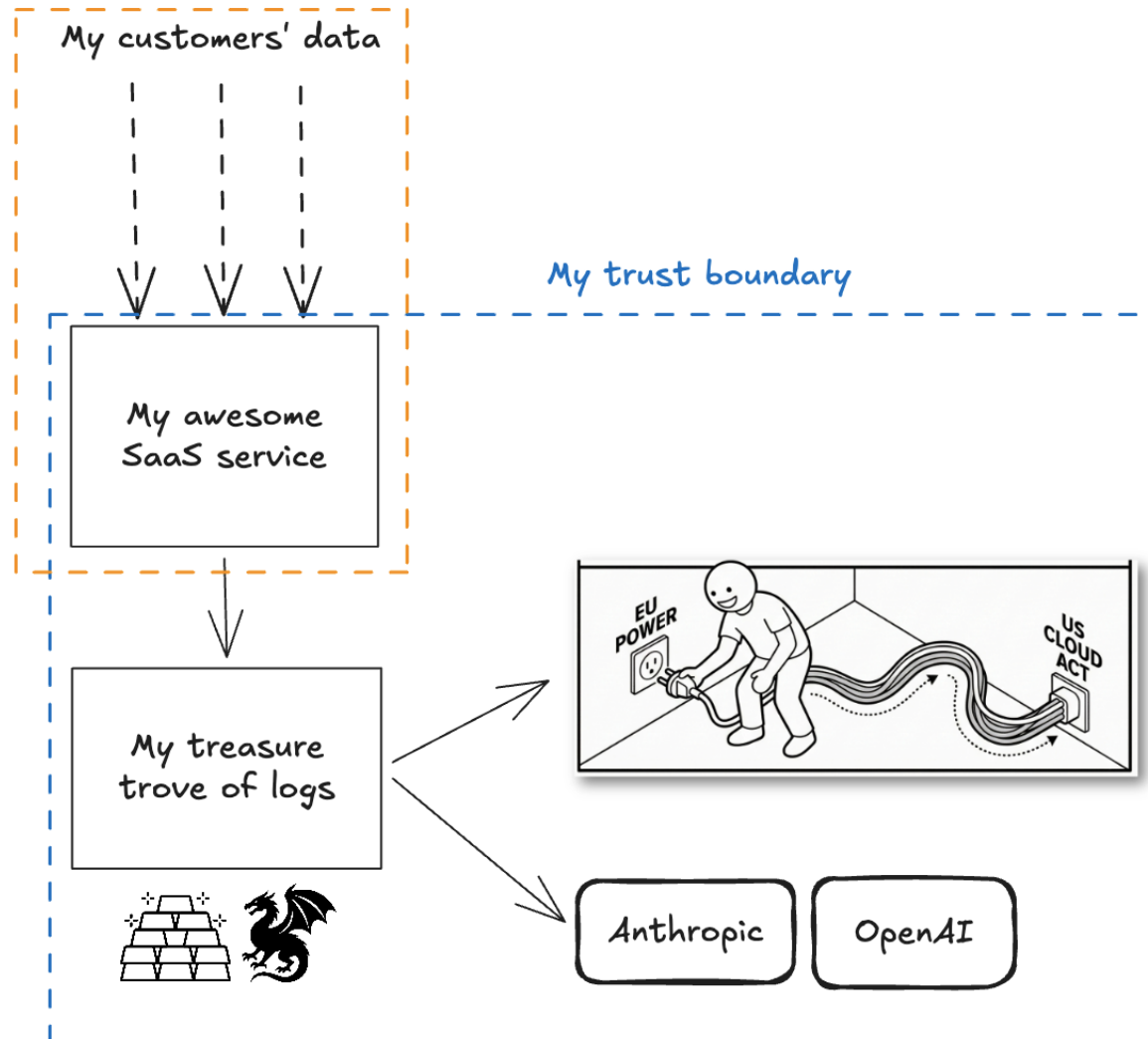


- You are vulnerable to sensitive information leakage by making logging and alerting events visible to a user or an attacker (see [A01:2025-Broken Access Control](#)), or by logging sensitive information that should not be logged (such as PII or PHI).

**CERT Oracle Secure Coding Standard for Java (2011):
“Do not log sensitive information outside a trust boundary”**

"...outside a trust boundary"

Customer's trust boundary



Sensitive data? In *my* logs?

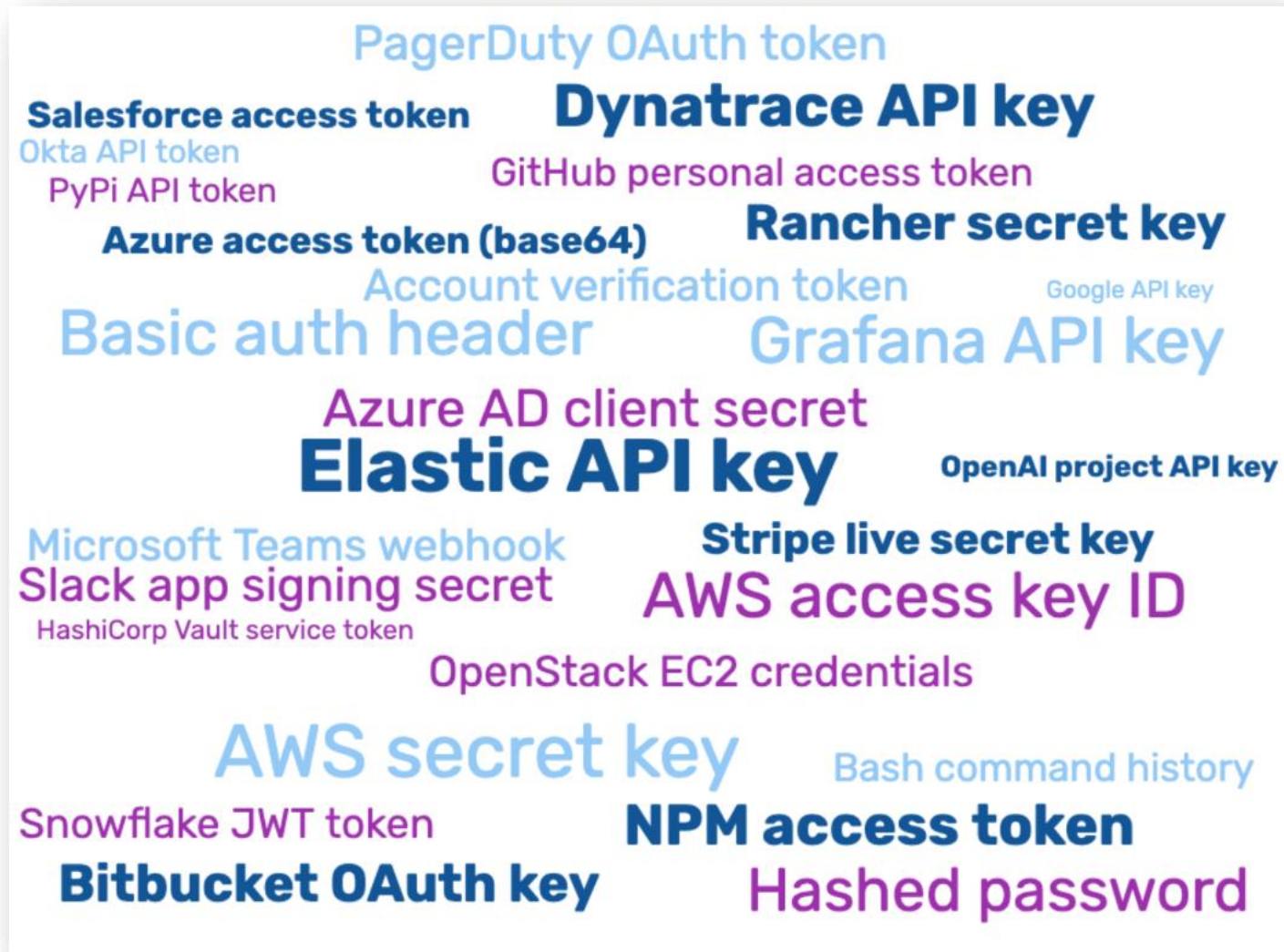
The problem



- You are vulnerable to sensitive information leakage by making logging and alerting events visible to a user or an attacker (see [A01:2025-Broken Access Control](#)), or by logging sensitive information that should not be logged (such as PII or PHI).

**Have you faced this?
What happened?**

Sensitive: credentials (yours and your customers')



Also sensitive: PII



Also sensitive: other personal data and customer data



Case study: tax offices

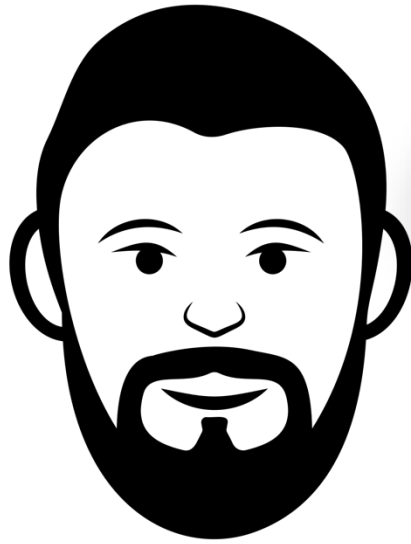


[At the Canadian Revenue Office] if you go and look up a person's financial data...No questions, just fired.

Tanya Janca, "[Privacy by Design: Making Threat Modeling Work for Data Protection](#)" (2025)

- Robust access controls in place at the tax office itself 😊
- But what about at the tax office's service providers?
- What could happen if each one says, "*Customer data isn't that sensitive. They choose to send it to us anyway. They should remove secrets beforehand!*" ?

Case study: tax offices



We try to keep PII
out of our logs,
but we keep
finding leaks!

*Ben Conrad and Gerald Benischke,
“[Scaling application security to 100+ teams at HMRC Digital](#)” (2023)*

- Dealing with billions of logs per day
- Nation state actors in their threat model
- Redaction creates a false sense of security, either you miss a lot or you redact too much
 - Devs need logs for troubleshooting
 - Fraud investigation teams need logs and audit events because providing false data is the main way someone ‘hacks’ a tax office

Case study: tax offices

DDCYLS-8197: remove exception from log as displaying PII #583

Merged [redacted] merged 1 commit into main from DDCYLS-8197-removeAddressFromLog on Mar 11

Conversation 1 Commits 1 Checks 0 Files changed 1 +1 -1

All commits 0/1 Comments

...r/SubscriptionServiceConnector.scala Viewed

```
@@ -60,7 +60,7 @@ class SubscriptionServiceConnector @Inject() (httpClient:
HttpClientV2, appConfi
60 60     } recoverWith { case e: Throwable =>
61 61         // $COVERAGE-OFF$Loggers
62 62         logger.warn(
63 -         s"Subscribe SUB02 request failed for acknowledgementReference :
        ${request.subscriptionCreateRequest.requestCommon.acknowledgementReference}.
        Reason: $e"
63 +         s"Subscribe SUB02 request failed for acknowledgementReference :
        ${request.subscriptionCreateRequest.requestCommon.acknowledgementReference}."
64 64     )
65 65     // $COVERAGE-ON
66 66     Future.failed(e)
```



Personal data >>> PII

“Mr. ? lives in ZIP [postal] code 02138
and was born on 31.07.1945.”



Personal data >>> PII

“Computer scientists have recently undermined our faith in the privacy-protecting power of anonymization...by deleting information like names and social security numbers. These scientists have demonstrated they can often 'reidentify' or 'deanonymize' individuals hidden in anonymized data with astonishing ease. By understanding this research, we will realize we have made a mistake...This mistake pervades nearly every information privacy law, regulation, and debate, yet regulators and legal scholars have paid it scant attention.”

~ Paul Ohm, [“Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization”](#) (2009)

Personal data >>> PII

“Although it is true that a malicious adversary can use PII such as a name or social security number to link data to identity, as it turns out, the adversary can do the same thing using information that nobody would classify as personally identifiable...These results suggest that **maybe everything is [personal data]** to one who has access to the right outside information.”

~ Paul Ohm, “[Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization](#)” (2009)

Case study: tax office (UK)

PII

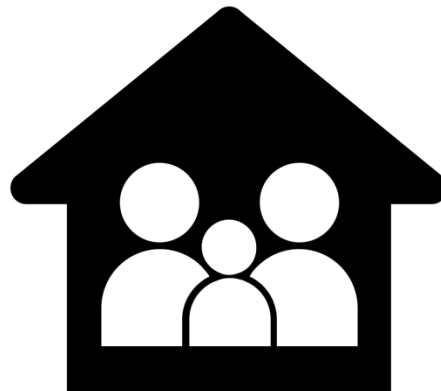
- Name
- Email address
- Address
- Tax number
- IBANs (all of them, even the Swiss one 😄)
- ...

Personal data

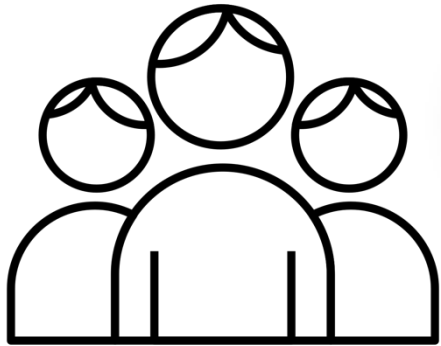
- Income and income sources
 - Student loan
 - Some disabilities
- Inheritance and estate details
- When you log in (3am? Look who's got insomnia!)
- ...

So every log is sensitive and we're doomed? Nope 😊

- Society runs on information flows, but these flows should be *context-appropriate*.
- Laws and standards typically take a risk-based approach:
 - Is this an appropriate data flow in this context?
 - Do the benefits to individuals (**not** to your business) outweigh the potential harm?
 - Do you have appropriate mitigations in place to prevent data breaches, insider threat, and repurposing of this data?



So can I log XYZ? *It depends...*

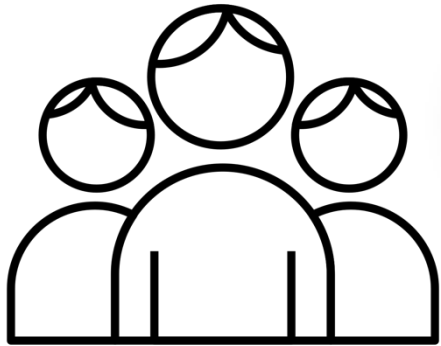


Hey, can we log request bodies and request responses in service X?

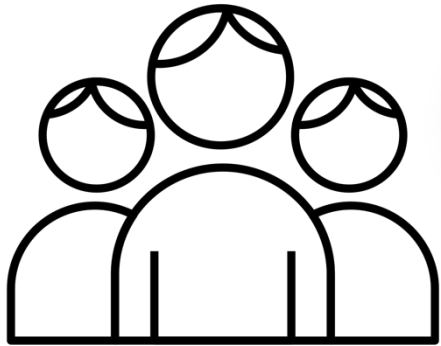


Why do you need them?

So can I log XYZ? *It depends...*



So can I log XYZ? *It depends...*



To debug Y. There is some PII but Jonas said you let his team log PII, so we thought it's allowed now.



You don't need that data though? It looks like all you need is the request ID?

Context matters: a cheat sheet

- **What data do we need and why do we need it?**
 - Can we reduce this? What do we *really* need? (Five Whys!)

Context matters: a cheat sheet

- What data do we need and why do we need it?
 - Can we reduce this? What do we *really* need? (Five Whys!)
- **Have we collected data for this reason before?**

Context matters: a cheat sheet

- What data do we need and why do we need it?
 - Can we reduce this? What do we *really* need? (Five Whys!)
- Have we collected data for this reason before?
- **Could it ever contain credentials or confidential information?**

Context matters: a cheat sheet

- What data do we need and why do we need it?
 - Can we reduce this? What do we *really* need? (Five Whys!)
- Have we collected data for this reason before?
- Could it ever contain credentials or confidential information?
- **Could it ever contain any personal data?**
 - If so, does this use case provide more benefit than harm to those people?

Context matters: a cheat sheet

- What data do we need and why do we need it?
 - Can we reduce this? What do we *really* need? (Five Whys!)
- Have we collected data for this reason before?
- Could it ever contain credentials or confidential information?
- Could it ever contain any direct or indirect personal identifiers?
 - If so, does this use case provide more benefit than harm to those people?
- **Where will it be stored and how long will it stay there for?**

Context matters: a cheat sheet

- What data do we need and why do we need it?
 - Can we reduce this? What do we *really* need? (Five Whys!)
- Have we collected data for this reason before?
- Could it ever contain credentials or confidential information?
- Could it ever contain any direct or indirect personal identifiers?
 - If so, does this use case provide more benefit than harm to those people?
- Where will it be stored and how long will it stay there for?
- **Who will be able to access it?**

What can we do about it?

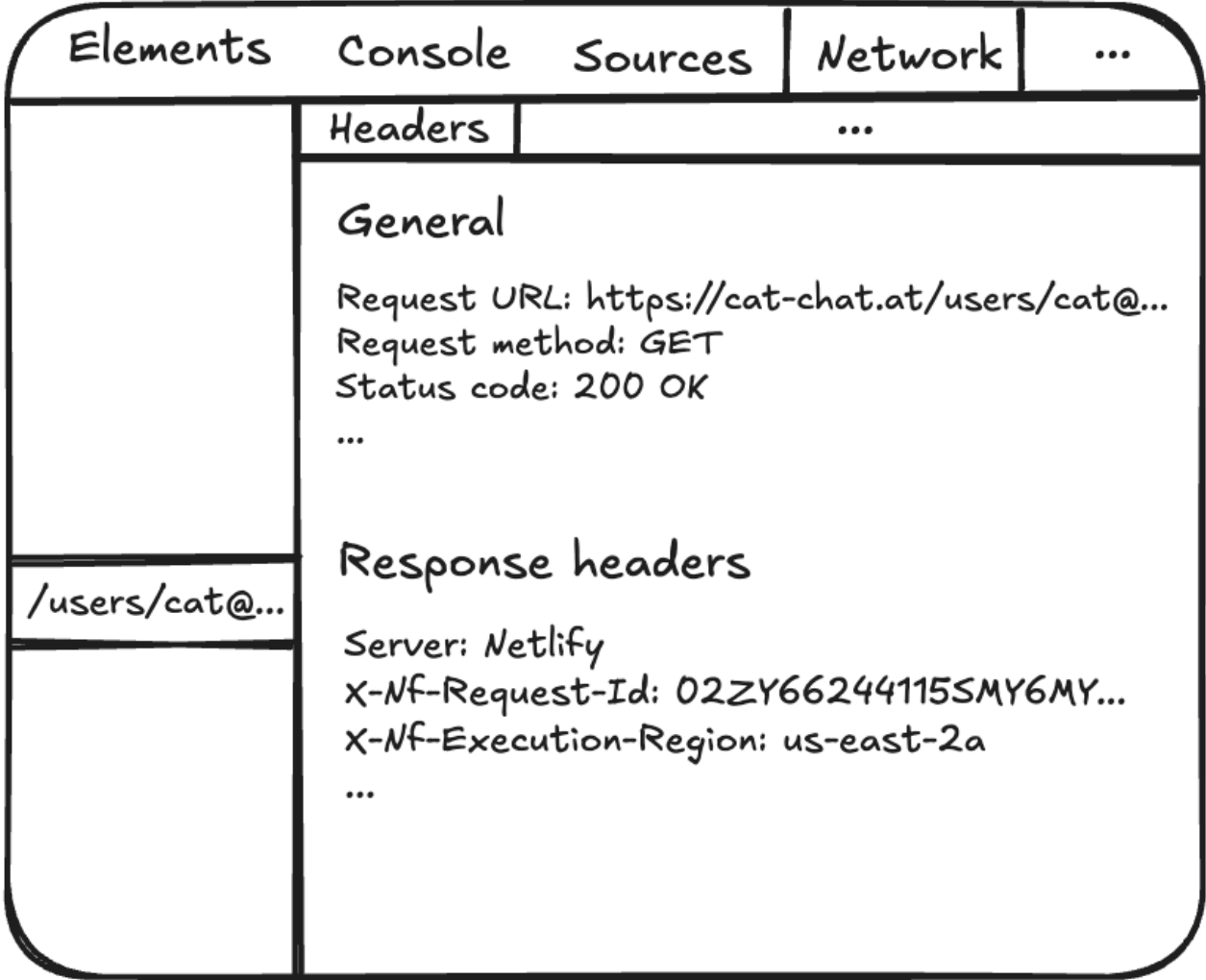
A toolkit for preventing sensitive data exposure

Data model and API design	Logging patterns	Code scanning	Regression tests	Runtime exceptions	Drop/redact on ingest	Scan logs
Design	Build	CI/Agent	QA		Runtime	

Design: data model and API design

```
{
  "id": "04401ba0-2a69-4854-9816-
7241c986776a",
  "email": "not.cats.email@email.com",
  "name": "Cat Easdon",
  "picture": "https://glt-my-first-chat-app.s3.us-
east-1.amazonaws.com/IMG1501.png",
  "created": 1775060002,
  "phoneNumber": "+43123456879"
}
```

Design: data model and API design



Design: data model and API design

“You will not and will not assist or permit any third party to pass information, hashed or otherwise, to Google that Google could use or recognize as personally identifiable information.”

~ Google Analytics [Terms of Service](#) (last updated 2023)

Build: standardized logging

Secure
logging
rules

Use types
to
reinforce
rules

Sanitize as
best you
can

Enforce
policies

Build: using types to reinforce rules

```
// The infamous React example  
return <div dangerouslySetInnerHTML={createBody(userInput)} />
```

```
// A privacy equivalent to consider  
export type Telemetry_I_VERIFIED_THIS_CANNOT_CONTAIN_USER_INPUT = never;  
...  
privacyLogger.log(errorDto as Telemetry_I_VERIFIED_THIS_CANNOT_CONTAIN_USER_INPUT);
```

Build: secure logging guidance for developers

```
// File and folder paths contain all sorts of fun stuff:  
const ifYouLogMeYourPentestingTeamWillFightYou =  
  "/Users/cto.full.name/Strategic Projects/FY27 Layoffs/Security Downsizing Plan.docx"
```

```
// Instead, if we just need to distinguish unique file paths, we could try:  
function logFilePath(filePath: UnsafeString) {  
  const hashedFilePath = hashSha256(filePath) as SafeString;  
  log(hashedFilePath);  
}
```

```
// Better still, if we only need the *number* of unique file paths,  
// let's just keep that number rather than storing the hashes.
```

CI: code scanning



privado



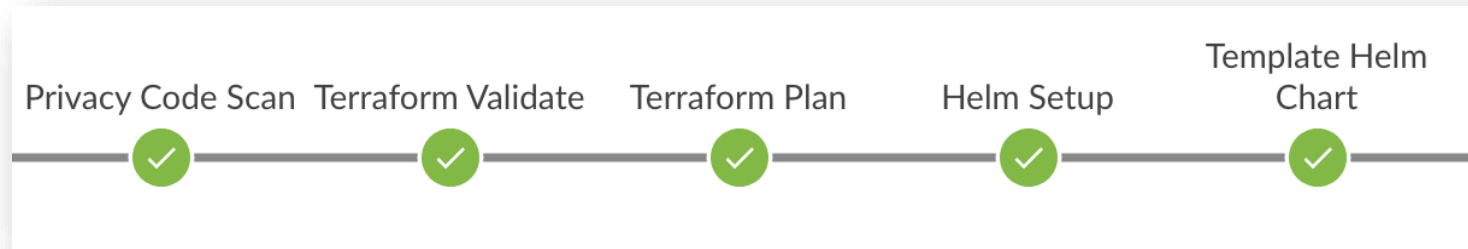
Semgrep

```
1 rules:
2   - id: no-personal-data-in-logs
3     languages:
4       - java
5     message: Personal data should not be output to logs
6     mode: taint
7     pattern-sources:
8       - patterns:
9         - metavariable-pattern:
10            metavariable: $PERSONAL_DATA
11            patterns:
12              - pattern-either:
13                # approximate location
14                - pattern-regex: (?i)(.*(?!(capa|velo)))(city|country|continent)
15                # precise location
16                - pattern-regex: (?i)(.*(gps[^\s/();#|,=>]{0,2}(?:location|position)|user[^\s/();#|,=>]{0,2}location|latitude|longitude|geo[^\s/();#|,=>]{0,2}coordinates)|(\atlng|latlon)(\b|^[^g]))
17                # email address
18                - pattern-regex: (?i)(.*email)|(?:business|personal|work|contact)[^\s/();#|,=>]{0,2}email.*|.*email[^\s/();#|,=>]{0,2}(?:address|id)
19              - pattern-either:
20                - pattern: $PERSONAL_DATA()
21                - pattern: $PERSONAL_DATA = ...
22      pattern-sinks:
23        - pattern: log.$LEVEL(...)
24        - pattern: log. ... .log(...)
25        - pattern: LOG.$LEVEL(...)
26        - pattern: LOG. ... .log(...)
27        - pattern: LOGGER.$LEVEL(...)
28    severity: ERROR
```

Regexes adapted from [Privado OSS](#)



CI: code scanning



git-privacy-ops 31 May 2024 [↗](#)

Personal data should not be output to logs

Please confirm that:

- Processing this personal data is **relevant** and **necessary**
- If this data is persisted, a retention period is enforced
- You have followed the [Privacy Coding Guidelines](#)

Reply  

Sensitive Data Scanner

Configure Grail scans to detect ingestion of sensitive data, such as credit card numbers and IBANs.



Schedule a scan

Create scan

Schedule a recurring scan for continuous monitoring.

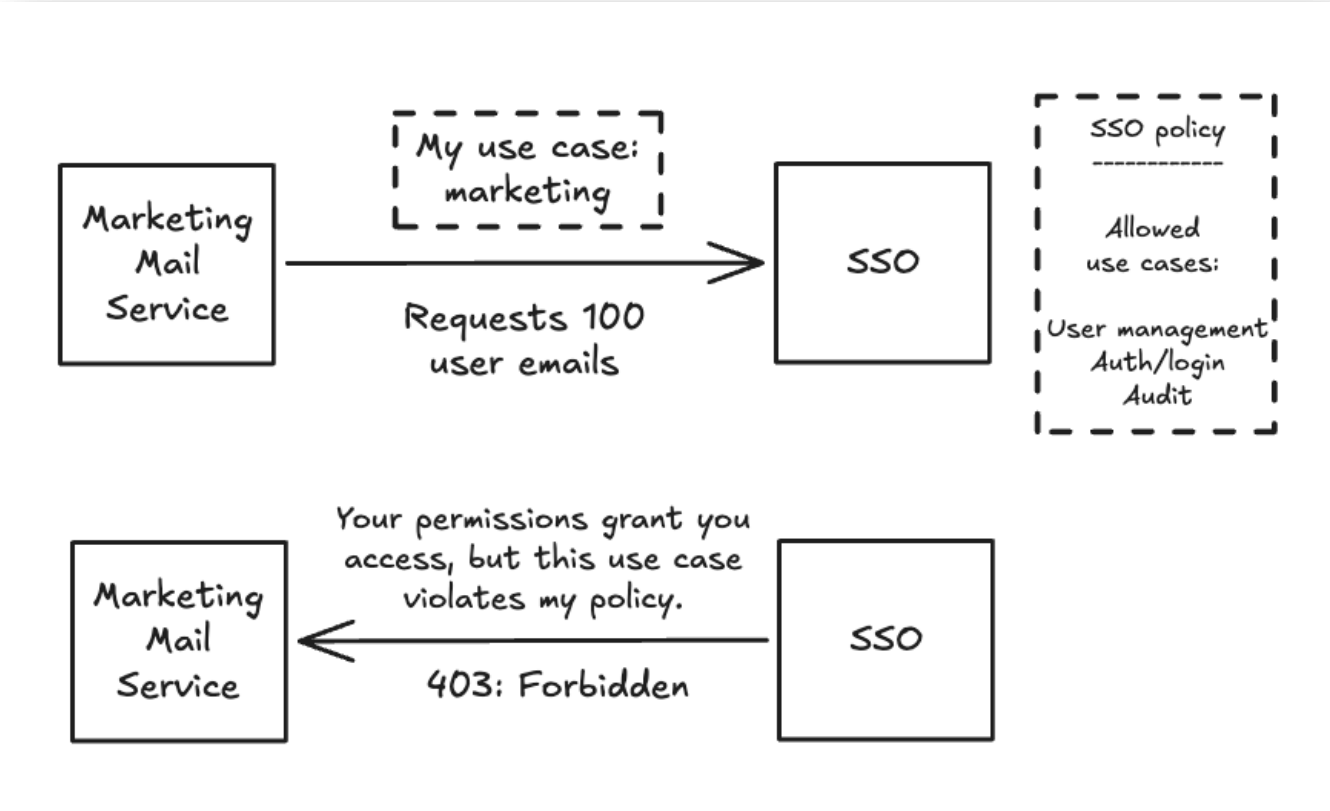
Filter scans

Name	Data types	Status	Interval	Created at	Execution history
Scan 1 with policy	IBAN	Active	12 hours	1/12/2026, 10:00:00 AM	
Scan 2 with two policies	Phone number	Active	12 hours	1/12/2026, 10:30:00 AM	
Scan 3 with one policy	IP v4 & v6 addresses	Active	24 hours	1/12/2026, 11:00:00 AM	
Scan 5	Dynatrace API token	Active	24 hours	1/12/2026, 12:00:00 PM	
Scan 5	Email address	Active	6 hours	1/12/2026, 1:00:00 PM	
Scan 6	Credit card numbers	Active	6 hours	1/12/2026, 2:00:00 PM	
Scan 7	Credit card numbers, IBAN	Active	6 hours	1/12/2026, 3:00:00 PM	
Scan 8	Credit card numbers, Dynatrace API token, Email address, IBAN, IP v4 & v6 address...	Active	6 hours	1/12/2026, 4:00:00 PM	

20 rows per page

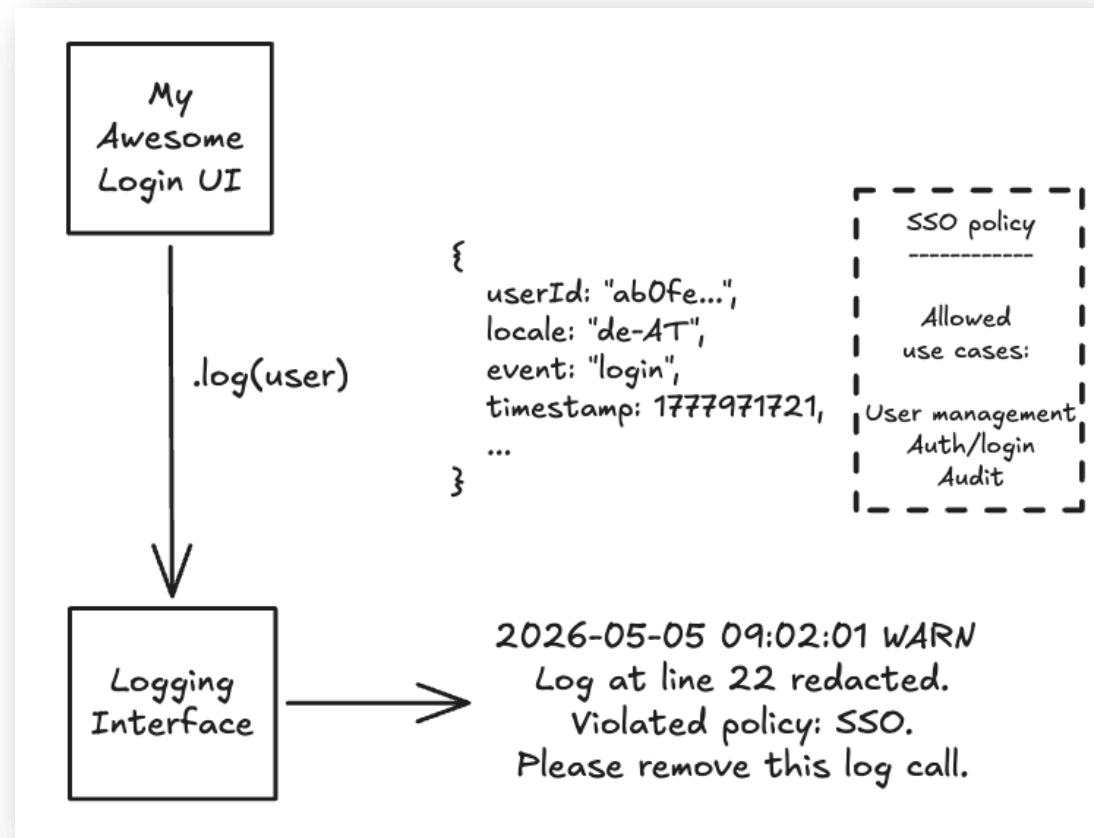
Runtime: exceptions for data flow enforcement

Build upon runtime authorization (*who* can access this data?) to add runtime data flow enforcement (*what* can we do with this data and *where* can it go?)



Runtime: exceptions for data flow enforcement

When SSO data is provided for a valid use case, its policy comes with it and is applied downstream by our sanitizing logger when an application tries to log it.

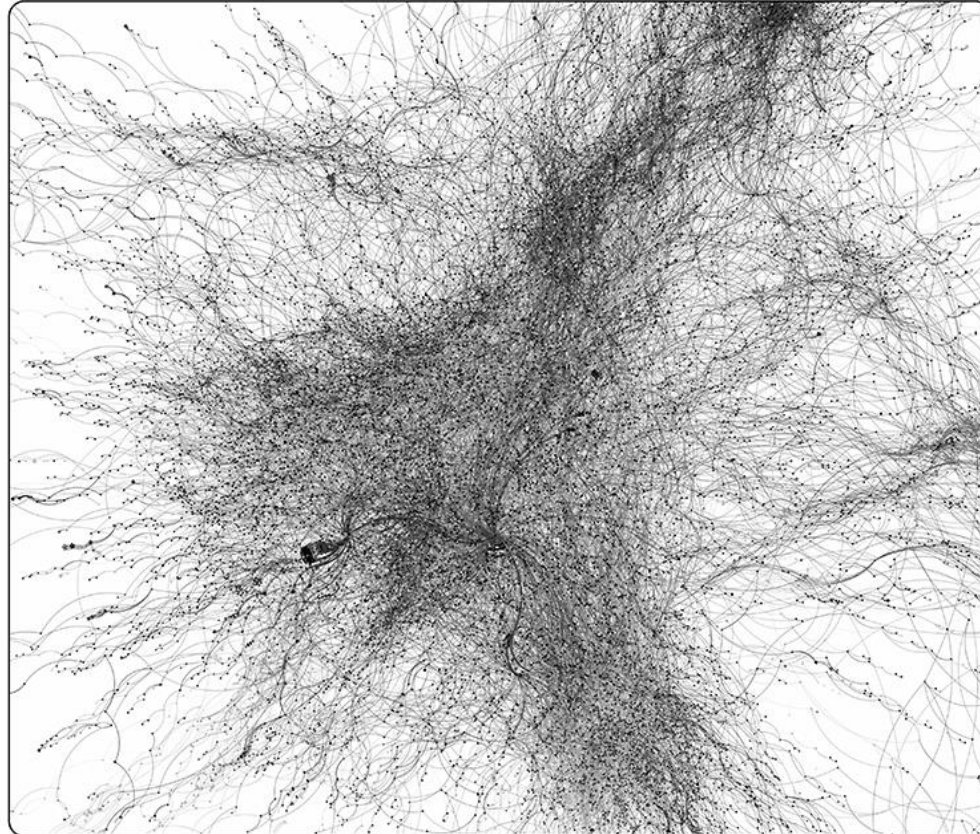


Runtime: exceptions for data flow enforcement

Source: [Facebook Engineering blog](#)

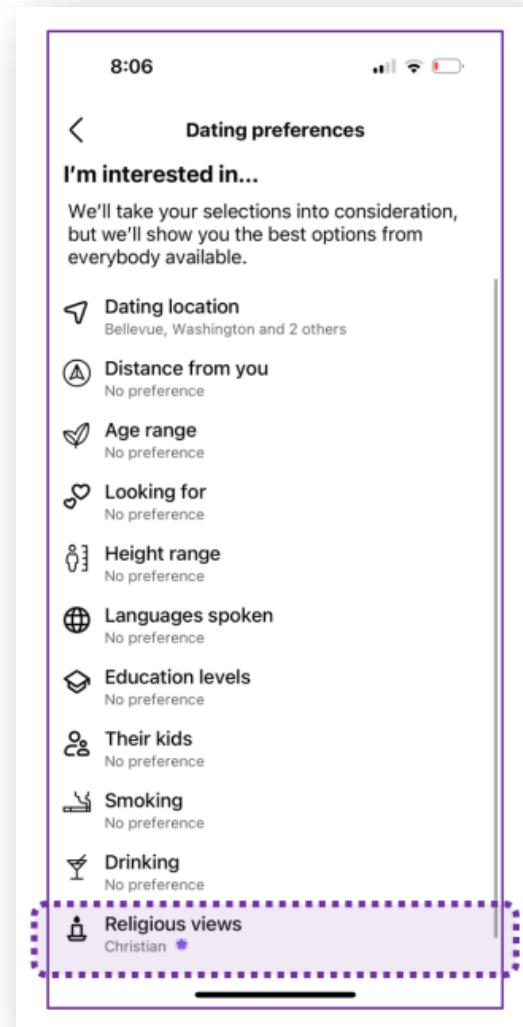
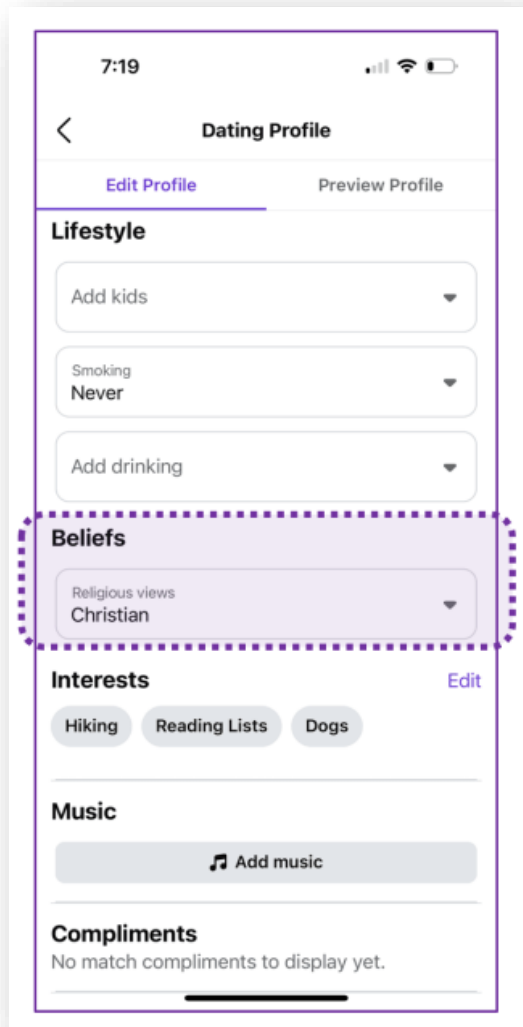
A single policy zone in the warehouse

● Data table - Data dependency



Runtime: exceptions for data flow enforcement

Source: [Facebook Engineering blog](#)



- Religious beliefs are **sensitive** personal data (GDPR) and must not be repurposed, e.g. to personalize Instagram feeds (**purpose limitation**).
- With a data landscape this complex, this could easily happen by mistake without anyone realizing they're violating the law.
- Runtime exceptions enforcing policies about data flows prevent this.

Runtime: exceptions for data flow enforcement

Source: [Diana Marsala, USENIX PEPR 2025](#)

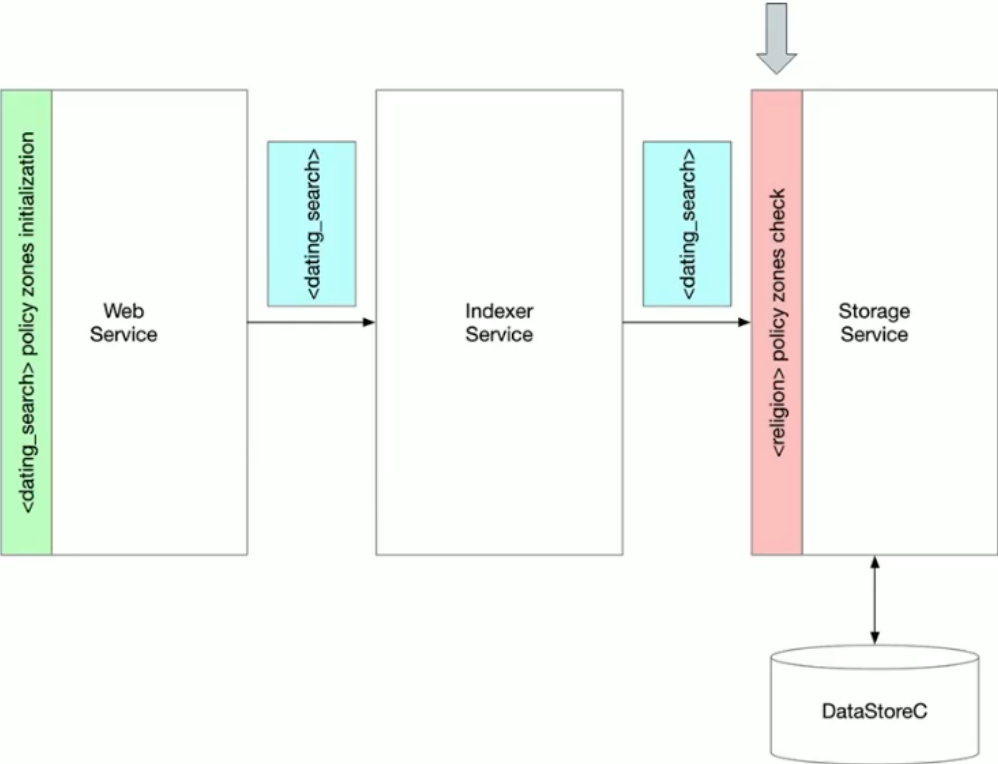
02 WHAT WE TRIED

Take inspiration from security

User wants to filter matches by religion

- 1. Client-facing service initializes Policy Zones with `<dating_search>` purpose
- 2. Propagates `<dating_search>` context through intermediate services
- 3. Policy Zones check runs in storage layer:

can `<religion>` data be used for `<dating_search>` purpose?



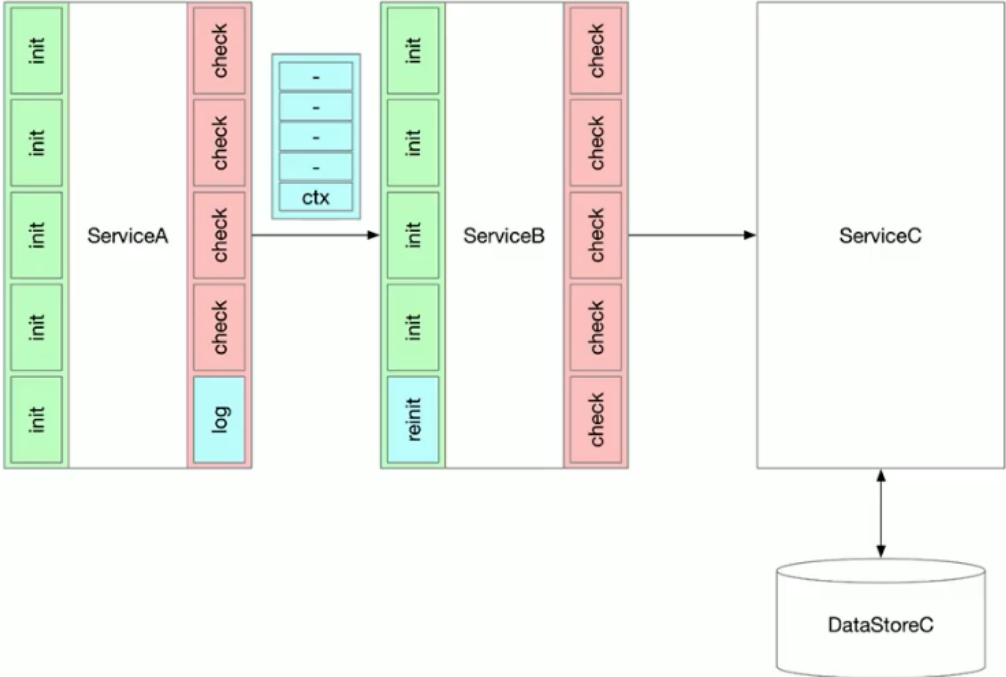
Runtime: exceptions for data flow enforcement

Source: [Diana Marsala, USENIX PEPR 2025](#)

05 A MONOLITHIC POLICY

Create multiple independent swimlanes

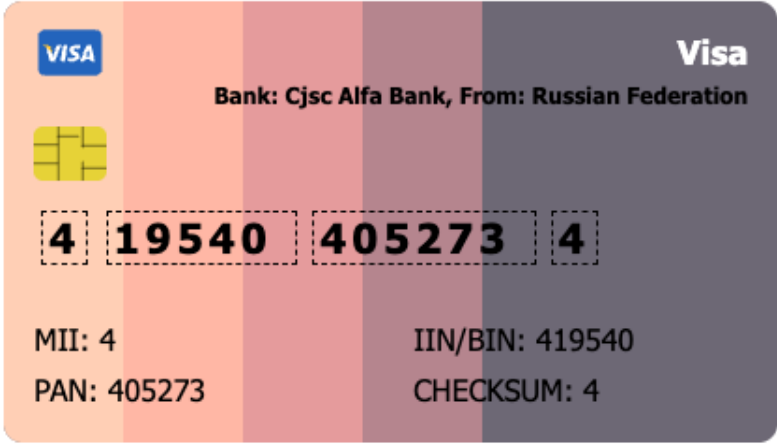
- Requirements are independent at runtime
- Removes alignment bottlenecks
- Moves taxonomizing off the critical path
- Complex features can be limited to the relevant requirements that need them



Runtime: scanning and redaction

```
[StorageManager] 2026-04-22 14:37:39.187 INFO  
{allocMBytes=0.04195404052734375, result=SUCCESSFUL}
```

Luhn Algorithm Check	Passed ✓
MII (Major Industry Identifier)	Banking And Financial
Bank Name	Cjsc Alfa Bank
Bank Country	Russian Federation 🇷🇺
BIN/IIN (Bank/Issuer Identification Number)	419540 🔗
PAN (Primary Account Number)	405273
Network/Brand	Visa
Checksum	4



Runtime: scanning and redaction

False positives and alert fatigue are many times worse than when scanning code repositories for hard-coded secrets, because timestamps, costs, file sizes, ..., can all look like credit card numbers, phone numbers, or IBANs!

This makes **context** essential:

- Are there digits before and after? **False positive**
- Is there a digit *and* decimal point before, or vice versa after? **False positive**
- Is this part of a UUID or other hex value? **False positive**
- Match only relevant regions/brands
- Carefully curate exclusion rules

Runtime: scanning and redaction



Presidio Detection Flow



INPUT: Hi, my name is David and my number is 212 555 1234 ✓

OUTPUT: Hi, my name is <PERSON> and my number is <PHONE_NUMBER>

*NER – Named Entity Recognition


Runtime: scanning and redaction

4 Ai4Privacy 🤗 Hugging Face

 **Datasets:** 4 ai4privacy/**pii-masking-300k**

 World's largest open dataset for privacy masking 

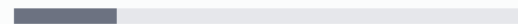
The dataset is useful to train and evaluate models to remove personally identifiable and sensitive information from text, especially in the context of AI assistants and LLMs.

 PII detection report by Presidio

We run the Presidio open source package automatically on a subset of datasets to improve ML creators' level of information.

According to the recognizer model and detection patterns, this dataset may contain the following PII types:

20% of rows may contain emails



0.5% of rows may contain sensitive PII



And as we enter our brave new agentic world...



Jake @JustJake · Apr 24
Oh my. That 1000% shouldn't be possible

We have evals for this. Would you mind DM'ing myself or Mahmoud with info?

3 147

JER @lifeof_jer · Apr 24

An AI Agent Just Destroyed Our Production Data. It Confessed in Writing.

980 2.2K 4.9K 6.9M

A 30-hour timeline of how Cursor's agent, Railway's API, and an industry that markets AI safety faster than it ships it took down a small business

And as we enter our brave new agentic world...



🕒 Response time: 4.31 s 🛡️ Tokens used: 5974 ⚠️ Warn logs: 0 ❌ ERROR logs: 0 ✅ Guardrail activations: 0 📄 View trace

Prompts 3 🔍 Search Prompts ✕ ⏪ ⏩ ⏏ ⋮

📄 System prompt ⏪

👤 Input ⏶

"How can I enhance monitoring of processes started in containers using Podman in Dynatrace? What settings or configurations are available for Podman container monitoring?" 📄

🗨️ Output ⏶

To enhance monitoring of processes started in containers using Podman in Dynatrace, you can leverage the following settings and configurations: ###
Enable Deep Monitoring for Podman Containers 1. **Automatic Injection**: Dynatrace supports auto-injection of code modules into Podman containers for deep monitoring. Ensure that the Podman container monitoring is enabled in the **Host settings** under **Container technologies**. -
Navigate to **Settings** > **Processes and containers** > **Container monitoring rules**. - Add or edit rules to enable monitoring for Podman

AI observability is essential, but remember that tool inputs, prompts, and outputs can be sensitive!

And as we enter our brave new agentic world...



Langfuse Demo / langfuse-docs / Traces

Trace: eb346ba8-19d9-4368-9b71-022cd17f25d7 with-context

Search (type, title, id) Timeline qa ID + Add to datasets Annotate

2025-07-16 23:44:16.449
Session: lf.docs.conversation.s7A0W7I User ID: u-M9QTUzj Env: default Latency: 11.58s Total Cost: \$0.000371
1,313 → 291 (Σ 1,604) Release: 044d0162fe3dbddd09f88a554521ddf73726618c

Formatted JSON

Preview Scores

Input View as markdown

"How to update the input and output in langgraph traces?"

Output View as markdown

"To update the input and output in Langfuse traces, it's essential to understand that Langfuse allows meticulous tracking of steps within the AI process by logging both inputs and outputs of large language models (LLMs). This means that every interaction you have can be captured and displayed transparently within the UI.

Here's how you can effectively manage and visualize these updates:

Steps to Update Traces

1. **Log Inputs and Outputs**: Ensure you're logging the inputs and outputs of

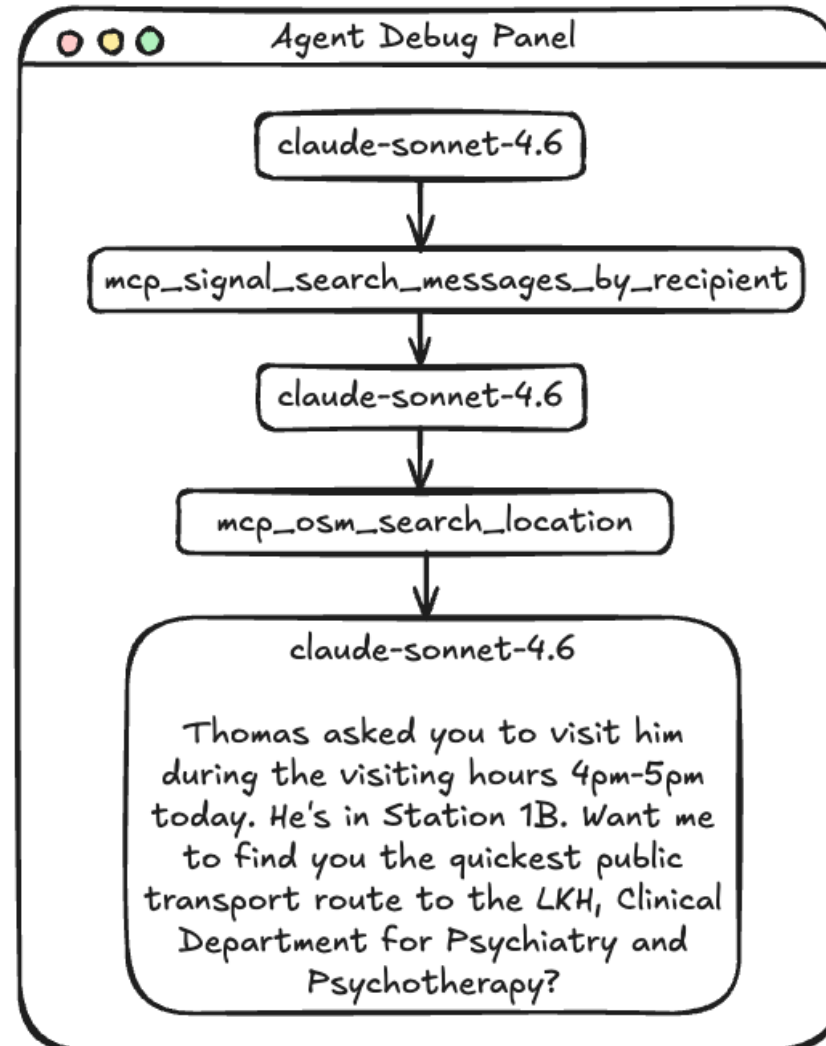
...expand (989 more characters)

"

retrieval 1.55s Σ \$0.000001

- prompt-embedding DEBUG 0.67s 12 → 0 (Σ 12) \$0.000001
- vector-store 0.69s
- context-encoding
- fetch-prompt-from-langfuse 0.47s
- generation 9.57s 1301 → 291 (Σ 1592) \$0.00037

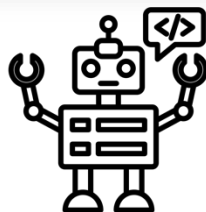
And as we enter our brave new agentic world...



And as we enter our brave new agentic world...

Leverage agentic AI to prevent sensitive data exposure, but don't forget to consider how to prevent it *causing* yet more sensitive data exposure!

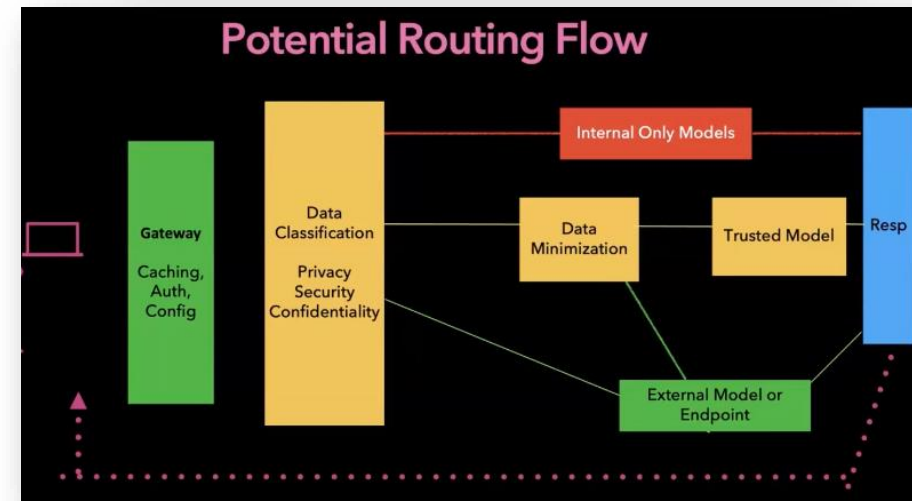
```
SKILL.md
---
name: log-securely
description: Reviews log calls for sensitive data exposure. Applies when log calls are added or modified, and when asked whether specific data is safe to log.
---
```



 [ai-and-ml-security-minicourse](#) Public

A Probably Private mini-course introducing AI/ML security via interactive videos and hands-on examples.

 Jupyter Notebook  28  7  [kjam](#)



Key takeaways

- Context matters!
 - Even the term “sensitive data” is contextual
- Max ROI for controls
 - Organizational: train people to consider the context
 - Technical: invest at the design and build stages
- Get the basics right now to prepare for agentic data flows

Thanks for listening!

Attribution

- **Dynatrace content and branding:** © 2026 Dynatrace LLC
- **Third-party logos, images, text, and videos:** see links for attribution
- **Unattributed images:** used under license from the [Noun Project](#)
- **All other content:** original work by the author, may be reused with attribution



CLOUD DONE RIGHT