CASE STUDY

110,000 FILES, ONE MISCONFIGURED BUCKET — HOW AI-POWERED BREACH RESPONSE KEPT A RETAIL COMPANY CCPA COMPLIANT IN 40 DAYS

Incident Summary

Our client Retail/ E-commerce company In April 2024, discovered that a misconfigured cloud storage bucket (Amazon S3) had been publicly accessible for several months. The exposed data included archived customer service chat logs, internal Excel reports, return processing documents, and customer registration forms.

Upon internal investigation, the company identified that files containing personal data of California residents may have been accessed by unauthorized third parties.

Objectives

- Identify all documents containing **personal information (PI)** as defined under CCPA
- Detect and validate the **affected California residents**
- Determine the **scope of the notifiable breach**
- Notify impacted users within the **45-day timeframe**
- Maintain a **defensible audit** trail for potential AG review

Review Workflow: Canopy AI + Manual Extraction

Data Ingestion & Pre-Processing

- Approx. **110,000 documents** (CSV exports, PDFs, chat logs, Excel sheets) pulled from misconfigured bucket
- OCR applied to scanned return forms and PDFs
- Uploaded into **Canopy AI's breach response platform** for analysis

Canopy AI – Automated Entity Discovery

Canopy scanned for CCPA-defined personal identifiers:

- Full names
- Phone numbers
- Email addresses
- Shipping addresses
- Credit card last 4 digits
- Customer support messages (which sometimes included usernames/passwords)
- **Identity Stitching Engine** grouped exposed data points to individual records
- 27,500 documents tagged with **probable personal data**

Manual Extraction & Validation

Human reviewers manually validated:

01 **Low-confidence AI hits** (e.g., names in marketing templates or HTML headers)

02 Redundant or templated documents (such as duplicate invoices

03 Embedded PI in chat logs—where customers typed credit card details or passwords

04 Documents where **AI misclassified placeholders** (e.g., test data)

Reviewers also validated California residency using:

01 ZIP codes and state fields

02 Shipping address mentions

03 Area codes (for partial phone numbers)

Final tags applied:

01 "Contains PI"

02 "CCPA Covered (CA Resident)"

03 "Notifiable"

Reporting & Notification

- Canopy dashboard generated:
  - ✓ Total count of affected CA residents: **5,800**
  - ✓ Notifiable consumers: 4,100 (PI + access confirmed)
  - ✓ High-risk indicators: password exposure, credentials reuse
- Data exported into structured CSV reports with individual profiles for notification letters
- Coordinated with a breach notification vendor to distribute customized alerts
- Notices delivered within **40 days of discovery**

Results Overview

Metric	Value
Total Files Reviewed	110,000
Documents with PI	27,500
CA Residents Identified	5,800
Notifiable under CCPA	4,100
Deadline Met	Yes (within 45 days)
Audit Trail Maintained	Via Canopy platform

Key Takeaways

- **Canopy's AI reduced triage time** by 70%, automatically identifying most PI with accuracy
- **Manual review remains critical** for interpreting contextual data like open-text chat logs and error-prone scanned documents
- **State-specific filtering** helps isolate obligations under CCPA versus other jurisdictions (GDPR, etc.)
- Use of **identity stitching** ensured duplicate entries were consolidated under single consumer records
- Clear audit logs enhanced **regulatory defensibility** and potential future risk assessments

CCPA Compliance Highlights

- Consumer Notification within 45 Days
- Audit Logs Maintained
- Transparent Consumer Access Provided Upon Request
- Ability to Track Individual Notification Scope & Content

Conclusion

This case demonstrates how a scalable, **AI-powered breach response platform like Canopy,** combined with targeted **manual** validation, can help retail and e-commerce companies meet CCPA breach response obligations quickly and defensibly. The hybrid approach protected the affected customers, preserved compliance posture, and mitigated reputational and regulatory risk.