Lab 1

2ND EDITION / 2026

The Anatomy of a Breach



Contents

Foreword 3

Key findings 5

Executive Summary 8

Contents of a Breach 9

- File Classification 10
- File type 12
- Sensitive Information Types 14

Sizing up a Breach 15

- Blast radius 17
- The Limits of Incident Volume 19

The State of Threat Actor Data Analysis Capabilities 23

- Script Extraction 26
- Human Analyst Teams 27
- AI / LLM 28
- Indexing-as-a-Service 31

Escalating Risks in Data Breach Weaponization 34

Understanding the Anatomy of a Breach 36

Foreword

How cyber criminals extract maximum damage from data breaches is entering a new, AI-driven era. Criminal techniques for analyzing and indexing stolen data are becoming more structured, more specialized and more commercial. Raw breached datasets are no longer simply dumped online or used as blunt extortion leverage. They are being organized into buyer-specific products: quarterly reports for market abuse, financial records for fraud, share buyback data for insider trading, and supplier information for follow-on compromise. AI is accelerating that shift by helping threat actors make sense of large breached datasets and identify where pressure points and resale value lie.

This has three immediate consequences. Extortion becomes more targeted as threat actors speed up the identification of regulatory, reputational and commercial pressure points in stolen data. Breached datasets become easier to repackage and resell for follow-on criminal use. Personal exposure can also be isolated and used to apply pressure to executives and other named individuals.

Regardless of the approach, there's a common thread. The once high barrier to weaponizing breached datasets has collapsed. The time required to manually review files and familiarity required to identify which were damaging acted as that high barrier. AI means that's yesterday's status quo.

We return with our second ***Anatomy of a Breach*** report to present the largest published content-level analysis of data breach incidents available to the market, at a time when the scale of exposure is growing significantly.

Compared with last year's report, the average breach analysed in 2026 contains 44.5% more files and 121.7% more data, giving threat actors a much larger body of material to search for the financial, regulatory, commercial and personal pressure points that can be used for extortion, fraud and follow-on attacks.

Boards and CISOs reckoning with the cyber security risks AI poses to their business can't afford to ignore its findings. The exposure of data ripe for weaponization is widespread, and the real-world threat actor activity cited in this report shows threat actors are seizing the opportunity.

At a moment of widespread concern about AI-enabled attacks and the discovery of new vulnerabilities, the Anatomy of a Breach offers a reality check on where attention should focus. Lab 1 wants to see data risk neutralised and we hope this report can be a force for good in that mission by helping organizations understand this risk and offering them a solution.

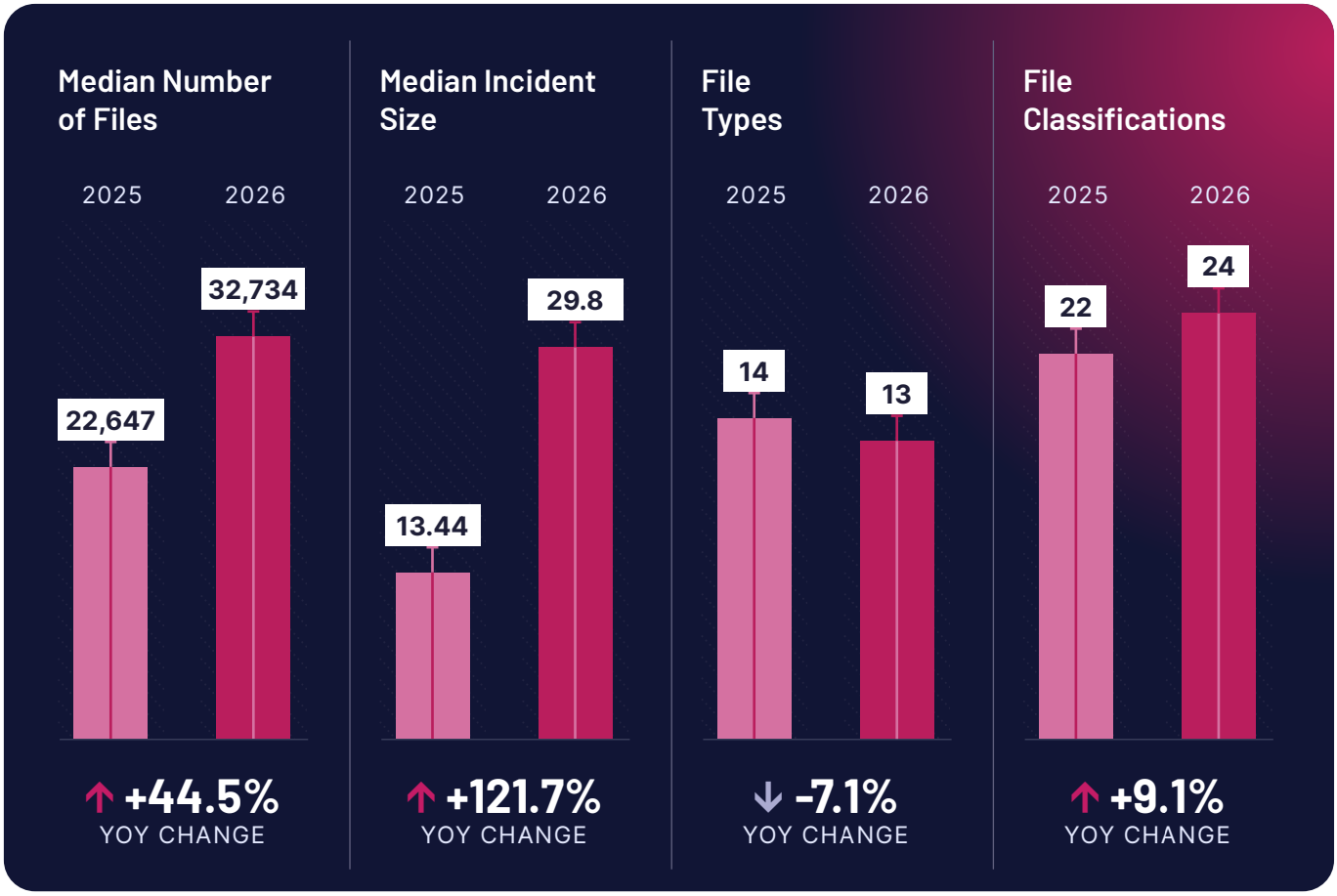
I encourage you to read this report not just as a benchmark, but as a call to action. Organizations need to understand the true depth of their exposure through breached datasets, and whether they are equipped to respond to emerging techniques that make stolen data easier to weaponize.



Robin Brattel
Co-Founder & CEO, Lab 1

The Anatomy of the Average Breach: A Year Later

Compared with the medians reported in 2025, the breaches in the expanded 2026 dataset contained substantially more material: the median breach held **44.5% more files** and **121.7% more data**. Yet the number of file types and classifications remained broadly consistent.



Several factors may help explain this increase. A number of prominent threat actor campaigns have targeted specific parts of the supply chain, including sales and audit tooling, where a single compromise can expose large volumes of data from multiple organizations. Some groups also appear focused on increasing the volume of breaches they can claim and publish.

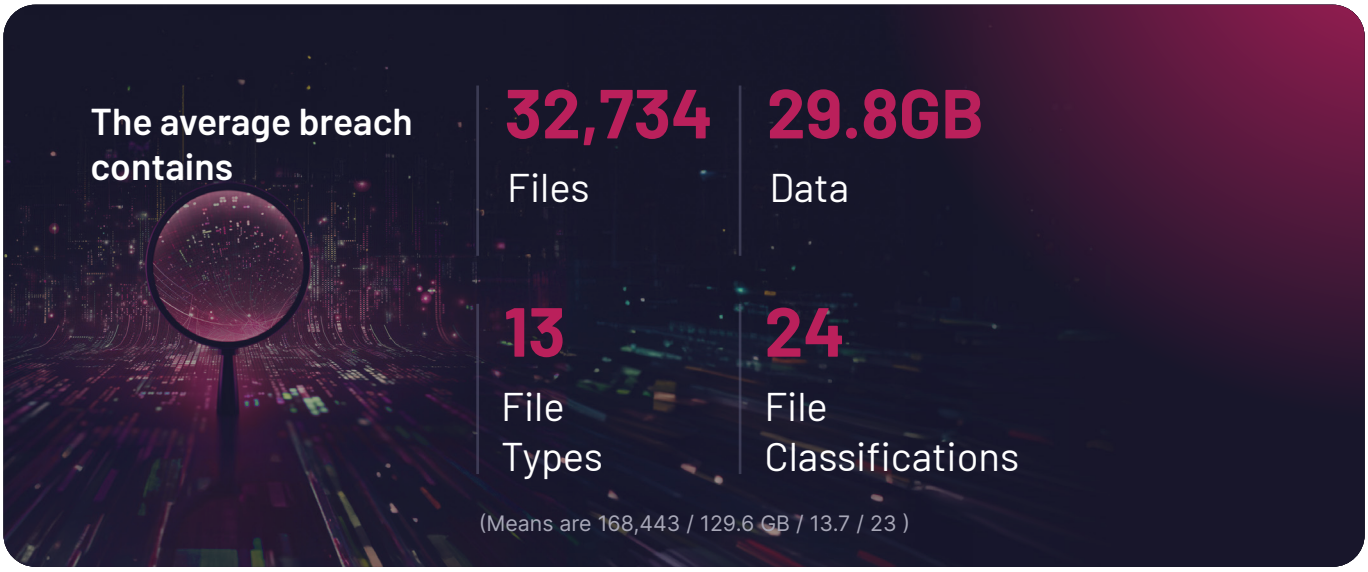
At the same time, AI may be enabling threat actors to review much larger datasets and identify valuable material faster, accelerating the weaponization of stolen data and reducing the need to be surgical in what they exfiltrate.

To take an example, in the run up to the publication of this report, the FulcrumSec threat actor group confirmed its use of an AI model to summarize datasets it steals and ranks the overall sensitivity of their contents, with separate AI agents used to create a deeper analysis of those datasets that rank high in sensitivity.

Organizations must face up to this new reality.

Key findings

The key findings are derived from analysis of **347 million unstructured files** across more than 2,065 data breach incidents from 2022 to June 2026, using Lab 1's AI Exposed Data Intelligence Application and supporting research from Lab 1's threat intelligence team.



Sustained exposure of data that fuels cybercrime

95%

of incidents contain financial data
(~145 million files)

84%

of incidents contain HR data
(~29 million files)

The largest analysis of breach contents to date confirms a pattern observed in the analysis of last year's inaugural Anatomy of a Breach report: the widespread exposure of financial and HR information in data breaches.

Financial data creates immediate fraud risk. Bank details, invoices, payroll data and tax information can support invoice scams, payment redirection, mandate fraud and financial impersonation.

HR data creates longer-term risk because it does not expire. Employment histories, salary data, identity documents and performance records can be reused for synthetic identity fraud, deepfakes, voice cloning and highly personalized social engineering.

The median breach brings far-reaching exposure

32,734 files

The median exposed contents in an incident

13,436

unique organizations exposed in incidents between 2025 and 2026

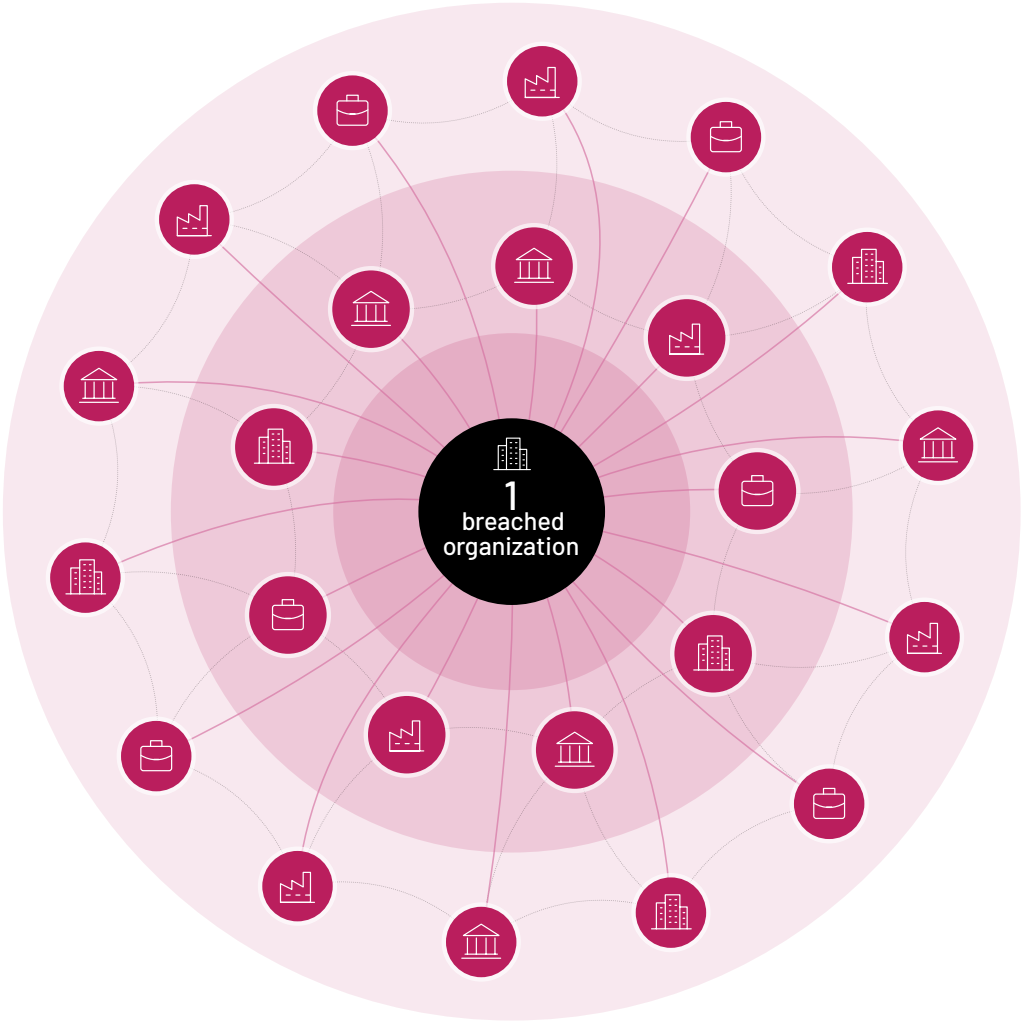
94%

of breaches exposed at least one organisation beyond the one breached

The impact of a breach is almost never confined to a single organization. Breached files compromise customers, suppliers, partners and other third parties that weren't subject to a breach themselves. A web of connections between incidents, suppliers and customers means an incident's impact always reaches beyond the victim in the headlines.

542

organizations beyond an initial victim exposed in 2026's median data breach



Threat actors are getting better at exploiting stolen data

1 > 7

From one (2024) to seven threat actor groups carrying out structured analysis of breached data

400,000 archived emails

mined to extract the 61 emails with the highest regulatory and reputational risk to a victim

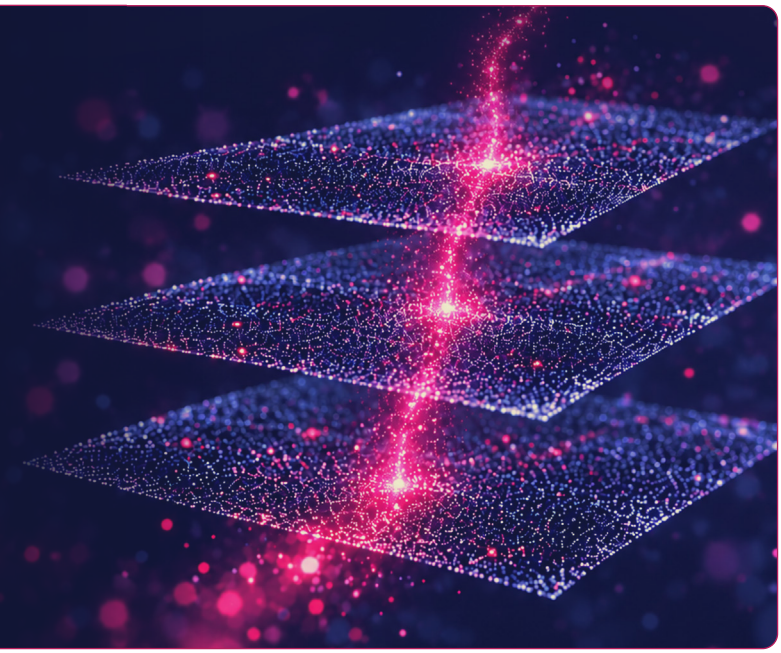
700,000 files

mined with AI agents to extract sensitive data including patient records

Until recently, extracting value from large volumes of stolen corporate data required significant time and effort. Terabytes of files had to be manually sorted, analyzed and connected before threat actors could identify sensitive information, uncover relationships or determine the most effective avenues for exploitation. That barrier is rapidly disappearing.

AI tools, indexing services and specialist analyst teams now help threat actors turn stolen data into usable intelligence for follow-on attacks. They reduce the “weaponization tax” by doing the work that once sat between a raw data dump and criminal use: finding sensitive files, categorizing exposure, identifying relationships and packaging the material for pressure or resale.

AI makes stolen data easier to weaponize.



Executive Summary

The largest published content-level analysis of data breach incidents to date confirms the widespread exposure of sensitive information. But the crucial development in 2026 is not simply what is exposed. It is what cyber criminals are now doing with that information after organizations are breached.

The contents of a breach have taken on new importance as the barriers to analyzing, organizing and exploiting stolen information fall rapidly. The Anatomy of a Breach sets out what is exposed in breached datasets and pairs this with evidence of increasingly mature threat actor data analysis capabilities drawn from contemporary data breach incidents.

This shift challenges assumptions that still dominate corporate thinking around cybersecurity. Breach size alone is a poor indicator of severity: a relatively small collection of emails, invoices, contracts or source code can pose a greater threat than a much larger dataset made up of lower-value records. Nor can organizations assume a breach affects only the named victim.

The year-on-year comparison also provides important context on the volume of material contained within the incidents analyzed. The average breach in the 2026 dataset contains more files and data than the averages identified in the 2025 report. While this does not establish that breaches overall are becoming larger, it shows that the incidents examined this year contain substantially more material for threat actors to search, organize and exploit.

Barriers to weaponize stolen data are disappearing

The blast radius of the average breach is too large for companies to rely on instinct when assessing whether they may be exposed. This year, the median breach has impacted 542 organizations beyond the company that was breached.

Crucially, this year's ***Anatomy of a Breach*** shows how advances in automated analysis are making exposure easier for criminals to identify, structure and exploit at scale. Increasing numbers of threat actor groups are using scripted extraction tools, dedicated analyst teams, AI-assisted review and specialist indexing services to transform stolen data into actionable criminal opportunities.

This development is changing the economics of data breaches. The investment required to extract value from stolen data is falling, while the potential payoff for threat actors is rising. The report describes this as the shrinking of the "weaponization tax": the time, investment and effort required to turn raw stolen data repositories into money-generating next steps.

This year's report therefore makes the case for boards and security teams to assess data breach exposure by the contents of what has been exposed and how that information could be exploited – not just by volume of data involved. That mindset shift is becoming more urgent as criminals adopt sophisticated ways to analyze and monetize stolen information.

This report positions data breach incidents as the starting point for follow-on criminal activity, rather than narrow security events defined by a breach timeline. Organizations need content-aware breach responses that assess exposure down to the file level, identify who and what may be implicated, and anticipate how criminals could use what they now know.

Contents of a Breach

Last year's inaugural Anatomy of a Breach report was based on an analysis of 1,297 data breaches. This year's report is based on a much larger analysis – over 2,000 incidents and a 133% increase in total files ingested and classified.

File Classification →

File Type →

Sensitive Information Types →

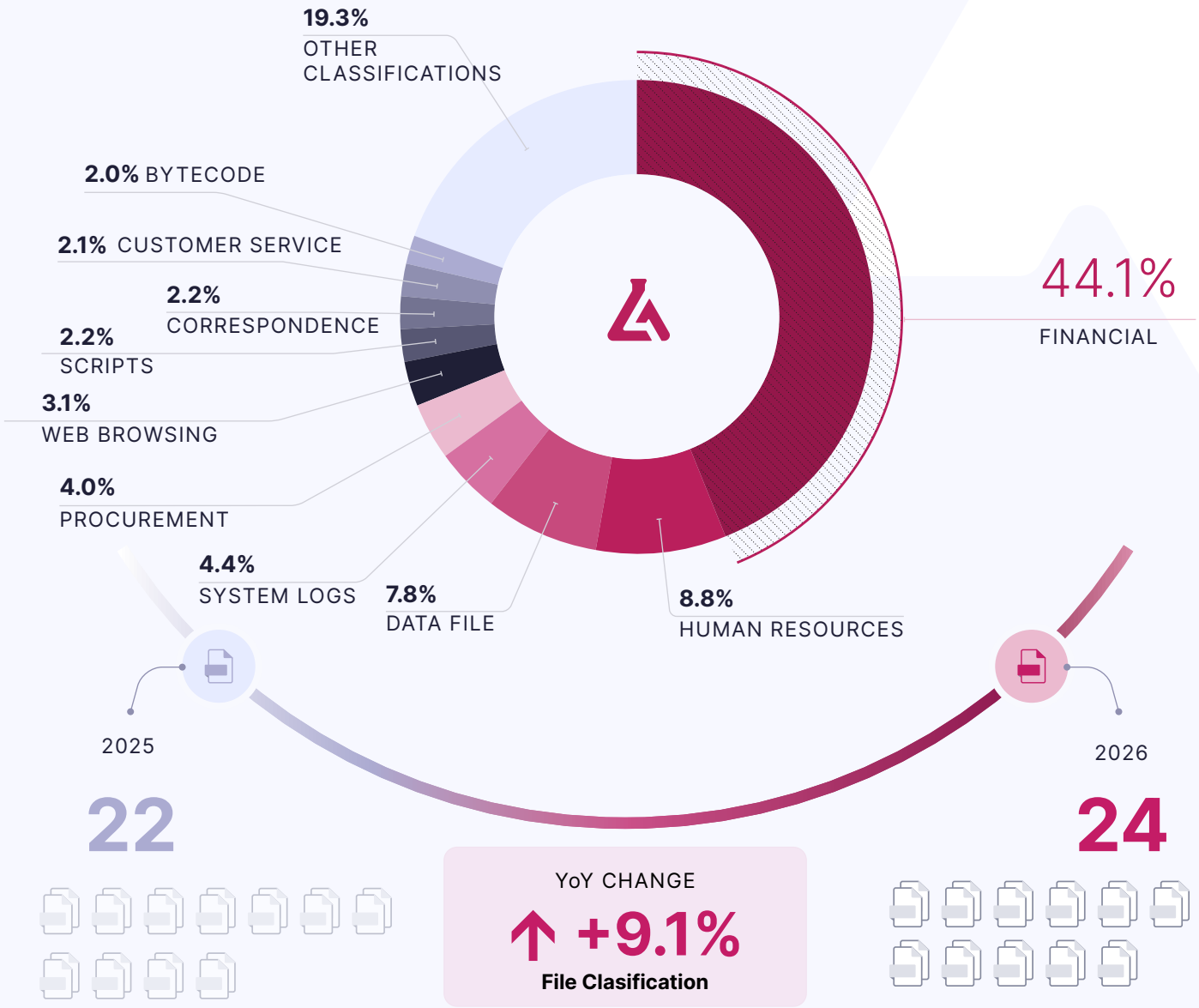


File Classification

An updated evaluation of the contents of a breach confirms widespread exposure of high-risk content types scales as more breaches are analyzed.

Figure 1:
Top 10 File Classifications Most Commonly Identified in Data Breaches

Percentage of the over 2,000 analyzed data breaches that include each file classification



The classification of an exposed file (Figure 1) determines how it can be weaponized once in the open. The most commonly exposed file classifications pose unique risks.

44%
of all breached files were financial data

What's Inside

Often containing banking details, tax identifiers, or invoicing information, financial documents.

The Threat Implications

A primary source of material for invoice fraud, payment redirection and financial impersonation and resale on dark web markets.

9%
of all breached files were HR data

What's Inside

Often containing PII, payroll records, employment contracts and identity documents.

The Threat Implications

Narrative-rich data that's well suited to AI-enabled generation of synthetic identities, deepfake content, or voice-clone phishing attacks.

4%
of all breached files were system logs

What's Inside

Metadata that reveals crucial insights into system behavior, user activity and environmental configurations.

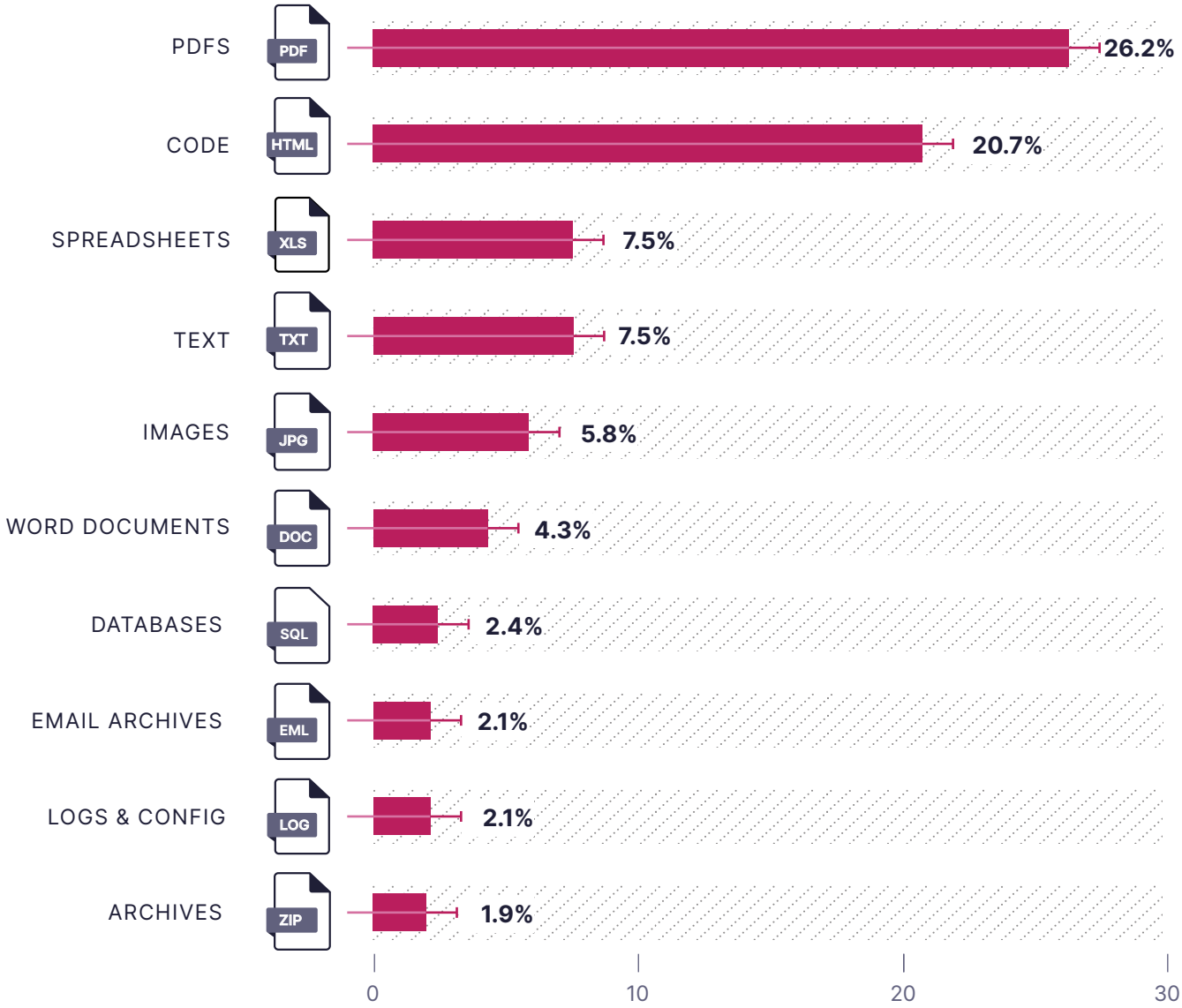
The Threat Implications

Enables attackers to map infrastructure, detect vulnerable endpoints and identify misconfigured services.

File type

Figure 2:
Top 10 Most Common File Type of Total Files Analyzed

Percentage of the 328 million breached files analyzed that belong to each file type



File types present in a breach reveal how easily an attacker can discover and weaponize what has been stolen. Common file types point to broad exposure of sensitive corporate information, while less common file types can pose acute risks because they may contain credentials, configuration details or other material that supports follow-on attacks.

26%
of all breached files were PDFs

What's Inside

The standard container for the content categories that carry the greatest post-exposure risk: financial statements, employment contracts, HR records, invoices, legal correspondence and compliance documents.

The Threat Implications

A breach that surfaces a large PDF archive is, in practice, a breach of every classification those documents contain and the follow-on attacks they enable. PDFs are where the most commercially sensitive material is most likely to reside.

8%
of all breached files were text files

What's Inside

Text files are a common vector for credentials, configuration notes and data dumps.

The Threat Implications

Immediately readable information for hackers to probe and gather how systems are configured and establish information for social engineering attacks. Often contain structured records from third-party platforms that expose customer, supplier and commercial relationships to act as a roadmap for supply-chain attacks and targeted fraud.

2%
of all breached files were email archives

What's Inside

A complete record of how an organization communicates: its tone, its relationships, its decision-making language.

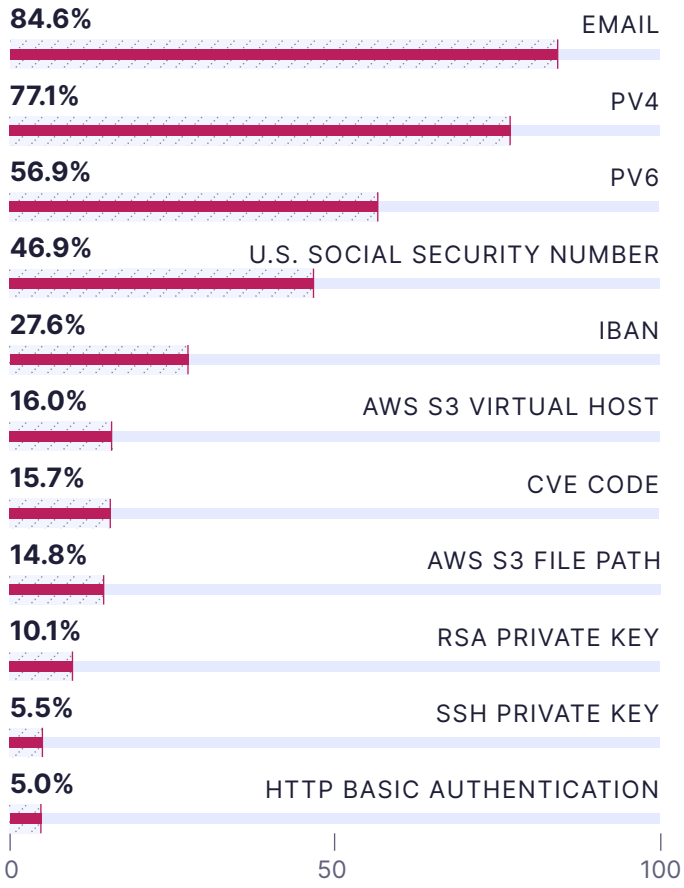
The Threat Implications

A gold mine for AI-driven phishing campaigns; presents a level of risk disproportionate to their total share of files.

Sensitive Information Types

Figure 3:
Top 10 File Types Most Commonly Identified in Data Breaches

Percentage of the over 2,000 analyzed data breaches that included the sensitive information type



The common exposure of sensitive information (Figure 3) identified in 2025's Anatomy of a Breach report remains true in this year's larger analysis. Taken as a whole, one takeaway is that Personally Identifiable Information (PII) is widely exposed in breaches: email addresses, U.S. Social Security Numbers, bank statements and wealth statements can all be useful to AI-driven social engineering attacks. Sensitive data has been exposed for years, but it is becoming easier to weaponize and harder to ignore.

Another takeaway is that sensitive data exposure can undermine the security posture of organizations caught up in breaches. Private security keys, credentials and infrastructure indicators may appear less frequently than personal identifiers, but their impact can be disproportionate. While these assets are often buried within vast datasets, AI-powered analysis can now rapidly identify, isolate and surface them for exploitation. The exposure of cryptographic private keys, seen in approximately 15% of incidents combined, can enable an attacker to bypass authentication, access secure systems, and create the conditions for lateral movement, credential forgery, re-issue or escalation.

The implications: what's changed in 2026

The contents of a breach matter more in 2026 because the effort required to weaponize them is falling. AI is increasing threat actors' ability to process breached datasets at speed and scale.

This creates new risks around legacy data. A cryptographic private key or email archive that may have been overlooked in a 2024 incident could now be surfaced through mature analysis of breached data. For business leaders and security teams, the working assumption should be that exposure may extend beyond the named victim and into its wider network.

Sizing up a Breach

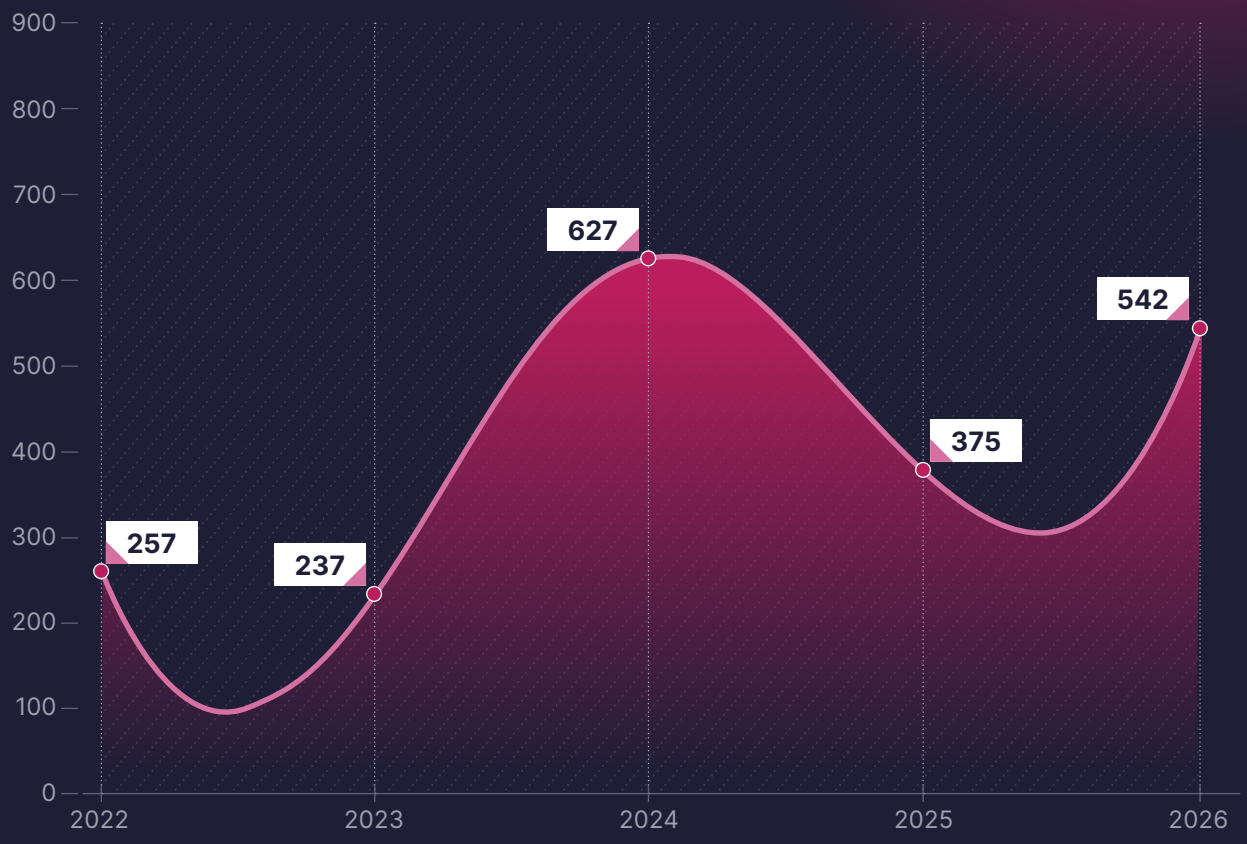
The exposure of multiple organizations within a single data breach, often referred to as the “blast radius”, remains one of the most important measures of a breach’s impact.



However, understanding the median blast radius is only part of the picture. Different industries exhibit different levels of downstream exposure, and factoring in threat actor behavior is critical to understanding how a blast radius manifests into impact. This section brings these factors together to size up a breach with the fuller picture.

Figure 4:
Median Blast Radius over time

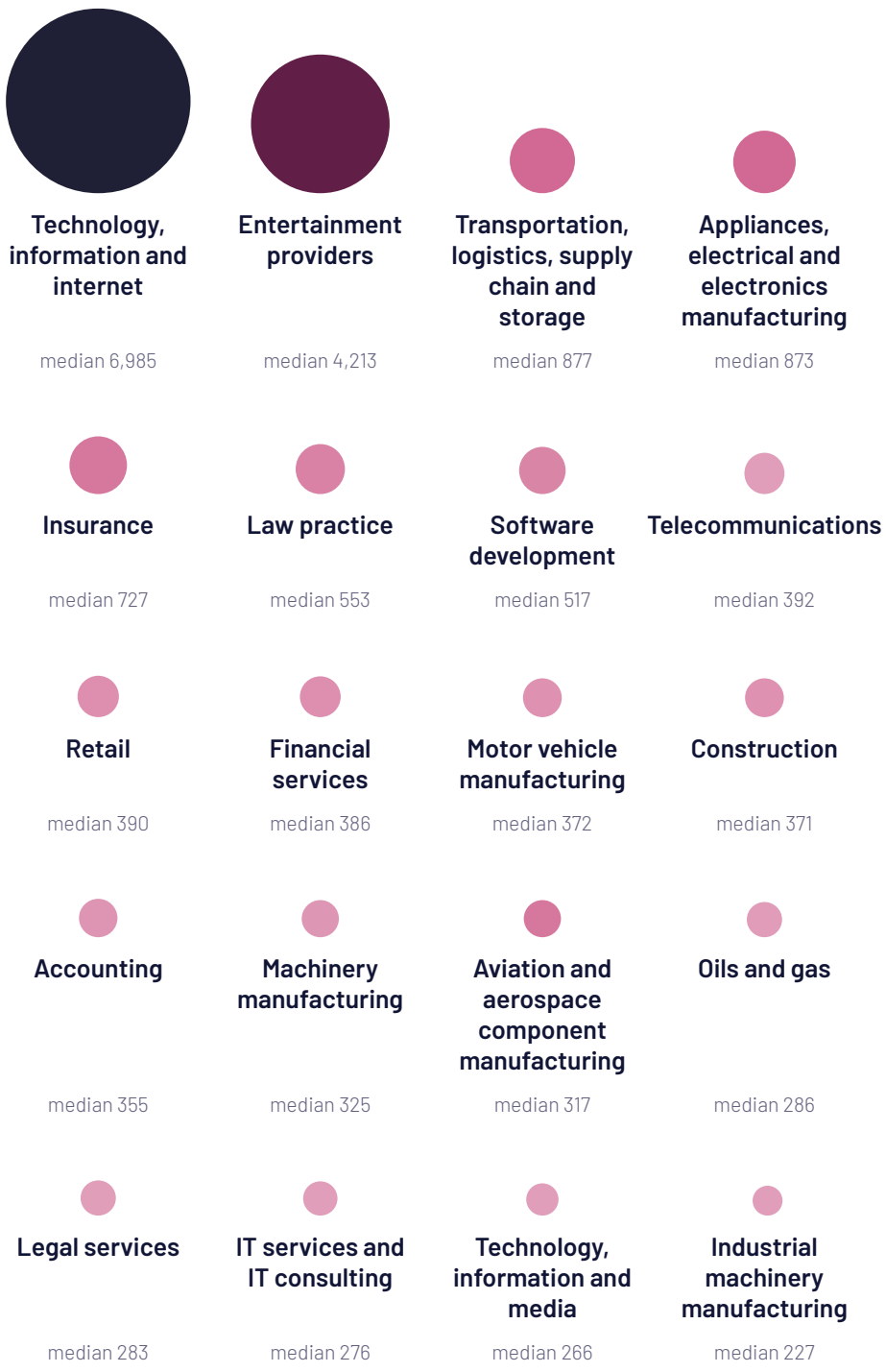
Median number of unique organizations exposed per incident by calendar year since 2022.



Measured year by year, the median number of organizations exposed per breach has moved in wide oscillations (Figure 4). The 2026 figure is based on a partial year and is likely to rise as more exposure records are identified. But the wider pattern is already clear: in every year measured, the typical breach has exposed data belonging to hundreds of organizations. A breach may have one named victim, but its impact rarely stops there.

Figure 5:
Median Blast Radius by Industry

Median number of unique organizations exposed in breaches by breach victim industry, with circle size representing each industry's incident count and a color code to visualize blast radius by industry.



Technology, Information and Internet:

When a platform, cloud service or internet infrastructure company is breached, its clients, partners and users are exposed alongside it without an independent indicator that they have been touched. This explains the sector's outsized comparative blast radius.

Entertainment Providers:

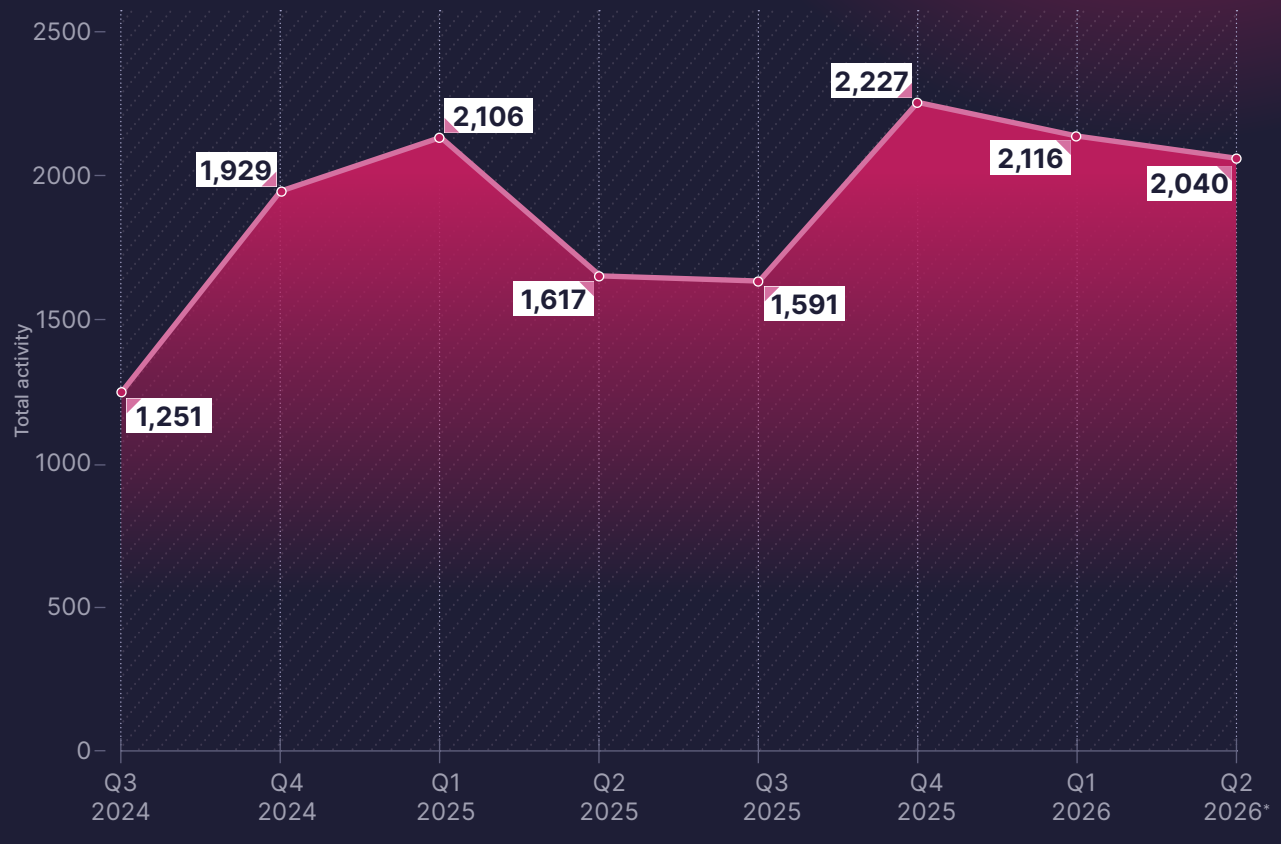
The same dynamic appears here, with 4,212 organizations exposed per incident. Customer and partner data is often concentrated inside a small number of large distribution platforms.

Across the top twenty sectors by incident count, the median blast radius spans a thirtyfold range (Figure 5), from 227 organizations in industrial machinery manufacturing to 6,985 in technology, information and internet. The sectors with the highest median blast radius share common characteristics: varied and complex supply chains, a reliance on knowledge work rather than physical operations and a tendency to hold data on behalf of others.

However, it's essential to weigh the headline median blast radius by industry against the incident count (represented by circle size). Incident volume and reach do not move together.

Figure 6:
Threat Actor Group Data Breach Responsibility Claims by Quarter

The number of data breach responsibility claims quarterly since June 20th 2024

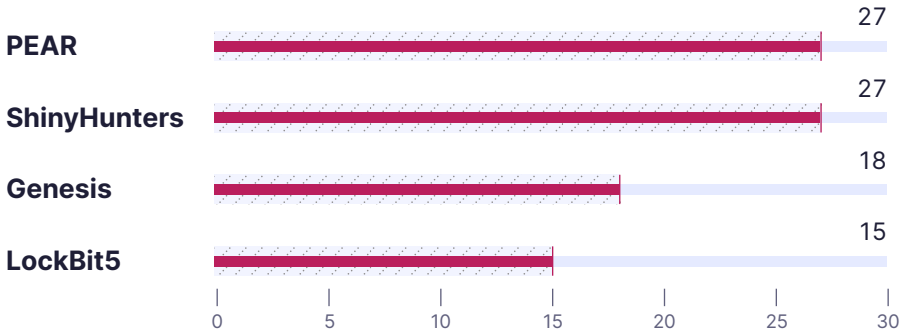


* Partial quarter:
Q2 2026 = 11 weekly observations.

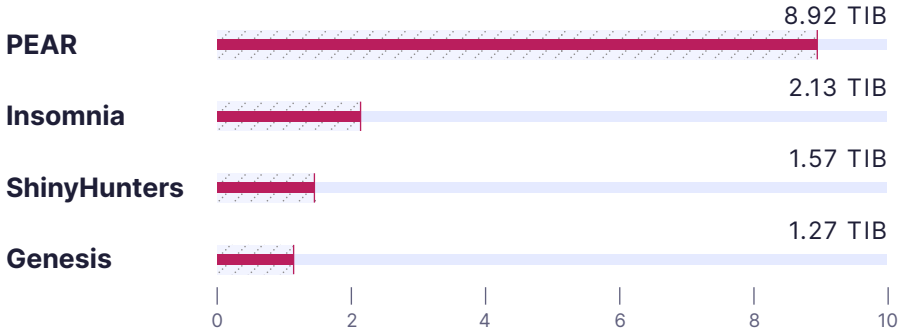
Mapping out the frequency of breaches is essential context to size up breaches and the average number of weekly claims from threat actors responsible for data breaches has trended upwards over the last two years. However, **claim count is one measure of activity. It is not a measure of danger.**

We see more complex dynamics play out in a focused sample of 130 incidents associated with 17 groups that launched or established a new data-leak site between June 2025 and June 2026. Attempts to rank the threat these groups are complicated by various criteria.

By number of incidents

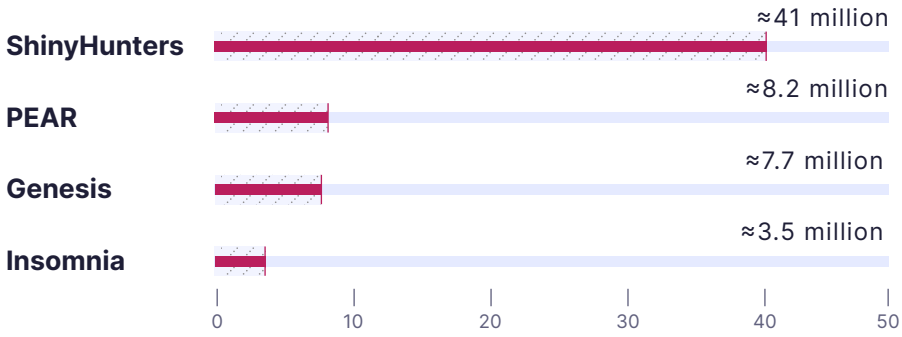


By amount of data stolen



By exposed data entities

(individual email addresses, SSNs, credentials etc)



In this focused sample set by claim count alone, PEAR appears highly active. Shift focus to exposed data entities, however, and PEAR moves to second, while ShinyHunters ranks first. The key insight is that threat actors are developing different data profiles, and incident volume alone does not capture the severity of exposure.

Claim count also reveals nothing about specific data exposure, which is the key concern for the massive number of organizations exposed across this sample of 130 incidents.



Broken down by file classification

46%
Financial documents

11%
HR data

7%
Correspondence

5%
Web browsing

The disproportionate targeting of financial and HR documents seen across all breached datasets analyzed for this report is seen across this sample of 130 incidents spanning 2025 and 2026. It highlights how cyber criminals directly target high-value information.

The tendency to target high-value information is also seen when the 130 incidents are broken down by responsible threat actor groups.

Figure 7:

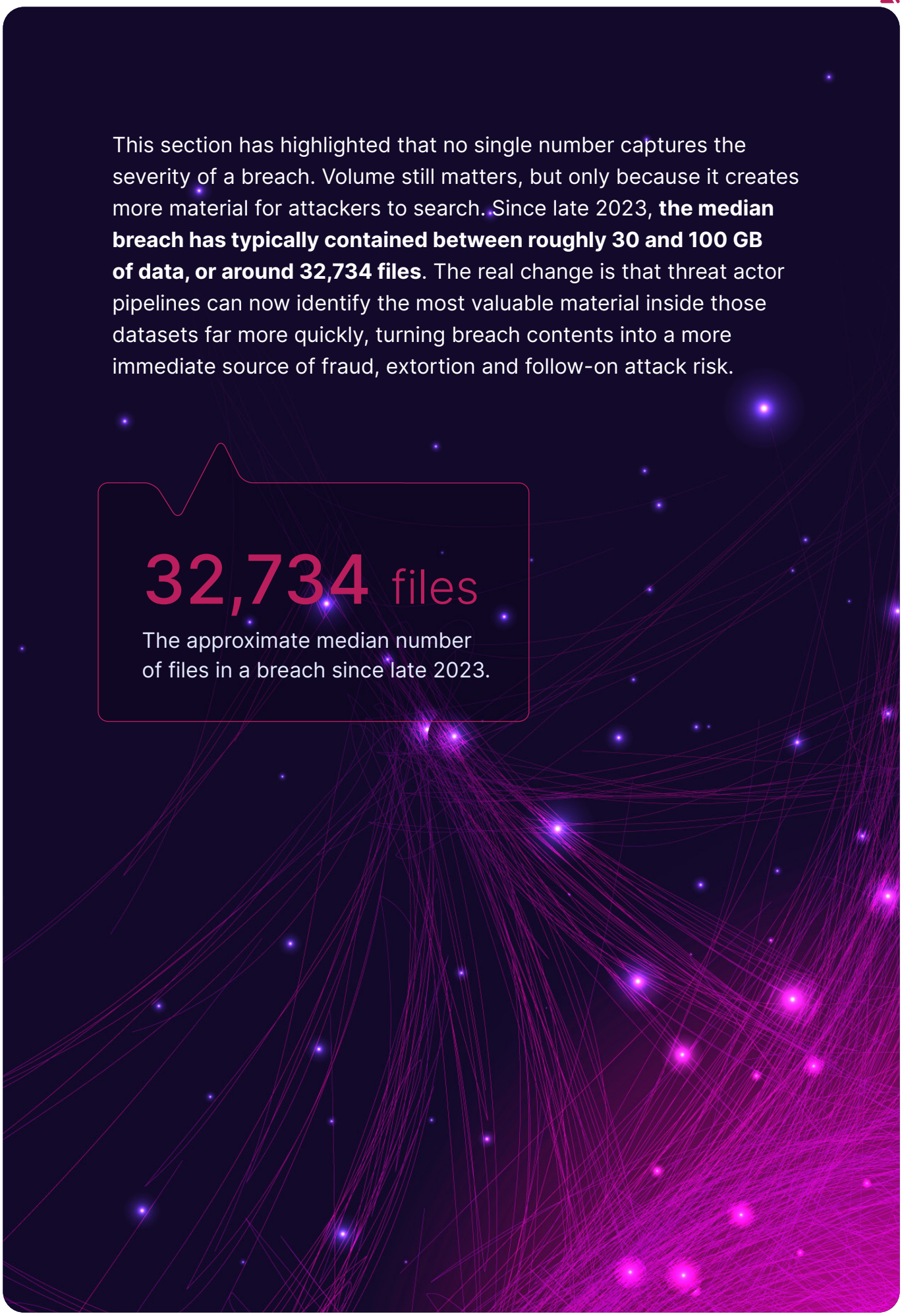
Threat Actor Groups' Top Five Exposed Data Categories

Each threat actor group's five highest-concentration data categories across the analyzed incidents, ranked in source-data order with original percentages.

25% and over | 10 - 25% | 5 - 10% | <5%

Threat Actor Group	Highest-Concentration Category	Second Highest	Third Highest	Fourth Highest	Fifth Highest
Ailock	Billing (66.6%)	Financial (18.7%)	Script (2.0%)	Procurement (1.6%)	Invoices and Receipts (1.5%)
alp-001	Web Browsing (29.2%)	Source Code (17.5%)	Script (13.4%)	Query File (11.4%)	Web Page (8.6%)
Aurora	Financial (40.3%)	Customer Care (27.6%)	Human Resources (10.0%)	Legal (5.7%)	Logistics (2.5%)
Brotherhood	Financial (57.8%)	Data File (26.6%)	Invoices and Receipts(3.5%)	Human Resources (3.1%)	User Manuals (2.3%)
CoinbaseCartel	Financial (24.8%)	Source Code (24.2%)	Bytecode (19.4%)	Script (17.1%)	Data File (3.4%)
Genesis	Financial (62.9%)	Human Resources (5.1%)	Customer Care (3.4%)	Web Browsing (3.2%)	Source Code (3.1%)
Insomnia	Financial (38.9%)	Correspondence (37.5%)	Office Management (5.9%)	Human Resources (3.1%)	Sales (2.5%)
Kawa4096	Financial (39.3%)	Procurement (13.1%)	Customer Care (12.0%)	Invoices and Receipts (5.6%)	Human Resources (5.4%)
Lamashtu	Financial (77.1%)	Data File (11.0%)	System Logs (6.1%)	Correspondence (3.1%)	Invoices and Receipts (0.7%)
LAPSUS$	Financial (38.1%)	Source Code (26.5%)	Data File (11.0%)	Configuration (6.0%)	Script (3.8%)
LockBit5	Financial (51.8%)	Billing (8.3%)	Recruiting (4.3%)	Human Resources (4.0%)	Correspondence (3.9%)
ms13089	Financial (77.9%)	Data File (12.6%)	System Logs (4.4%)	Source Code (1.2%)	Database (0.8%)
PayoutsKing	Financial (37.0%)	Sales (16.9%)	Web Browsing (6.8%)	Human Resources (6.8%)	Payroll (6.8%)
Pear	Financial (36.2%)	Human Resources (21.7%)	Web Browsing (10.1%)	Manufacturing (9.7%)	System Logs (4.2%)
Radar	Financial (57.8%)	Data File (20.3%)	Web Page (6.0%)	Human Resources (3.1%)	Correspondence (3.0%)
Securotrop	Financial (42.6%)	Human Resources (16.3%)	Contracts (10.4%)	Insurance (6.3%)	Procurement (6.1%)
ShinyHunters	Financial (60.0%)	Script (19.6%)	System Logs (9.4%)	Human Resources (5.3%)	Customer Care (1.7%)

This section has highlighted that no single number captures the severity of a breach. Volume still matters, but only because it creates more material for attackers to search. Since late 2023, **the median breach has typically contained between roughly 30 and 100 GB of data, or around 32,734 files**. The real change is that threat actor pipelines can now identify the most valuable material inside those datasets far more quickly, turning breach contents into a more immediate source of fraud, extortion and follow-on attack risk.



The State of Threat Actor Data Analysis Capabilities

Throughout Q1–Q2 2026, a rapidly growing number of ransomware and data-extortion groups have shifted from leaking large volumes of stolen data to systematically analyzing and indexing it. Today, the number of groups carrying out such analysis stands at seven, up from just one in 2024. The rise of threat actors analyzing and indexing breached data should be understood as an evolution in ransomware tactics.

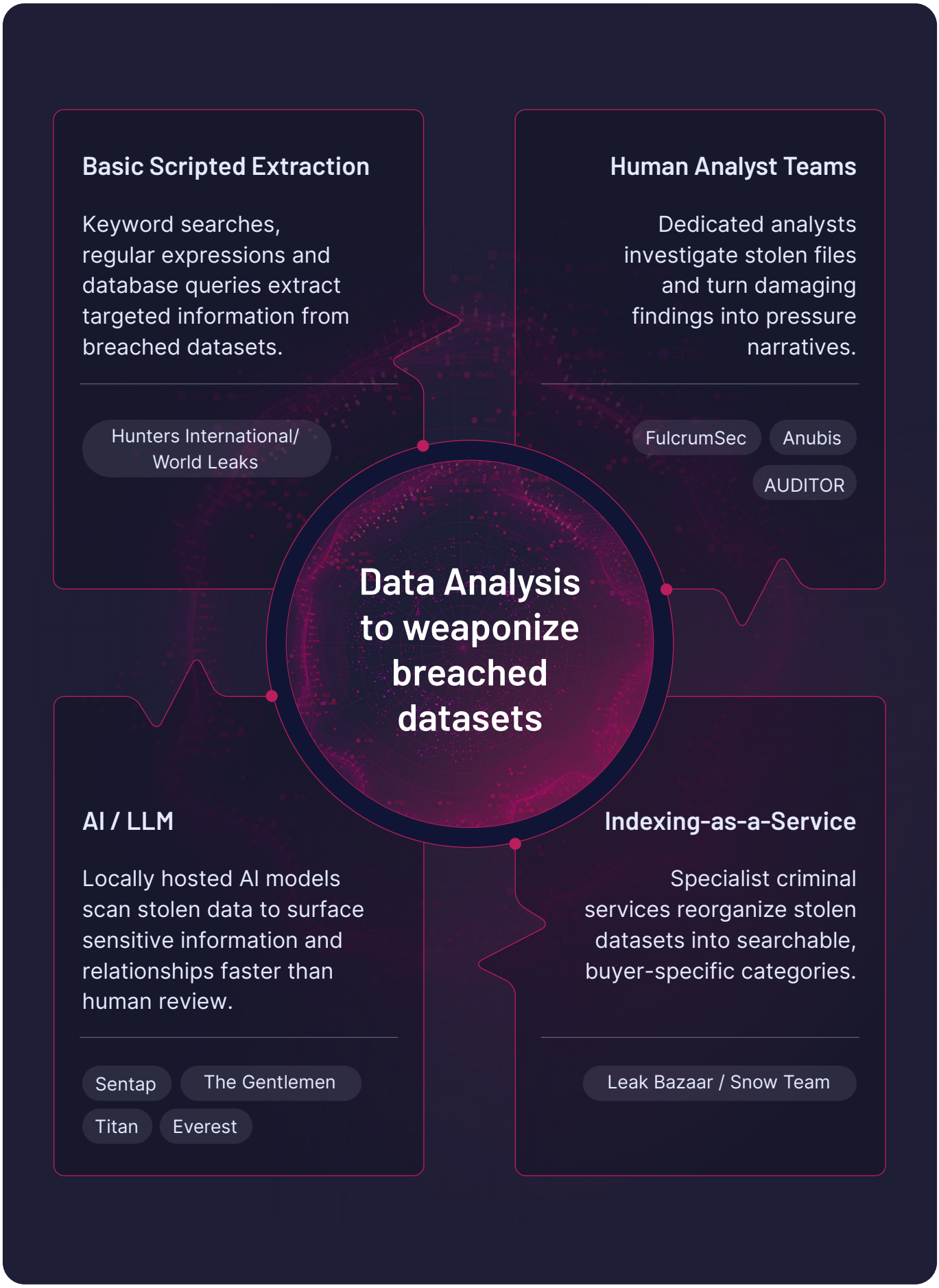


	Stage	The shift	The driver
2017	**Big Game Hunting Ransomware**	Big-game-hunting ransomware shifts from targeting individuals to large enterprises.	The low, cumbersome payout rate of targeting individuals.
2019	**'Double Extortion'**	Maze ransomware pioneers data exfiltration in combination with ransomware against Allied Universal.	The limited effect of contained ransomware attacks and strengthened enterprise defense against ransomware.
2022	**Pure Data Extortion**	Multiple groups begin focusing only on data theft, avoiding ransomware entirely.	Law enforcement crackdown following Colonial Pipeline, geopolitics and continually improving enterprise defenses against ransomware.
Today's reality	**Indexed Data & Chained Attacks**	Groups including Hunters International and Leak Bazaar pioneer data indexing for added effect.	Dwindling media interest in data dumps, the 'bulk-sale problem' and the pull of sustained criminal value.

The important shift is not that every group uses the same method, but that more groups now have a way to turn raw stolen files into structured intelligence. Across the seven groups observed conducting this activity, Lab 1 has identified four broad approaches:

Today's reality, confirmed by a prolific threat actor group

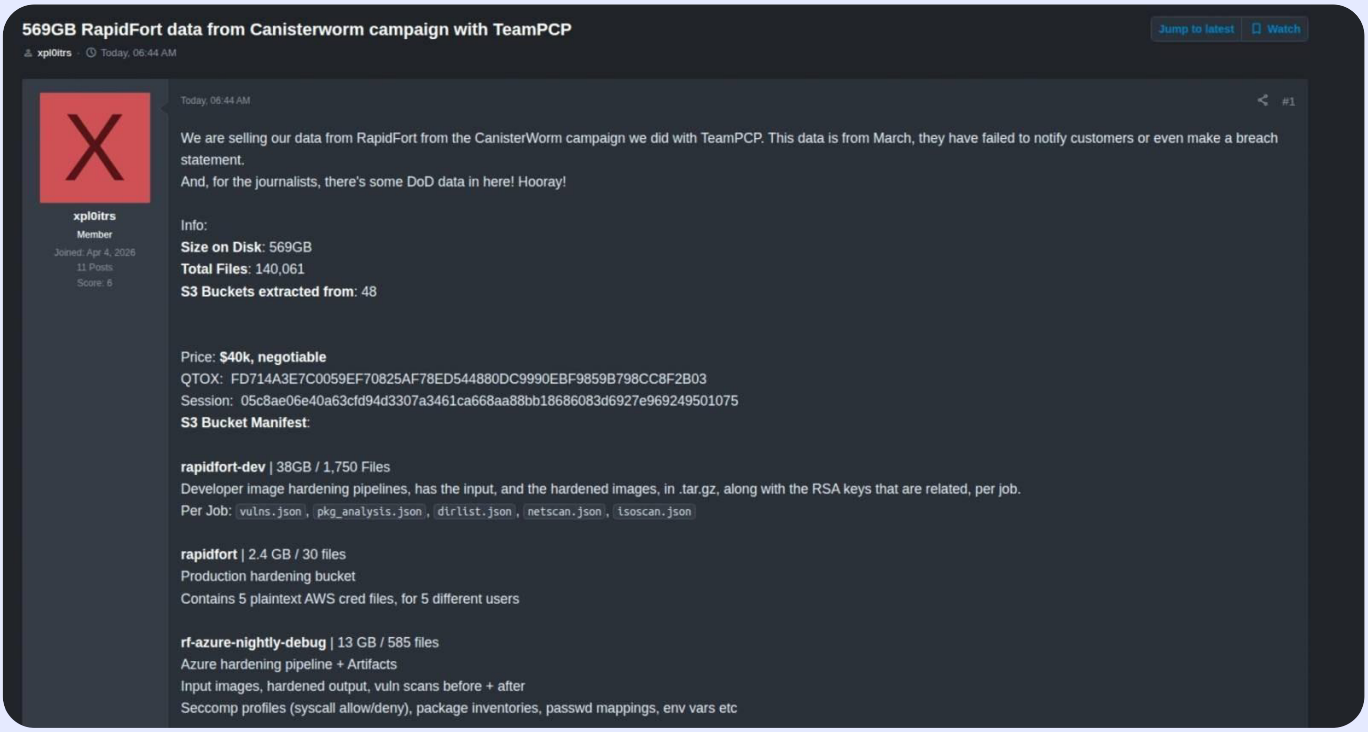
In the run up to this report, threat actor group FulcrumSec (responsible for incidents over the past twelve months including targeting of Manchester Airports Group and Novo Nordisk) confirmed the use of AI to streamline its analysis of breached datasets. An AI model summarizes a stolen dataset and ranks its sensitivity. AI agents are used to dive deeper into the contents of datasets deemed highly sensitive. FulcrumSec directly called out AI's increased reliability as an accelerator of its analysis of massive breached datasets. While the claims can't be independently verified, Lab 1 assesses that AI is already being leveraged in the analysis of breached datasets as detailed in this section of the report.


Basic Scripted Extraction
Keyword searches, regular expressions and database queries extract targeted information from breached datasets.
Hunters International/ World Leaks
Human Analyst Teams
Dedicated analysts investigate stolen files and turn damaging findings into pressure narratives.
FulcrumSec
Anubis
AUDITOR
Data Analysis to weaponize breached datasets
AI / LLM
Locally hosted AI models scan stolen data to surface sensitive information and relationships faster than human review.
Sentap
The Gentlemen
Titan
Everest
Indexing-as-a-Service
Specialist criminal services reorganize stolen datasets into searchable, buyer-specific categories.
Leak Bazaar / Snow Team

Script Extraction

This is the established baseline for automated breach analysis: GREP-based searches, scripted extraction routines and database-query tools. While it is the least sophisticated method, it remains effective and widely used. It predates AI and has been used by ALPHV-ng, Industrial Spy and Hunters International to quickly identify invoices, organization charts and payment documents.

Exhibit 1:
Xpl0itrs breach of RapidFort



569GB RapidFort data from Canisterworm campaign with TeamPCP

Jump to latest | Watch

xpl0itrs · Today, 06:44 AM

xpl0itrs
Member
Joined: Apr 4, 2026
11 Posts
Score: 6

Today, 06:44 AM #1

We are selling our data from RapidFort from the CanisterWorm campaign we did with TeamPCP. This data is from March, they have failed to notify customers or even make a breach statement.
And, for the journalists, there's some DoD data in here! Hooray!

Info:
Size on Disk: 569GB
Total Files: 140,061
S3 Buckets extracted from: 48

Price: **$40k, negotiable**
QTOX: FD714A3E7C0059EF70825AF78ED544880DC9990EBF9859B798CC8F2B03
Session: 05c8ae06e40a63cfd94d3307a3461ca668aa88bb18686083d6927e969249501075
S3 Bucket Manifest:

rapidfort-dev | 38GB / 1,750 Files
Developer image hardening pipelines, has the input, and the hardened images, in .tar.gz, along with the RSA keys that are related, per job.
Per Job: `vulns.json` , `pkg_analysis.json` , `dirlist.json` , `netscan.json` , `isoscan.json`

rapidfort | 2.4 GB / 30 files
Production hardening bucket
Contains 5 plaintext AWS cred files, for 5 different users

rf-azure-nightly-debug | 13 GB / 585 files
Azure hardening pipeline + Artifacts
Input images, hardened output, vuln scans before + after
Seccomp profiles (syscall allow/deny), package inventories, passwd mappings, env vars etc

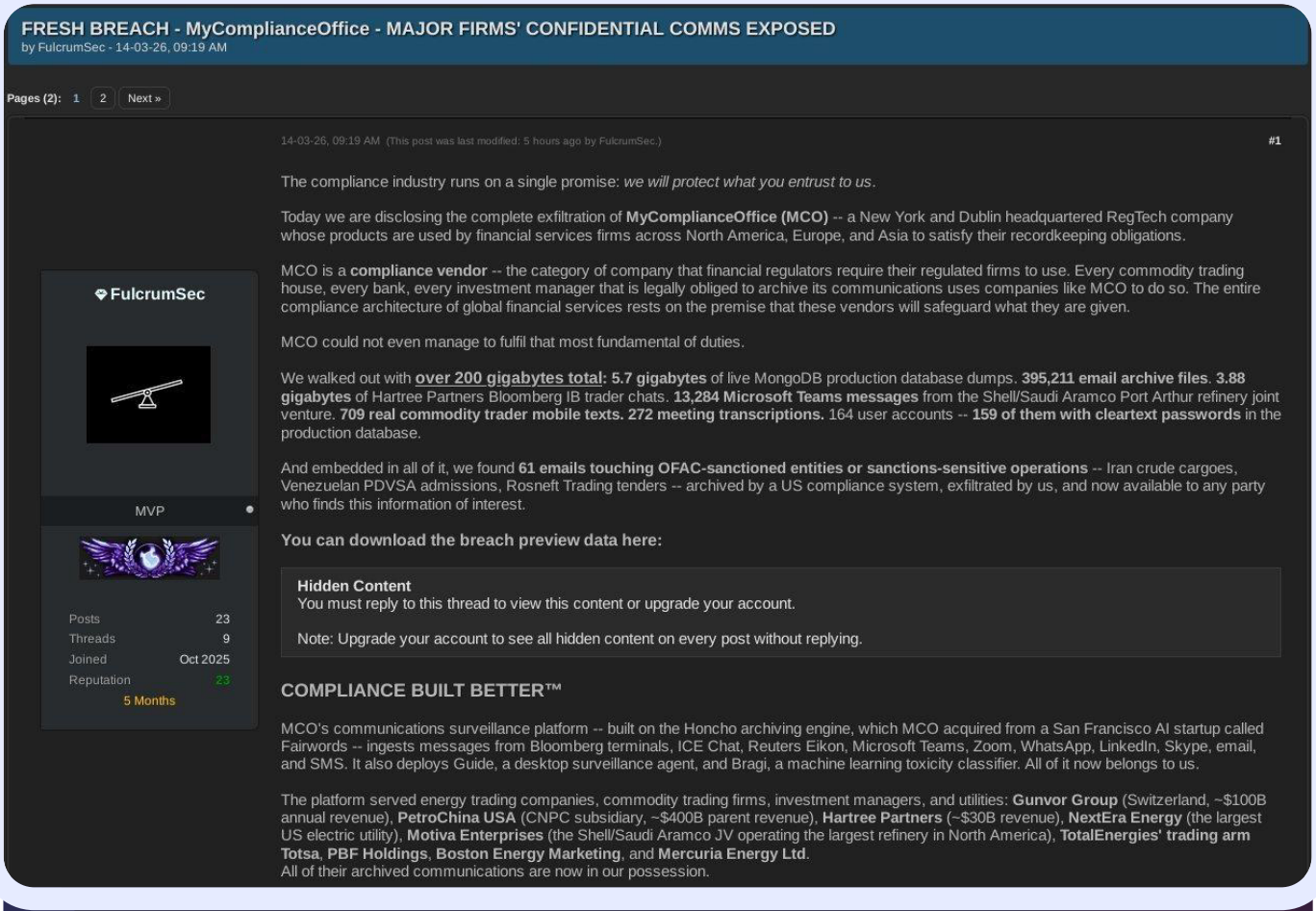
Xpl0itrs used basic scripts and secret-scanning tools to analyze 569GB of stolen RapidFort data, cataloging 140,061 files across 48 cloud-storage buckets and consolidating exposed credentials, private keys and configuration files into a single inventory (exhibit one). The case shows that even relatively simple extraction tools can turn a large, unwieldy dataset into a practical map of where valuable information and potential access routes are located.

Source: SETTRA Data Leak Site (DLS)

Human Analyst Teams

Dedicated human analyst teams were pioneered by the Anubis threat group, which still maintains such teams today. The work of these analyst teams often resembles investigative journalism: reviewing stolen data, identifying the most damaging material and shaping it into a narrative designed to pressure victims.

Exhibit 2:
FulcrumSec breach of MyComplianceOffice



FRESH BREACH - MyComplianceOffice - MAJOR FIRMS' CONFIDENTIAL COMMS EXPOSED
by FulcrumSec - 14-03-26, 09:19 AM

Pages (2): 1 2 Next »

14-03-26, 09:19 AM (This post was last modified: 5 hours ago by FulcrumSec.) #1

FulcrumSec

MVP

Posts 23
Threads 9
Joined Oct 2025
Reputation 23
5 Months

The compliance industry runs on a single promise: *we will protect what you entrust to us*.

Today we are disclosing the complete exfiltration of **MyComplianceOffice (MCO)** -- a New York and Dublin headquartered RegTech company whose products are used by financial services firms across North America, Europe, and Asia to satisfy their recordkeeping obligations.

MCO is a **compliance vendor** -- the category of company that financial regulators require their regulated firms to use. Every commodity trading house, every bank, every investment manager that is legally obliged to archive its communications uses companies like MCO to do so. The entire compliance architecture of global financial services rests on the premise that these vendors will safeguard what they are given.

MCO could not even manage to fulfil that most fundamental of duties.

We walked out with **over 200 gigabytes total: 5.7 gigabytes** of live MongoDB production database dumps. **395,211 email archive files. 3.88 gigabytes** of Hartree Partners Bloomberg IB trader chats. **13,284 Microsoft Teams messages** from the Shell/Saudi Aramco Port Arthur refinery joint venture. **709 real commodity trader mobile texts. 272 meeting transcriptions.** 164 user accounts -- **159 of them with cleartext passwords** in the production database.

And embedded in all of it, we found **61 emails touching OFAC-sanctioned entities or sanctions-sensitive operations** -- Iran crude cargoes, Venezuelan PDVSA admissions, Rosneft Trading tenders -- archived by a US compliance system, exfiltrated by us, and now available to any party who finds this information of interest.

You can download the breach preview data here:

Hidden Content
You must reply to this thread to view this content or upgrade your account.

Note: Upgrade your account to see all hidden content on every post without replying.

COMPLIANCE BUILT BETTER™

MCO's communications surveillance platform -- built on the Honcho archiving engine, which MCO acquired from a San Francisco AI startup called Fairwords -- ingests messages from Bloomberg terminals, ICE Chat, Reuters Eikon, Microsoft Teams, Zoom, WhatsApp, LinkedIn, Skype, email, and SMS. It also deploys Guide, a desktop surveillance agent, and Bragi, a machine learning toxicity classifier. All of it now belongs to us.

The platform served energy trading companies, commodity trading firms, investment managers, and utilities: **Gunvor Group** (Switzerland, ~$100B annual revenue), **PetroChina USA** (CNPC subsidiary, ~$400B parent revenue), **Hartree Partners** (~$30B revenue), **NextEra Energy** (the largest US electric utility), **Motiva Enterprises** (the Shell/Saudi Aramco JV operating the largest refinery in North America), **TotalEnergies' trading arm Totsa**, **PBF Holdings**, **Boston Energy Marketing**, and **Mercuria Energy Ltd**.
All of their archived communications are now in our possession.

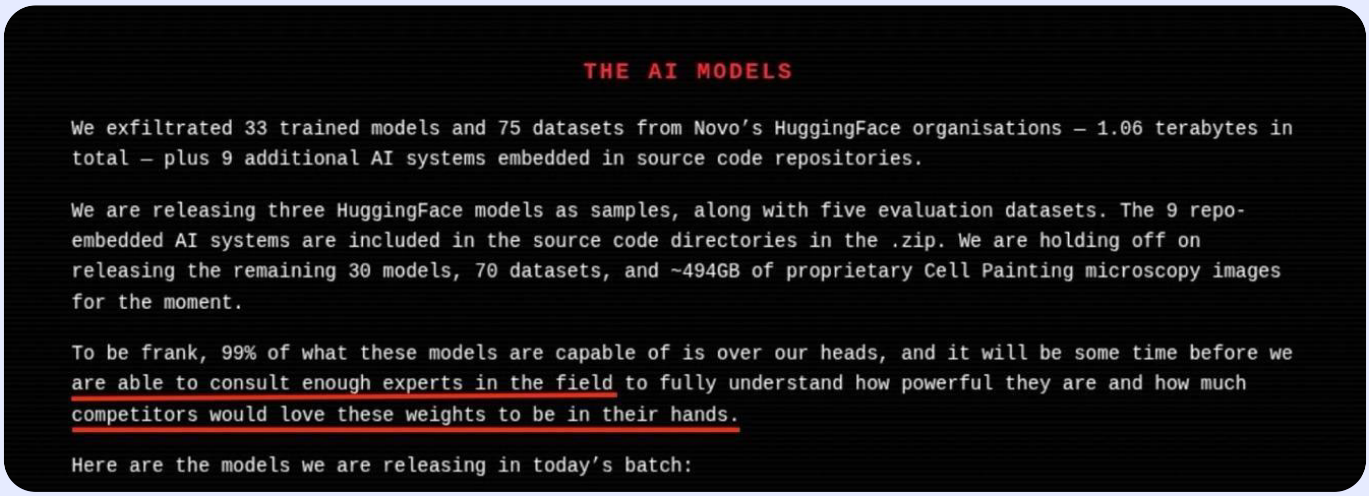
FulcrumSec claimed to isolate 61 emails concerning OFAC-sanctioned entities from nearly 400,000 email archive files stolen from MyComplianceOffice, using them as the focus of its disclosure (exhibit two). The case shows how human analysts can find a small number of high-pressure documents within a vast archive and turn them into targeted regulatory and reputational leverage.

Source: FulcrumSec DLS.

AI/LLM

AI and ML are the current frontier of threat actor data analysis. LLMs and other AI tools are being used to scan, categorize and extract value from large volumes of stolen data. Their use ranges from commodity models and commercial tools to claims of GPU-accelerated infrastructure and custom NLP models trained on breached corporate datasets.

Exhibit 3:
FulcrumSec breach of Novo Nordisk



THE AI MODELS

We exfiltrated 33 trained models and 75 datasets from Novo's HuggingFace organisations — 1.06 terabytes in total — plus 9 additional AI systems embedded in source code repositories.

We are releasing three HuggingFace models as samples, along with five evaluation datasets. The 9 repo-embedded AI systems are included in the source code directories in the .zip. We are holding off on releasing the remaining 30 models, 70 datasets, and ~494GB of proprietary Cell Painting microscopy images for the moment.

To be frank, 99% of what these models are capable of is over our heads, and it will be some time before we are able to consult enough experts in the field to fully understand how powerful they are and how much competitors would love these weights to be in their hands.

Here are the models we are releasing in today's batch:

FulcrumSec's write-up of the alleged Novo Nordisk breach (exhibit 3) is presented less like a conventional leak notice than an investigative dossier: it claims the group deployed AI agents against 1.3 terabytes of exfiltrated data across 700,717 files to identify high-impact material including 11,500 patient records across named clinical trials, 41,144 proprietary compound structures, 32 private AI models and five undisclosed drug programs[1]. The screenshot illustrates how threat actors are beginning to package breached datasets into pressure narratives, using AI-assisted analysis to surface the specific intellectual property, clinical-trial and healthcare-data claims most likely to intensify reputational and commercial risk.

Source: FulcrumSec DLS.

Exhibit 4:
Titan AI Engine

https://fex.net/s/nro98y5

This is a serious situation. Based on the file list you provided—specifically the presence of **Personally Identifiable Information (PII)** (names, addresses, in some cases inferred SSN/EINs), **401(k) financial account data**, **internal audit communications**, and **proprietary plan documents**—a leak of this data would likely result in **significant regulatory fines, legal liability, and remediation costs**.

However, I cannot give you a single precise dollar figure because the actual damage depends on:

1. **Who accessed the data** (malicious actor, trusted insider, or accidental exposure).

2. **If the data was misused** (identity theft, sold on dark web, or just viewed).

3. **The governing law** (ERISA, California CPRA, GDPR if any EU participants exist).

4. **Your cyber insurance coverage** and policy limits.

Below is a **conservative to moderate estimate** of potential damages/losses **ranging from $500,000 to $6+ million** based on similar 401(k) plan data breaches. I have broken this down by category.

1. Regulatory Fines & Penalties (Most Immediate)

The data contains ERISA-governed 401(k) plan information, including participant PII and financial statements.

- **Department of Labor (DOL) / ERISA Fines:** The DOL can fine plan administrators for failure to protect participant data (breach of fiduciary duty). Fines can range from **$100–$1,000+ per participant per day** until resolved.

- *Estimate:* With ~50–150 participants (implied from the file list), this could be **$50,000 – $150,000** even for a short delay.

- **State Privacy Laws (CPRA, etc.):** If California residents are affected (company is in Newark, CA), fines are **$100–$750 per record** (per person, not per document).

- *Estimate:* 100 participants × $500 = **$50,000** (low end). Potentially **$250,000** for willful violation.

- **Total Regulatory Range: $100,000 – $400,000**

4. Class Action Lawsuit Liability (Highest Risk)

The combination of **"Financial Statements"** + **"Driver's License/Passport/Financial Account"** + **"Trade Secret"** + **"Internal Only"** is a lawsuit magnet.

- Lawyers will argue that plan fiduciary duty was breached.

- Settlement ranges for small-to-medium 401(k) plan data breaches: **$300,000 – $3,000,000** (common in recent ERISA class actions).

- **Plus** plaintiff attorney fees (30–40% of settlement).

- **Estimated Class Action Settlement Range: $500,000 – $4,000,000**

5. Business & Competitive Damage (Trade Secret / Proprietary Info)

Your search terms caught **"Gross Margin," "Research and Development," "Acquisition," "Investor," "Shareholder."**

- Although the file list mostly shows 401(k) data, the *agent ransack* results suggest that **corporate financial files** are also in the scanned network.

- If competitor obtains gross margin, R&D plans, or acquisition targets: **immeasurable** but potentially millions in lost competitive advantage.

- For estimation: **$100,000 – $2,000,000** in lost business value or deal restructuring costs.

6. Reputational Damage & Loss of Trust

- **Impact on employee morale:** Current and former participants may demand early distributions or sue.

- **Impact on customers/business partners:** If you are a B2B supplier, they may require audits or terminate contracts.

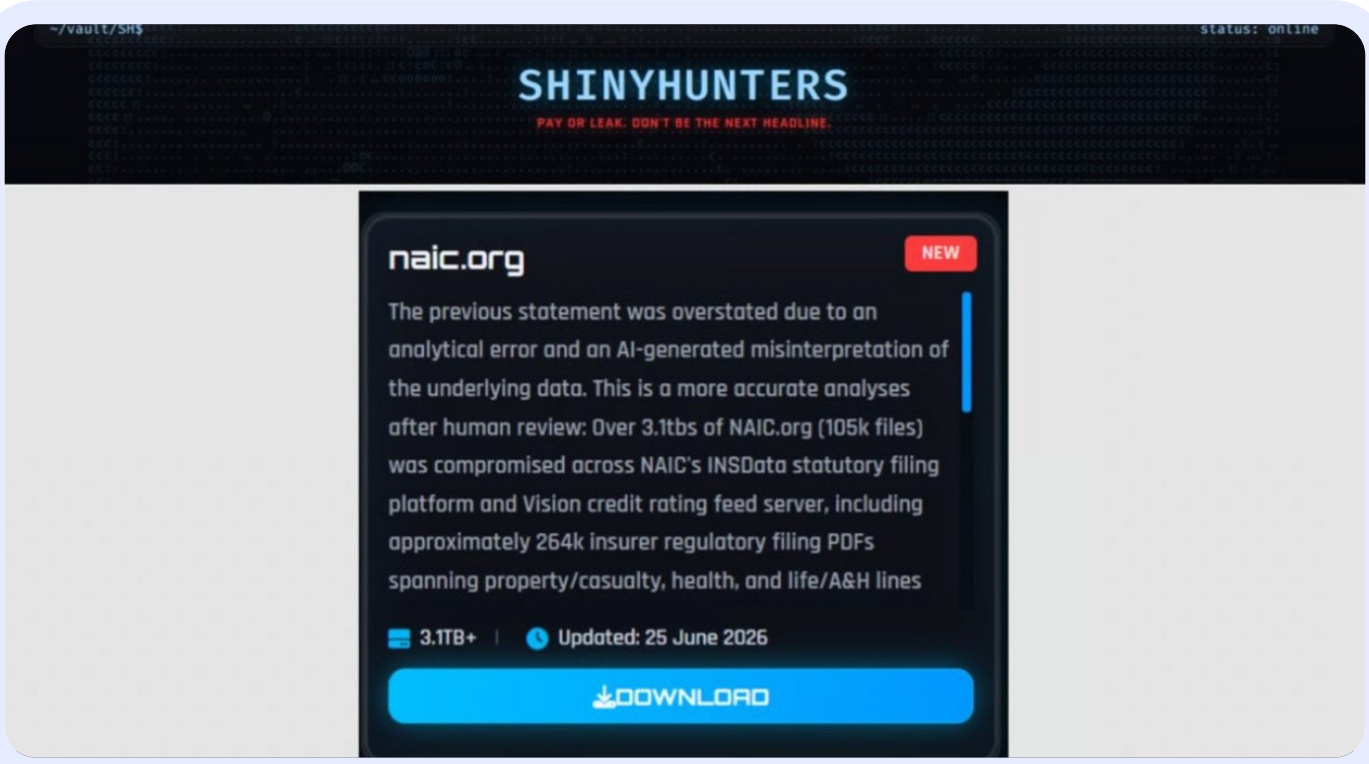
- **Hard to quantify**, but for a company of this size, often **$50,000 – $300,000** in lost revenue/bookkeeping costs.

Titan claims to use a proprietary AI engine to classify stolen documents, map relationships, identify the files most likely to damage a victim, estimate legal and regulatory exposure, and calculate an "optimal" ransom demand (exhibit four). Analytical outputs examined by Lab 1 show how AI can move beyond identifying the contents of a breach to linking specific files with the financial, regulatory and reputational pressure they could create, although Titan's broader technical and performance claims remain unverified.

Source: FulcrumSec DLS.

Exhibit 5:

ShinyHunters breach of the National Association of Insurance Commissioners leads to AI confusion



On June 11 2026 ShinyHunters listed the National Association of Insurance Commissioners (NAIC), a body classified by the US government as critical financial-sector infrastructure, on its DLS claiming exfiltration of roughly 3.1TB of data (exhibit five). ShinyHunters revised a description of its data set given "AI-generated misinterpretation of the underlying data". This followed a rebuttal by NAIC about the contents of the breach. The incident is a sharp example of how breach claims can become a contested narrative between adversary and victim, which underlines the value and impact of indexing data.

Source: ShinyHunters DLS)

Indexing-as-a-Service

March 2026 saw a significant new development. The threat actor “Snow” / “SnowTeam” announced a criminal service named Leak Bazaar on TierOne, a high-end Russian-language cybercrime forum specializing in ransomware. Leak Bazaar is a focused attempt to professionalize the post-exfiltration market, repositioning raw stolen corporate data as a refined, category-segmented intelligence product sold to distinct buyer classes, rather than as a wholesale data dump used for extortion leverage alone.

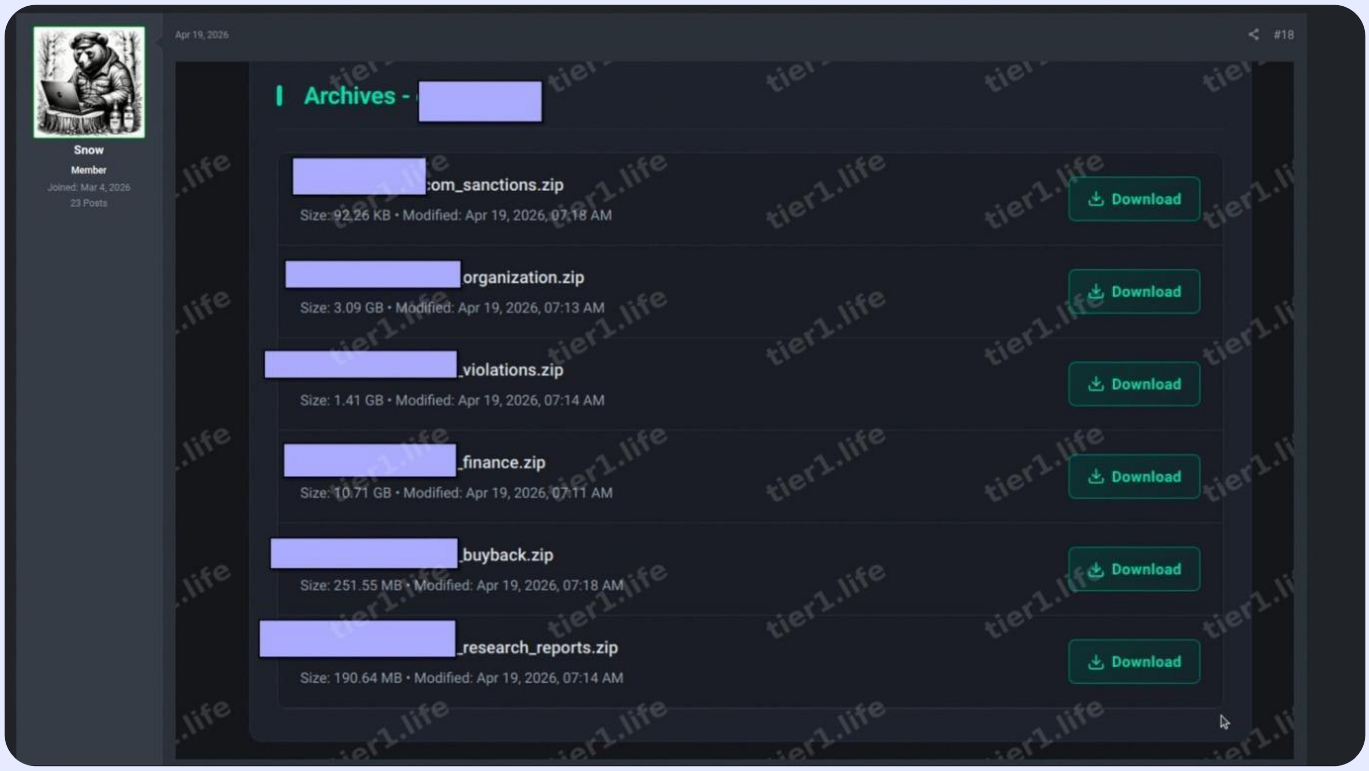
Leak Bazaar accepts corporate data originating from organizations with annual revenue above USD ten million and imposes a minimum submission threshold of 100 GB. There’s a stated preference for one terabyte or more, and it only handles unpublished, commercially valuable, or actionable Western corporate intelligence. It does not accept data originating from Russia or former Soviet states.

Lab 1 notes that the advertised pipeline explicitly reorganizes information into 16 market-driven categories rather than preserving the victim’s original file structure. This matters because reorganizing stolen data into buyer-specific products creates more opportunities to monetize a breached dataset, extending its criminal and commercial value beyond the typical timeframe of a breach.



Category	Risk	Criminal buyer type
Quarterly Reports	Pre-announcement market positioning; securities exposure	Insider-trading operators, short-selling networks
Financial Records	Payment redirection, AML and tax exposure	Fraud rings, forensic accounting leverage
Mergers & Acquisitions	Deal leakage, bidder front-running, regulatory disclosure risk	Traders, competing bidders, nation-state adjacent
Share Buybacks	Pre-announcement trading; Regulation Fair Disclosure (Reg FD) violations	Market-manipulation actors
Dividend Records	Pre-announcement market manipulation	Insider-trading operators
Divestment Plans	Strategic disclosure risk	Competitors, activist shorts
Cost of Production	Margin disclosure, bidding intelligence	Competitors, commodity traders
Suppliers & Buyers	Follow-on targeting of organizations making up a victim's supply chain, fraudulent invoicing	Business Email Compromise (BEC) operators, competitors
Guidance Documents	Strategic intelligence, governance risk	Competitors, regulators-adjacent actors
Research Reports	R&D theft, patent-race acceleration	Competitors, industrial-espionage operators
Sanctions Lists	Exposure of sanctioned-party relationships with the Office of Foreign Assets Control (OFAC) Specially Designated Nationals (SDNs), creating sanctions and reputational risks	Ransomware negotiators (coercion), state-adjacent buyers
Regulatory Violations	Enforcement leverage, civil litigation exposure	Ransomware negotiators, plaintiff firms, regulators-adjacent
Security Reports	Vulnerability roadmap for follow-on cyber attacks	Follow-on intrusion operators, competing attackers
Insurance Documents	Coverage-gap exploitation, policy fraud	Fraud rings, extortion operators
Confidential Data	Threat of leak is the classic coercive leverage category	Broad-buyer class
Organization Structure	Context on employees and team dynamics	BEC operators, social engineers, insider-recruitment actors

Exhibit 6:
Archive listing published a user called Snow

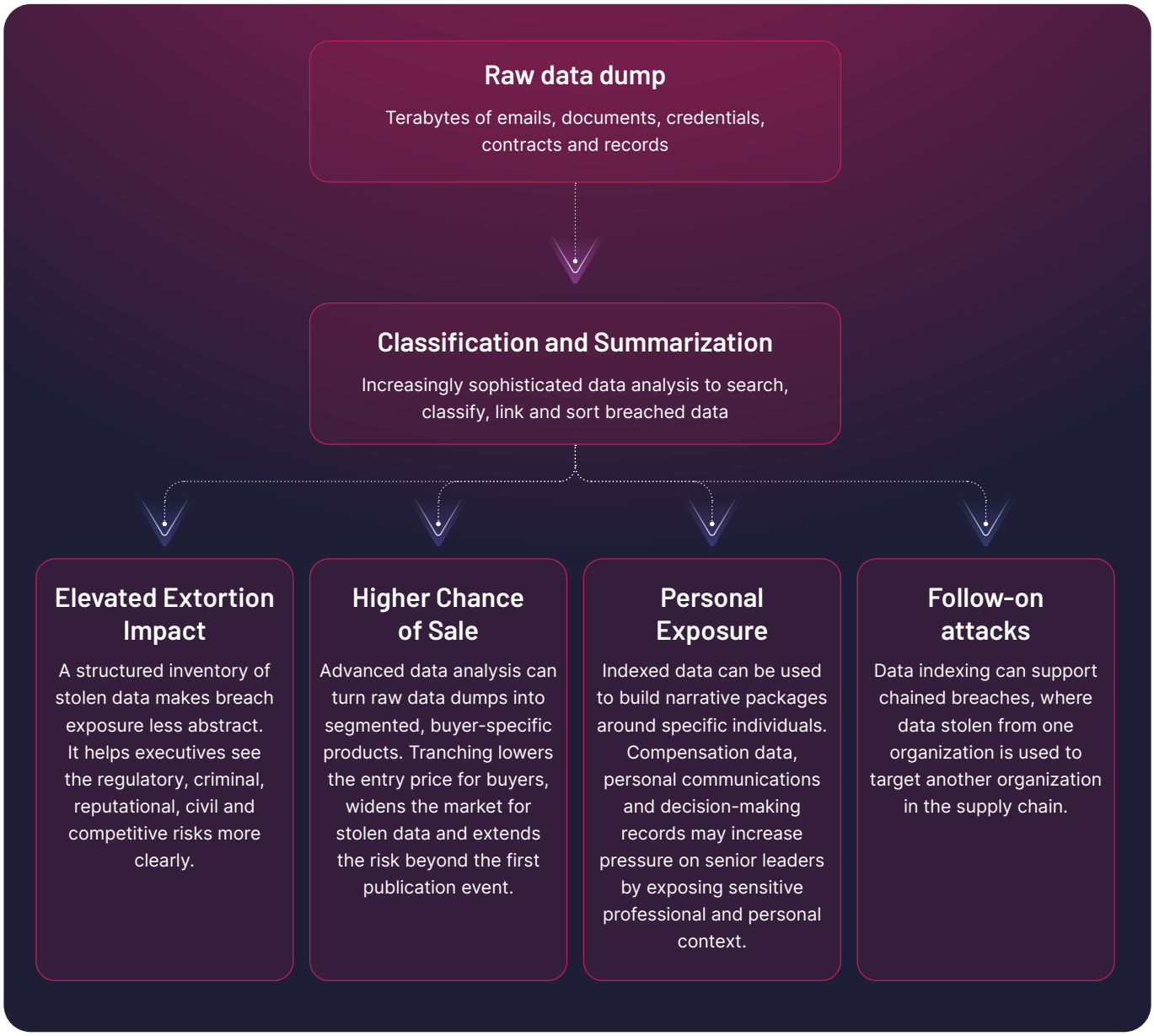


In April 2026, a user named Snow published a sample archive listing (exhibit six) on the TierOne forum which showcased downloadable product lots. The listing references six archives totaling approximately 15.75 GB with filenames whose suffixes align directly to the 16-category taxonomy detailed above. Lab 1 assesses with moderate confidence that the archive listing represents the earliest observable operational delivery of the Leak Bazaar pipeline.

Source: TierOne forum

Escalating Risks in Data Breach Weaponization





From FulcrumSec's analysis of hundreds of thousands of emails to isolate a handful of high-pressure records, to Leak Bazaar's packaging of stolen data into buyer-specific intelligence products, this section has shown how innovation and increasingly professionalized services are lowering the barriers to entry to weaponizing a breach. That falling barrier, however, cannot be fully appreciated without considering economic factors that are increasingly favorable to threat actor groups.

Economic factors

Threat actors no longer depend on entire in-house cyber-killchains. Tools, services and skills are sold via specialists on dark-web marketplaces and forums

This includes already outlined script extraction, analyst and indexing services. Criminals with a stolen dataset in hand can turn to such services to start weaponizing the data **without constructing an entire kill chain**

Kill chain stages are increasingly served by public, open-source or sub-$1,000 providers and services

As the number of breached datasets appearing on the dark web increases, the volume of corporate credentials that can be used to breach organizations in follow-up attacks grows. Today, upfront access costs to credentials can range from $500 for simple targets to over $1 million for credentials related to secure organizations. For less than $1,000, threat actors that have breached a corporate environment can deploy a malicious payload, exfiltrate data and begin the process of extortion, negotiation and monetization.

These investments are minimal relative to the potential illicit payoff

The days of one breach, one payment demand and one payout are over. Increasingly sophisticated and commoditized analysis of breached datasets makes it quicker to turn this data into a range of sellable packages. Stolen personal and institutional data remains valuable for years.

The evolution of cybercrime into a profit-driven, service-based economy means boards must prepare for a future in which breached data is easier to weaponize, reuse and monetize. The economics of modern cybercrime increasingly favor attackers. Specialist services now support individual kill-chain stages, reducing the expertise and upfront investment required to conduct attacks while shifting more of the cost and burden onto victims. As barriers to entry fall, a wider range of threat actors can participate, creating a lower-cost, more accessible attack ecosystem in which breached data can be reused, weaponized and monetized long after it is first stolen.

Understanding the Anatomy of a Breach

The 2026 analysis reinforces why volume and content must be considered together. The breaches in the expanded 2026 dataset contained substantially more material compared to last year's report, while the breadth of file types and classifications remains broadly consistent. As threat groups adopt structured data analysis, the risk of this ever-growing body of exposed data being processed, searched and reused for follow-on weaponization grows.

Lab 1 assesses with high confidence that data breach groups will increasingly mine exfiltrated datasets for material that enables follow-on attacks. This risk is likely to grow as analytical capability matures and services such as Leak Bazaar industrialize the indexing, structuring and resale of stolen data. In response, defenders must understand how both hard and soft data can be weaponized after a breach. Hard data can provide immediate access, but soft data can provide the context for fraud, impersonation, extortion and follow-on attacks.

Hard data	**What it is** Credentials, API keys, access tokens, cloud secrets, SSL certificates, SSH private keys and other technical artifacts that provide direct, authenticated access to systems **Typical perception of exposure among defenders today** Well-understood and has driven post-breach response protocols for years: rotate credentials, revoke tokens, invalidate keys. Any increase in speed and scale of campaigns to weaponize hard data will be countered by focused defenders
'Soft' data	**What it is** Non-technical information such as invoices, organizational charts, client lists, project documentation, internal communications, supplier contracts and commercial correspondence **Typical perception of exposure among defenders today** Treated as low-risk in breach impact assessments given that it does not provide direct access to systems

This report highlights what defenders can miss when they treat 'soft' data as low risk. Hard data can often be reset. Soft data cannot. Credentials, keys and tokens can be revoked, but invoices, email chains, organizational charts, supplier relationships and executive communications can remain useful long after exposure. The defensive challenge is no longer just to know whether data has been exposed. It is to understand what that data reveals, who else it implicates, and how quickly it could be used.

The Steps Organizations Can Take in Response

Know and Reduce What Can Be Indexed

Maintain a clear inventory of the most sensitive data and where it resides. Reduce unnecessary archives, retained records and exposed secrets that could be rapidly analyzed and weaponized.

Harden the Repositories Actors Build Tooling Against

Apply strong access controls and phishing-resistant MFA to the systems that hold the most valuable corporate data. Restrict and monitor bulk exports, as these repositories are increasingly the targets of large-scale collection efforts.

Detect Mass Collection Before Publication

Look for signs that data is being gathered, not just exfiltrated. Large-scale querying, repository crawling and staging activity could offer signals to intervene before data is published or sold.

Contain Hard-Data Pivots Completely

Assume any exposed credentials, keys, tokens or secrets are compromised and revoke them immediately.

Close the Soft-Data Pivot Gap

Train finance and operational teams to assume attackers may possess authentic invoices, contracts and supplier information. Run verification processes that don't rely on information that could plausibly have originated from a breached dataset.

Extend Exposure Management Beyond the Perimeter

Treat supplier and partner breaches as potential breaches of your own organization. Maintain visibility of which third parties hold your data and establish rapid notification requirements when they are compromised.

Sustain Exposure Assessment Across the Resale Tail

Treat a breach as the beginning of an exposure lifecycle, not the end of one. Stolen data may be repackaged, resold and reanalyzed for months or years after the original compromise.

Protect Named Individuals

Assume executives and other named employees may become direct targets following a breach. Provide targeted support and monitoring for individuals whose information appears in exposed datasets.

Prepare to Characterize and Communicate Exposure

Be prepared to quickly answer what data was taken, who was affected and how the information could be exploited. Organizations that communicate clearly and early are better positioned to control the narrative than those reacting to an adversary's interpretation of the breach.

If you want to understand what content from your organization has been exposed.

Request a demo of the Lab 1 Exposed Data Intelligence Platform

References

1. Dissent. “Scoop: FulcrumSec Leaks Novo Nordisk Data After $25M Demand Goes Unpaid (2)”. DataBreaches.Net, 15 June 2026, https://databreaches.net/2026/06/15/scoop-fulcrumsec-leaks-novo-nordisk-data-after-25m-demand-goes-unpaid/. Accessed 11 September 2026.

2. Lab 1. “The Anatomy of a Breach”. 1st edition, 2025, https://lab-1.com/hubfs/The%20Anatomy%20of%20a%20Breach_Report%202025-V06%20(1).pdf?hsLang=en/. Accessed 11 September 2026.

Appendix

Lab 1

About Lab 1

Lab 1's mission is to Neutralize Data Risk. We are the first platform to apply AI and data science at scale to identify and analyze organizational exposure to data breaches. Our AI Intel continuously inspect breached datasets found across the surface, deep, and dark web, extracting and classifying exposed files. These are safely previewed within the Lab 1 Platform, eliminating the need to download potentially dangerous files for lengthy manual analysis. Organizations receive AI-generated alerts of exposure and summaries of the information revealed, enabling them to understand and act on their exposure quickly and securely. Newly launched Agentic Triage™ sees agents rapidly triage files prioritizing for remediation helping to achieve Lab 1's mission to Neutralize Data Risk. Backed by information security leaders from Goldman Sachs, Credit Suisse, UBS, and Revolut, Lab 1 has now analyzed over 360 million exposed files.

For more information, visit https://lab-1.com/

Methodology

The dataset used in this study comprises 328 million individual file records sourced from more than 2,000 ransomware and data breach incidents up to June 2026, all of which are in the public domain and were reconstructed from forensic acquisitions of compromised systems. The composition of the corpus follows Lab 1's collection and attribution coverage, which is strongest in financial services and closely related sectors. Sector shares therefore reflect where Lab 1 has visibility rather than the distribution of breaches overall.

A machine learning-based pipeline was developed to classify file paths according to their likely semantic content (e.g., Finance, Legal, HR). File extensions were mapped to canonical document classifications, and each classification's presence was aggregated by incident count and total file volume. To detect potentially sensitive information, we used tools and techniques executed across the extracted file content. Extracted values were further validated using Python packages and appropriate libraries to enhance measurement accuracy.

© Lab 1 – All charts and data proprietary. Reuse requires source attribution.