Constella

2026 Identity Breach Report

The Industrialization of Identity Risk & The AI Agent Frontier

2026 | Annual Report

Constella

Table of Contents

3. **Foreword**
4. **About Constella's Identity Breach Report**
6. **Executive Summary**
7. **Key Data Insights**
9. **Key Threat Dimensions: Mapping the 2026 Identity Risk Landscape**
    - Identity at Machine Scale: The Industrialized Attack
    - The Infostealer Surge: 51.7 Million Packages
    - The Plaintext Crisis: Identity Risk as a Boardroom
    - Eternal Exposure: The Persistent Recycling of Identity
    - Targeted Exploitation: Executive Digital Exposure
    - The Identity Domino Effect: Cascading Supply Chain
    - Machine-Scale Warfare: The Rise of Autonomous Identity Exploitation
31. **2025 Data**
    - Total Breached Identity Metrics
    - Top 10 Breaches & Leakages
    - Top PII Exposed from Breaches
    - Top Social Media Attributes
    - Password Security & Algorithms
    - Geographic Distribution
    - Most Impacted Sectors
55. **Appendix & Data Sources**

The Identity Frontier: Navigating an Automated Threat Landscape

Identities have become the ultimate gateway to organizational and personal assets. Over the past year, we have witnessed the growth and expansion of the identity attack surface accelerate at a pace previously unseen, driven by relentless innovation in adversarial tactics and the rapid industrialization of digital risk. In the **2026 Identity Breach Report**, Constella provides a critical deep dive into this evolution, fueled by findings from the complete January through December 2025 data set.

The findings in this year's report are both a validation of our technological progress and a stark warning. In 2025, our team utilized **AI Agent automation** to expand our detection capabilities by **159%**, hunting over **567,000 breaches**. However, this same level of automation is being mirrored by our adversaries. The "industrialization" of cybercrime is no longer a concept; it is a reality manifested in the **51.7 million infostealer packages** we processed in 2025, a 72% increase that targets the very heart of the modern workforce.

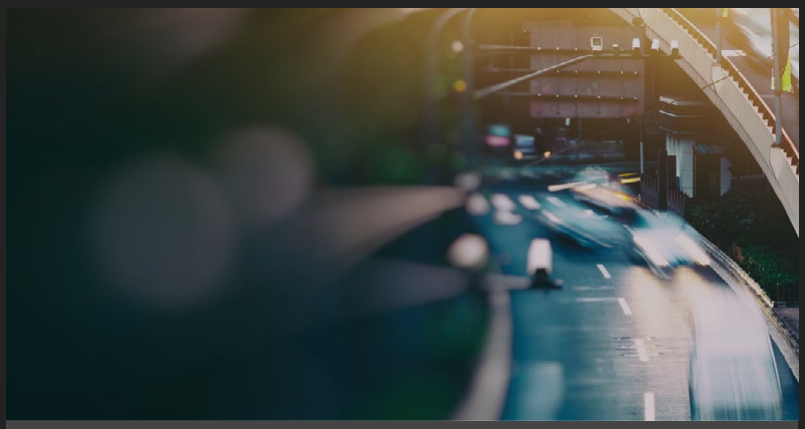
Perhaps the most alarming insight from our 2025 research is that identity breach exposures are virtually permanent (meaning the identity artifacts remain exploitable). A lot of this data is tied to a "never expires" exposure model. Nearly 60% of the breach datasets we ingested were recycled credential compilations (up from 43% the year prior), and only 7.4% represented truly unique breach or leak events. In other words, breaches effectively never end: even after password resets, the same emails, names, and partial PII continue circulating and remain weaponizable, as evidenced by the 4.2B unique email addresses observed across our monitoring. This is why identity centric defense must shift to continuous exposure monitoring and intelligence-led controls that assume yesterday's data will be exploited again tomorrow.

At Constella, our mission is to provide the visibility required to defend this new frontier. Our Data Lake now holds over **54.6 billion curated records** and **429 billion curated attributes,** providing our partners with the most comprehensive view of identity risk available. By consolidating daily "Combo" breaches and focusing on net-new unique records, we are delivering the high-fidelity intelligence necessary to stay ahead of automated threats.

As we move through 2026, the challenge for security leaders is clear: we must meet machine-scale threats with machine-scale intelligence. This report is designed to provide the data-driven foundation you need to strengthen your identity posture and protect your most valuable assets.

Andres Andreu
CEO, Constella Intelligence

About Constella's Identity Breach Report



The Definitive Map of the Identity Attack Surface

The **2026 Identity Breach Report** is Constella's comprehensive examination of the global identity-based risk landscape. This report is fueled by data collected from January 2025 through December 2025, providing security leaders, researchers, and organizations with the visibility required to defend against the industrialization of digital risk.

At the core of this report is the **Constella Identity Breach Data Lake,** the world's most extensive repository of breached and exposed personal information. Our intelligence is gathered from a vast array of sources across the surface, deep, and dark web, ensuring a 360-degree view of how identities are harvested, traded, and exploited.

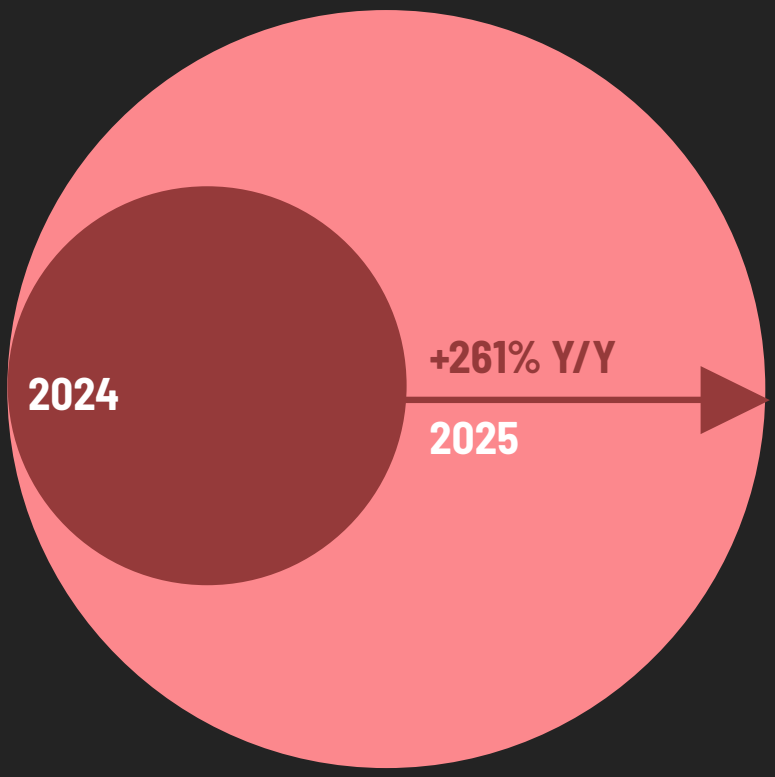
Unprecedented Scale and AI-Driven Precision In 2025, the scale of identity exposure reached a new peak. The Constella Data Lake now holds over **54.6 billion records** (+80% YoY) and **429 billion curated attributes** (+49% YoY). This massive expansion was made possible by our deployment of **Agentic AI automation**, which allowed our teams to hunt **159% more breaches** and curate **135% more records** than the previous year.

Redefining Data Quality

This report marks a shift in our data processing philosophy. We have moved toward high-fidelity curation:

- **High-Density Consolidation:** Total record volume exploded by 135%, reflecting a strategic technical shift. Adversaries are moving away from fragmented datasets in favor of synthesizing thousands of individual leaks into unified, high-fidelity **delta compilations** designed for machine-scale exploitation.
- **PII Enrichment:** A **+661% increase** in breaches containing PII demonstrates our focus on capturing data sets rich in Personally Identifiable Information (PII).
- **Infostealer Intelligence:** We processed **51.7 million infostealer packages** (+72% YoY), identifying **24.8 million unique infected devices.**

Credentials Stored in Plaintext



A Critical Security Alert:

The Plaintext Crisis

The 2026 report highlights a critical escalation in identity risk: **68.89% of all breached credentials were found in plaintext,** marking a staggering **261% increase year-over-year**.

Rather than a broad regression in organizational hygiene, this surge reflects the industrialization of adversarial tradecraft. Threat actors have mastered the recovery of plaintext passwords through high-velocity cracking farms and infostealer exfiltration, which strips away security layers at the point of infection. With only **5.26% of credentials** remaining properly hashed, the risk of immediate, automated account takeover (ATO) has reached an unprecedented peak.

EXECUTIVE SUMMARY

The Industrialization of Identity Risk

In 2025, the identity attack surface reached a critical tipping point. Driven by **Agentic AI** and automated harvesting, cybercrime has transitioned from manual exploitation to machine-scale industrialization. Constella's threat intelligence, powered by AI automation, identified **567,061 hunted breaches**, representing a **159% increase**, resulting in the curation of 27.9 billion high-fidelity records.

The Plaintext Crisis & Infostealer Surge

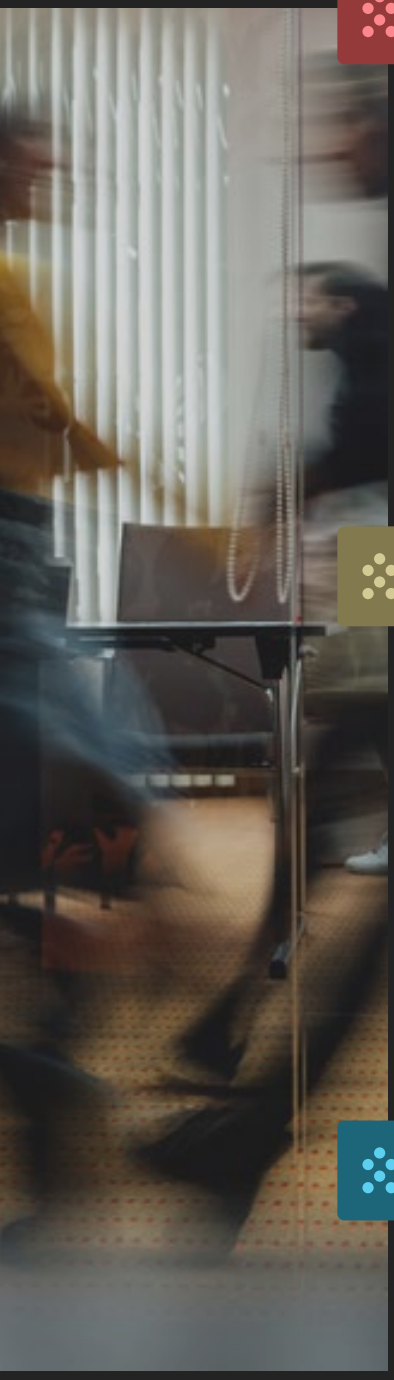
Our findings reveal a critical escalation in identity risk: **68.89% of all breached credentials were found in plaintext**, representing a staggering **261% increase year-over-year.** Rather than a regression in hygiene, this reflects the industrialization of adversarial tradecraft; threat actors have mastered the recovery of clear-text passwords through high-velocity cracking farms and infostealer exfiltration, allowing for instantaneous, automated Account Takeovers (ATO).

Simultaneously, infostealers have become the primary engine of identity theft. Constella processed **51.7 million packages (+72% YoY)**, identifying **24.8 million unique infected devices**. These logs are particularly lethal because they often contain session cookies, enabling adversaries to perform session hijacking and bypass Multi-Factor Authentication (MFA) entirely.

High-Fidelity Intelligence at Scale

To combat data recycling, Constella prioritized net-new, PII-rich data, resulting in a **661% increase in breaches containing PII data**. This refined approach has grown our Data Lake to **54.6B+ curated records** and **429B+ curated attributes**. As we navigate 2026, organizations must meet this speed of automation with machine-scale intelligence. This report provides the roadmap to defend the modern digital identity against an industrialized adversary.

Key Data Insights



1 | THE PLAINTEXT CRISIS: A CRITICAL SECURITY EVOLUTION

The most critical discovery in the 2025 data is the staggering prevalence of exposed credentials: **68.89% of all identified passwords were found in plaintext,** representing a **261% increase year-over-year.**

Rather than a broad regression in organizational security practices, this surge reflects a fundamental shift in adversarial tradecraft. Threat actors are increasingly evolving their methods to recover plaintext passwords from compromised datasets and operationalizing previously leaked credentials at scale. This industrialized pipeline, fueled by high-velocity cracking farms and infostealer exfiltration, ensures that even credentials once protected by legacy hashing are being converted into actionable, clear-text weapon libraries.

2 | AI-DRIVEN DETECTION: SCALING THREAT INTELLIGENCE

The exponential growth in our detection metrics is a direct result of the deployment of **Agentic AI automation** within Constella's threat-hunting pipeline. In 2025, our automated agents enabled a **159% increase in hunted breaches** and a **135% increase in curated records,** allowing our team to monitor the deep and dark web at a speed and depth that human analysts alone could not achieve.

This transition to machine-scale intelligence ensures that we are not only identifying more threats but also extracting higher-quality data. By automating the identification, normalization, and attribution of breach events, we have significantly shortened the "time-to-intelligence," providing our partners with a decisive advantage in the race against industrialized cybercrime.

3 | DATA FIDELITY: PII ENRICHMENT

While recycled lists are declining in our Data Lake, **breaches containing PII surged by 661%,** reflecting our expanded success in identifying and collecting PII-rich data sets directly from the source. This prioritization of high-value, verified leakages ensures that the intelligence we provide is actionable, current, and deeply relevant to the protection of individual and corporate identities.

Key Data Insights (Continued)

4 | MASSIVE GROWTH OF THE IDENTITY DATA LAKE

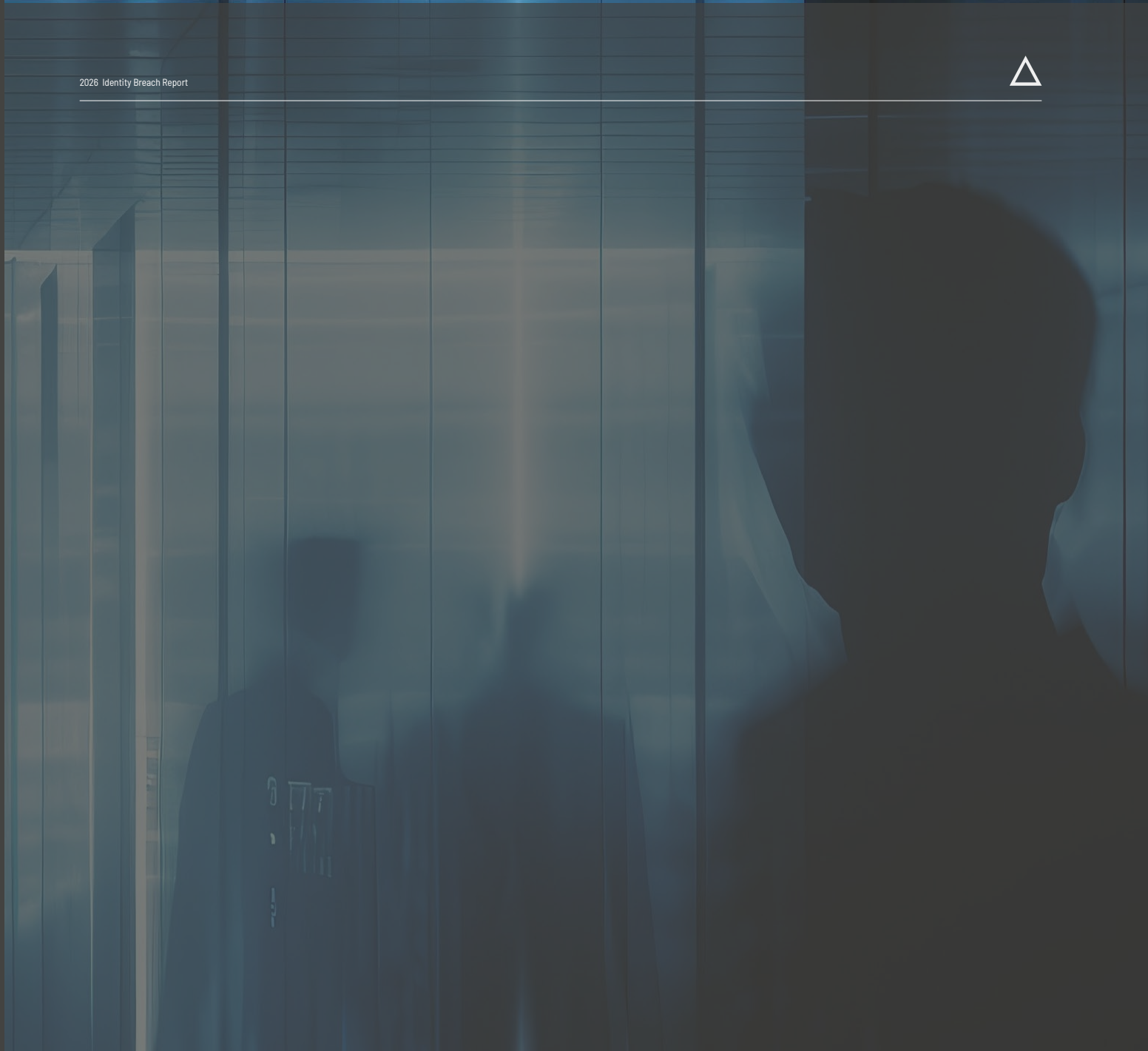
The Constella Data Lake has solidified its position as the world's most comprehensive repository of identity risk, now holding over **54.6 billion curated records**, an **80% increase year-over-year.** This growth is matched by a **49% increase in curated attributes**, which now total over **429 billion**. This expansion is not merely a matter of scale but of depth: each record is enriched with multifaceted attributes that allow for complex link analysis and the mapping of entire identity clusters.

This unprecedented visibility into the global identity attack surface provides the foundation for our predictive analytics, allowing organizations to visualize their exposure across 125 countries and dozens of languages.

5 | THE INFOSTEALER PANDEMIC: 51.7 MILLION PACKAGES

Infostealers have become the primary engine of the modern identity threat landscape, with Constella processing **51.7 million packages** in 2025, a **72% increase YoY**. Beyond simple credential theft, these packages are used to harvest live session cookies, system metadata, and autofill data, identifying **24.8 million unique infected devices** (+77% YoY). This surge reflects the industrialization of "Malware-as-a-Service," where even low-skilled actors can deploy sophisticated stealers to bypass Multi-Factor Authentication (MFA) via session hijacking.

The geographic distribution of these infections, led by regions like Brazil, Turkey, and India, highlights a globalized threat that targets the modern remote workforce at its most vulnerable endpoints.



Key Threat Dimensions:

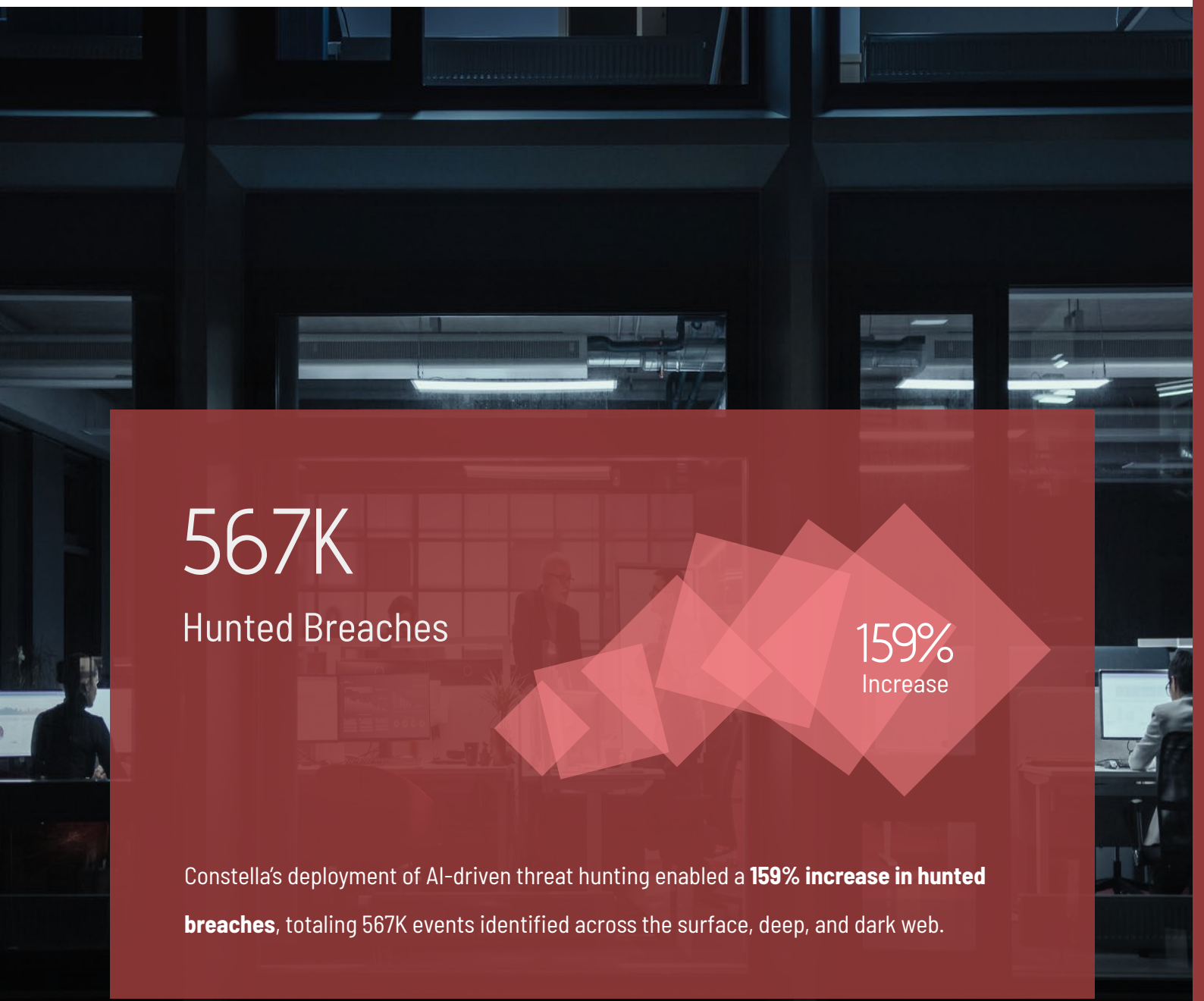
Mapping the 2026 Identity Risk Landscape

01

Identity at Machine Scale:

The Industrialized Attack

In 2025, the identity attack surface shifted from a collection of isolated incidents into a fully industrialized, machine-scale threat economy. Driven by the rapid adoption of **Agentic AI** and automated harvesting, cybercriminals are no longer "breaking in"—they are simply logging in with stolen, high-fidelity credentials.



567K

Hunted Breaches

159%

Increase

Constella's deployment of AI-driven threat hunting enabled a **159% increase in hunted breaches**, totaling 567K events identified across the surface, deep, and dark web.

The Scale of Exposure

The sheer volume of data circulating in 2025 underscores this industrialization. Constella's deployment of AI-driven threat hunting enabled a **159% increase in hunted breaches**, totaling more than **567K events** identified across the surface, deep, and dark web. This operation resulted in the curation of **27.9 billion records** (+135% YoY) and **141.2 billion attributes** (+173% YoY) being ingested into the Data Lake.

- **Records Curated:** 27.9 Billion (+135% YoY).
- **Total Data Lake Volume:** 54.6 Billion+ curated records and 429 Billion+ curated attributes.
- **Verified Breaches:** More than 8K high-fidelity breaches inserted after ML-based deduplication.

Trend Alert:

The Rise of "Combo" Consolidation

While raw threat volume has surged, the composition of identity data is shifting toward higher density and greater precision. The 66% decrease in Combo Breaches reflects a technical move toward high-density consolidation, synthesizing thousands of individual leaks into unified delta compilations to eliminate "zombie" noise while increasing total record density. Conversely, breaches containing PII surged by 661%, indicating that as technical consolidation streamlines historical data, the availability of fresh, PII-rich intelligence directly from primary breaches has reached an all-time high.

The Threat Perspective

"The identity perimeter has effectively evaporated. In an era where nearly 70% of breached credentials are found in plaintext, the challenge is no longer just preventing unauthorized access, it's outrunning the automated systems that weaponize these identities at machine speed."

— Andres Andreu, CEO, Constella

Key Recommendations

Adopt Identity-Centric Security: Transition defense strategies from network perimeters to continuous identity threat monitoring.

Assume Compromise: Given that unique emails in the Data Lake grew significantly, organizations must assume a portion of their credentials are already exposed and mandate unique, strong passwords managed via enterprise tools.

Leverage Machine-Scale Defense: Meet automated threats with automated responses. Use AI-driven intelligence to detect and block credential stuffing and anomalous login patterns in real-time.

Identity at Machine Scale:

The Industrialized Attack

Summary

The transition to an industrialized attack surface means that identity risk is now a permanent, high-velocity liability. Constella's Data Lake now encompasses over **1 Trillion raw identity attributes** identified globally, with **429 Billion high-fidelity, curated attributes** currently indexed and deduplicated for real-time monitoring. This unprecedented depth of exposure means the average employee or executive is no longer just a single data point, but a fully mapped digital persona.

430 Billion

Curated Attributes Indexed

02

The Infostealer Surge:

51.7 Million Packages

Infostealers have solidified their position as the primary engine of the modern identity threat landscape. No longer limited to simple password theft, these sophisticated malware strains harvest entire digital personas, including live session cookies, system metadata, and autofill data, allowing adversaries to bypass Multi-Factor Authentication (MFA) and gain persistent access to corporate environments.

The Scale of the Pandemic

In 2025, Constella processed **51.7 million infostealer packages,** marking a **72% increase** year-over-year. This surge in volume is accompanied by a significant expansion in the number of unique infected devices identified, which rose to **24.8 million**, a **77% increase** from 2024.

Packages Processed: 51,731,348 (+72% YoY).

Unique Infected Devices: 24,809,114 (+77% YoY).

Corporate Exposure: **78% of recently breached companies had corporate credentials appearing in infostealer logs** within six months of their breach.

Trend Alert:

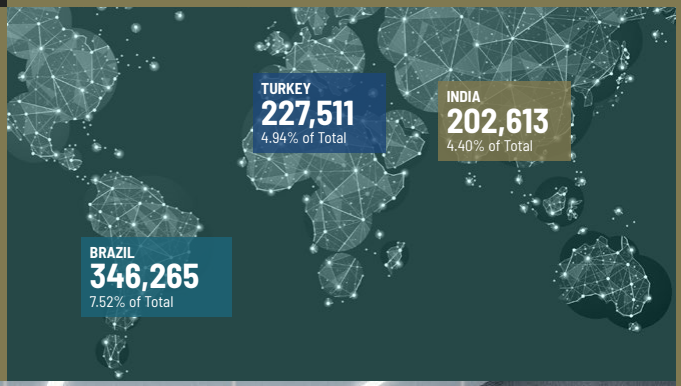
Geographic Shifts and "Malware-as-a-Service"

The democratization of cybercrime via the "Malware-as-a-Service" model has led to a globalized threat. Infostealer infections are highly concentrated in regions with high software piracy and aggressive social engineering campaigns.

The top three most impacted countries in 2025 by % of total packages: Brazil, Turkey, and India.

24,809,114

Unique Infected Devices



The Threat Perspective

"The identity perimeter has effectively evaporated. In an era where nearly 70% of compromised credentials are found in plaintext, the challenge is no longer just preventing unauthorized access, it's outrunning the automated systems that weaponize these identities at machine speed."

— Andres Andreu, CEO, Constella

Key Recommendations

Monitor Personal Device Exposure: With the rise of remote work, ensure threat intelligence extends to personal devices where employees may sync corporate credentials.

Invalidate Compromised Sessions: When an infostealer alert is triggered, security teams must move beyond simple password resets and proactively invalidate all active session tokens.

Deploy EDR and MFA Hardening: Use Endpoint Detection and Response (EDR) to catch stealers early and implement phishing-resistant MFA (such as FIDO2) to mitigate the risk of session hijacking.

The Infostealer Surge:

51.7 Million Packages

Summary

The **72% growth** in infostealer packages represents a shift toward a "silent" identity pandemic. By harvesting multiplexed attack vectors, credentials, cookies, and system tokens, from **24.8 million devices**, adversaries have built an automated pipeline for bypassing traditional security perimeters. Organizations must adopt continuous, identity-centric monitoring to detect these exposures before they are weaponized into full-scale breaches.

*Crowdstrike 2025 Global Threat Report

03

The Plaintext Crisis:

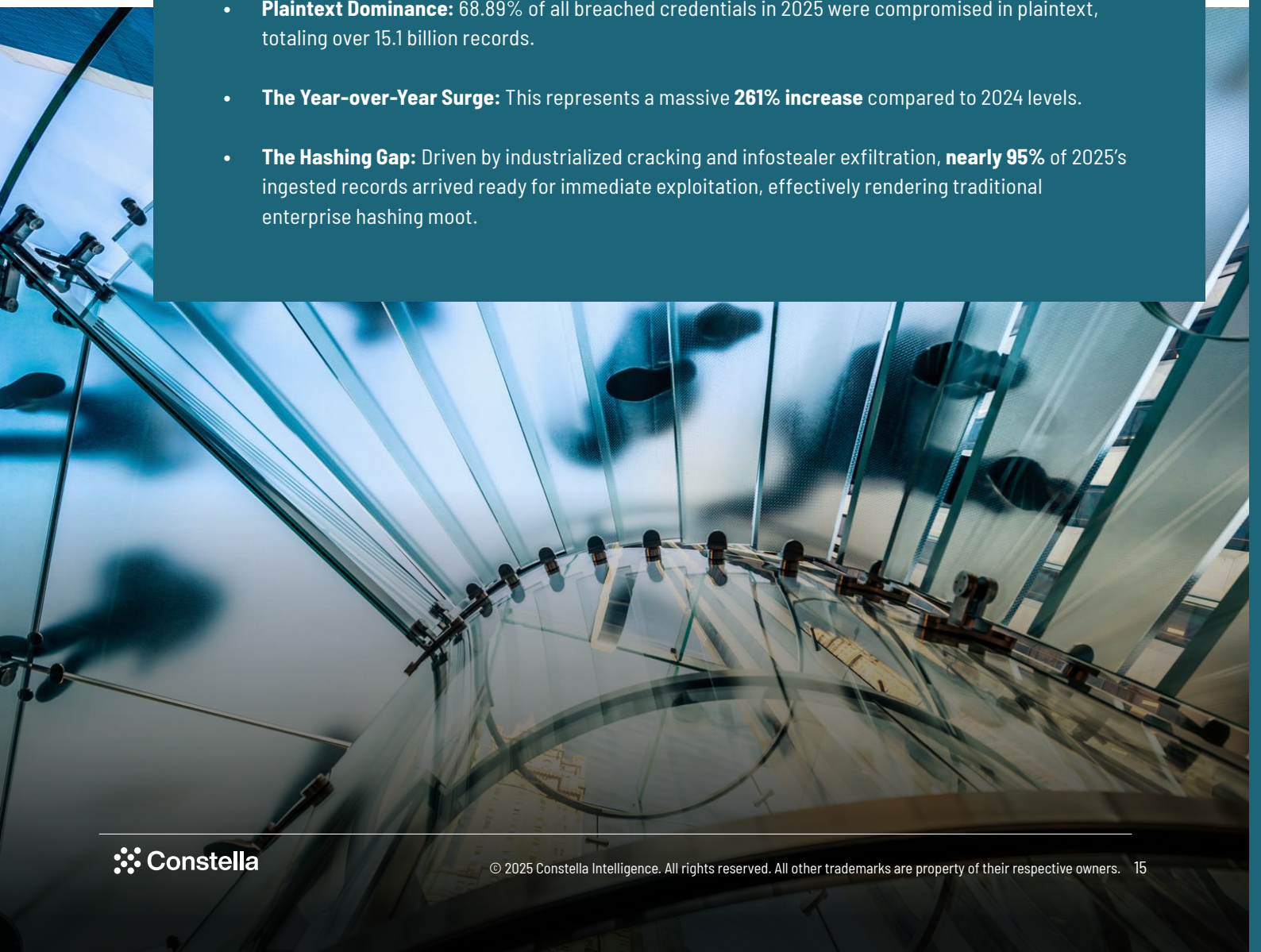
Identity Risk as a Boardroom

The 2025 landscape was defined by the **industrialization of credential recovery and operationalization.** The observed surge in plaintext material is not a reflection of declining enterprise standards, but rather the growing efficiency of the cybercriminal ecosystem in converting hashed data into actionable weapons. Through **GPU-accelerated cracking**, optimized wordlists and rulesets, and the rise of **Infostealers**, which harvest credentials directly from memory post-authentication, adversaries have effectively bypassed traditional hashing hurdles.

The Scale of Exposure

Constella's 2025 metadata analysis uncovered a staggering increase in the volume of unprotected credentials entering the threat economy. The majority of breached identities now provide attackers with immediate, clear-text access to sensitive accounts.

- **Plaintext Dominance:** 68.89% of all breached credentials in 2025 were compromised in plaintext, totaling over 15.1 billion records.
- **The Year-over-Year Surge:** This represents a massive **261% increase** compared to 2024 levels.
- **The Hashing Gap:** Driven by industrialized cracking and infostealer exfiltration, **nearly 95%** of 2025's ingested records arrived ready for immediate exploitation, effectively rendering traditional enterprise hashing moot.



TREND ALERT

Weak Algorithms & The "Instant-ATO" Era

Even when encryption is present, it is often obsolete. Among the 15.7 billion hashing algorithm detections analyzed, weak and easily reversible methods dominate the landscape. Legacy and Weak Algorithms (including MD5 and MySQL323) dominate the landscape of recoverable data. The high prevalence of MD5-based hashes reflects the ongoing recycling of historical datasets and the speed at which modern attackers can collide and resolve these legacy protections.

The Threat Perspective

"The 'plaintext problem' has evolved; it is no longer a primary metric of insecure enterprise storage. Instead, it reflects an industrialized pipeline where infostealers exfiltrate credentials directly from browsers and endpoint memory—bypassing hashing entirely—while GPU-optimized cracking farms convert historical hash sets into actionable, clear-text weapon libraries at a global scale."

— Cybersecurity Insiders, 2026 State of Identity Security Report

Key Recommendations

Mandate Adaptive Hashing: Organizations must enforce modern hashing standards such as Argon2, Bcrypt, or scrypt for all internal and customer-facing databases.

Implement Phishing-Resistant MFA: Given the prevalence of plaintext passwords, traditional MFA is no longer a sufficient fail-safe. Move toward FIDO2-compliant hardware keys or biometrics to mitigate the risk of credential-based entry.

Executive Exposure Monitoring: Prioritize continuous monitoring for high-value targets. Because executive identities are frequently targeted for BEC and impersonation, any plaintext exposure for these individuals is a "zero-hour" emergency.

The Plaintext Crisis:

Identity Risk as a Boardroom

Summary

The **261% surge** in plaintext password exposure represents a critical turning point for organizational risk. In 2025, attackers no longer needed sophisticated decryption tools to access corporate networks; they simply utilized the billions of clear-text credentials made available by poor security hygiene. To protect the organization and its leadership, security teams must treat identity exposure as a systemic risk, moving toward a posture of continuous verification and automated identity protection.

261% Surge

In Plaintext Password Exposure

04

Eternal Exposure:

The Persistent Recycling of Identity

The digital threat landscape is increasingly characterized by the "zombie" nature of identity data. Once a set of credentials or personal identifiers is leaked, it enters a cycle of perpetual reuse, being repackaged and redistributed across the deep and dark web for years. While the volume of raw data continues to explode, the underlying pool of unique human identities remains finite, leading to a landscape defined by high redundancy and persistent risk.

The Scale of Recycling

Data from 2025 highlights a significant divergence between the growth of raw record ingestion and the growth of unique identity discovery. This gap underscores the extent to which threat actors rely on recycled data to fuel their operations.

- **Divergent Growth:** While total curated records inserted into the Data Lake increased by **135%** in 2025 (reaching 27.9 billion), the number of **unique emails ingested grew by only 11%.**
- **Unique Identity Ratio:** This indicates that Constella is seeing an average of nearly **six records for every one unique identity**, providing deeper context for each persona but also demonstrating the repetitive nature of credential leaks.
- **Quality Filtering:** While Hunted Breaches grew by 159%, we now synthesize thousands of individual Combo (-66%) and URL Log (-79%) leaks into single **delta compilations** prior to ingestion. This process eliminates "zombie" data and recycled noise, ensuring that while the total volume of unique records actually increased, the Data Lake remains streamlined and high-fidelity.

STRATEGIC CONSOLIDATION

Enhancing Record Density

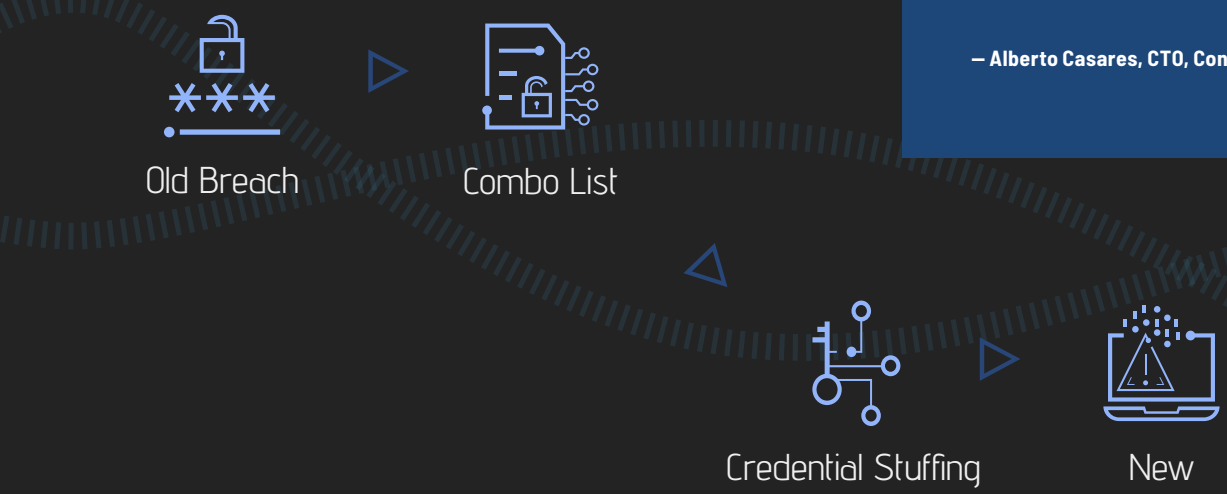
The 2025 landscape was defined by a technical shift toward **high-density data compilations.** Total records increased Year-over-Year as we synthesized thousands of individual leaks into unified **delta compilations**. By aggregating these massive collections of credentials into daily and annual files, the data enters the Data Lake with higher fidelity and less "zombie" noise, effectively weaponizing recycled identities for high-velocity credential stuffing and account takeover campaigns.

The Threat Perspective:

The Persistence of Credential Reuse

"The real danger isn't just the consolidation of data, but the persistence of human behavior. Password reuse across personal and professional accounts turns a single, years-old leak into a recurring and scalable attack opportunity. In the eyes of modern security and regulations like NIS2, an exposed password is a compromised password, regardless of its strength."

— Alberto Casares, CTO, Constella



*2024 identity breach report: https://constella.ai/2024-identity-breach-report

Key Recommendations

Enforce Password Rotations After Known Leaks: Do not rely on "standard" rotation cycles; force credential resets immediately when an identity is found in a new compilation or combo list, even if the source data is aged.

Deduplicate Threat Intelligencee: Organizations should ensure their threat intelligence providers use rigorous deduplication (like Constella's NetNew pipeline) to avoid "alert fatigue" caused by recycled data.

Implement Continuous Identity Monitoring: Because data is recycled indefinitely, monitoring must be persistent. A "clean" scan today does not account for an identity being re-released in a different compilation tomorrow.

Eternal Exposure:

The Persistent Recycling of Identity

Summary

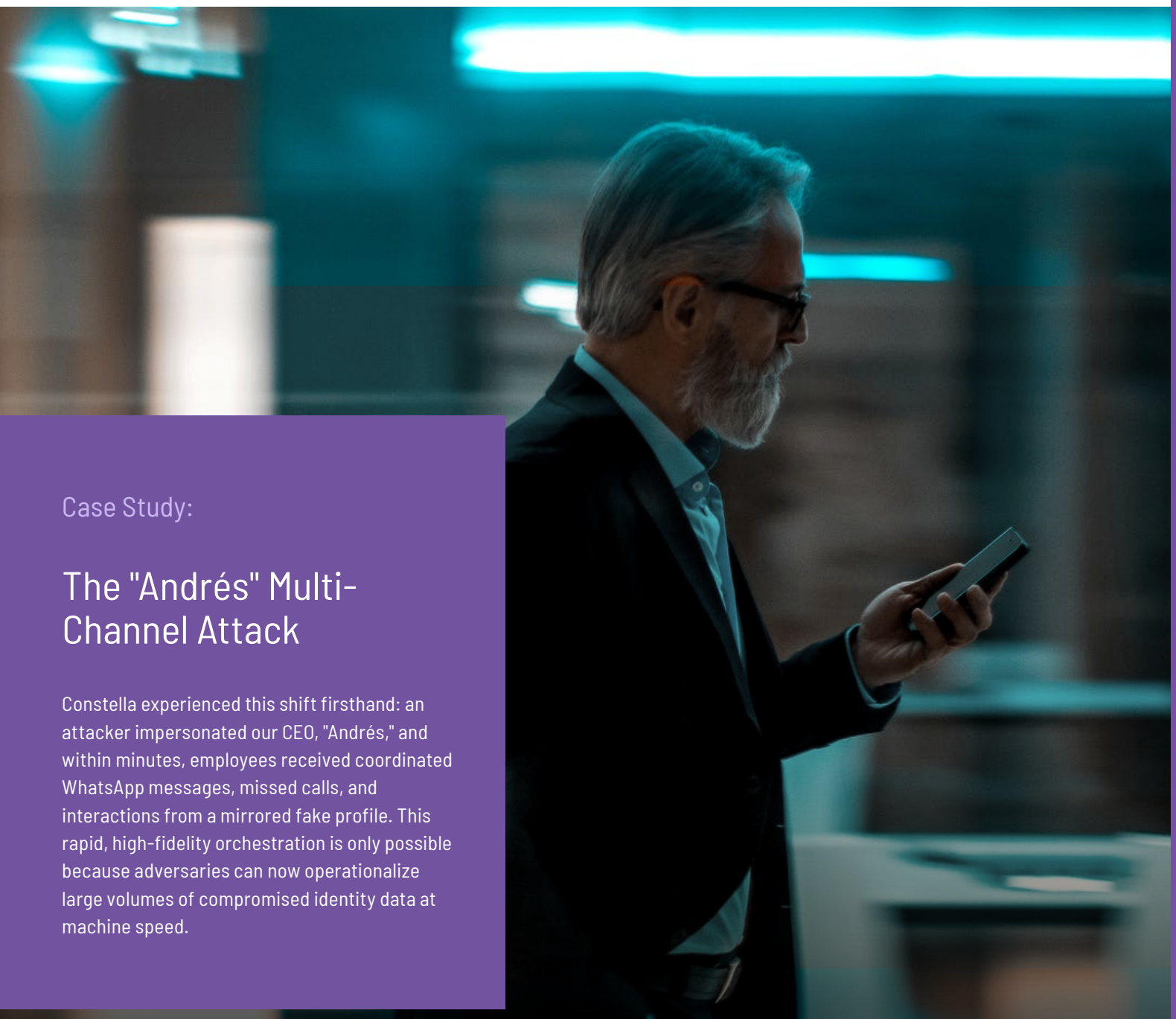
The divergence between **11% growth in unique emails and 135% growth in total records** reveals a dangerous expansion of the digital footprint per identity rather than simple data recycling. This growing density provides adversaries with a high-fidelity map of individual behavior, integrating historical passwords, device signals, and location patterns to enable AI-driven **password prediction** and sophisticated identity correlation. As the volume of attributes per individual swells, each identity becomes progressively more "attackable," shifting the threat from simple credential stuffing to industrialized, personalized impersonation that demands a model of continuous identity vigilance.

05

Targeted Exploitation:

Executive Digital Exposure

As the identity attack surface becomes increasingly industrialized, threat actors are refining their focus on high-value targets. Executive Digital Exposure (EDE) has moved beyond simple privacy concerns to become a critical security vulnerability. CXOs and senior leadership are now the primary targets for highly automated, identity-based attacks that leverage leaked PII to facilitate Business Email Compromise (BEC), financial fraud, and corporate espionage.



Case Study:

The "Andrés" Multi-Channel Attack

Constella experienced this shift firsthand: an attacker impersonated our CEO, "Andrés," and within minutes, employees received coordinated WhatsApp messages, missed calls, and interactions from a mirrored fake profile. This rapid, high-fidelity orchestration is only possible because adversaries can now operationalize large volumes of compromised identity data at machine speed.

The Sophistication of Leadership Risk:

Beyond Basic Impersonation

In 2025, the risk to leadership transitioned from simple spoofing to a high-velocity, **AI-enabled orchestration.** Adversaries no longer rely solely on email/password pairs; they now weaponize a rich tapestry of **Identity Intelligence,** including phone numbers, corporate hierarchies, personal interests, and real-world contact points to craft attacks that are indistinguishable from legitimate business operations.

- **The Weaponization of Professional Identity:** While LinkedIn attributes represent **6.32%** of social media exposures, their true value lies in mapping organizational structures. When combined with AI automation, threat actors can instantly identify an executive's colleagues and photo, obtaining corporate and personal phone numbers to launch scalable, multi-channel attacks.
- **High-Fidelity PII as an Impersonation Engine:** The **661% surge** in compromised breaches provides the granular signals needed for "Deep-Dive" social engineering. Attackers use this data to move beyond basic phishing, instead utilizing compromised legitimate accounts, often via infostealer infections, to communicate from real inboxes and trusted channels.
- **The Failure of Traditional Perimeters:** With **68.89%** of credentials available as actionable material, the threat is increasingly coming from inside the organization. Traditional defenses often fail because the "sender" is a verified corporate account, making an Identity Risk Posture, continuously checking if communications are linked to recently compromised identities, the only viable defense.

The Threat Perspective

"Cybercriminals are increasingly moving away from broad-based attacks to focus on high-impact targets. Executives are the 'crown jewels' of identity; a single compromised CXO identity can bypass millions of dollars in technical controls through social engineering alone."

— FBI Internet Crime Complaint Center (IC3) - 2025 Cyber Trends Analysis

Key Recommendations

Implement Executive Shadow Monitoring: Security teams must monitor the deep and dark web for the personal email addresses, phone numbers, and home addresses of senior leadership, as these are often the first points of compromise.

Mandate Phishing-Resistant MFA: For all executive accounts, move beyond SMS or app-based MFA to hardware security keys (FIDO2) to prevent session hijacking via infostealers.

Establish Out-of-Band Verification: Create strict protocols for financial transactions or sensitive data transfers that require verbal or secondary verification, neutralizing the impact of a compromised identity.

Targeted Exploitation:

Executive Digital Exposure

Summary

Executive identities are the high-yield currency of the modern threat economy. With the **135% increase** in curated records and the prevalence of **plaintext passwords**, the barrier to impersonating a leader has never been lower. Organizations must recognize that protecting the digital footprint of their leadership is no longer a luxury, it is a foundational requirement for enterprise resilience in 2026.

The barrier to impersonating a leader has never been lower.

06

The Identity Domino Effect:

Cascading Supply Chain

In an interconnected digital economy, a security failure at a single entry point rarely stays contained. The "Identity Domino Effect" describes how a breach at a third-party vendor or service provider cascades through the supply chain, weaponizing stolen credentials to compromise downstream partners and customers. As organizations tighten their own perimeters, adversaries are increasingly targeting the softer targets within the supply chain to gain "trusted" access to larger enterprise targets.

The Scale of Exposure

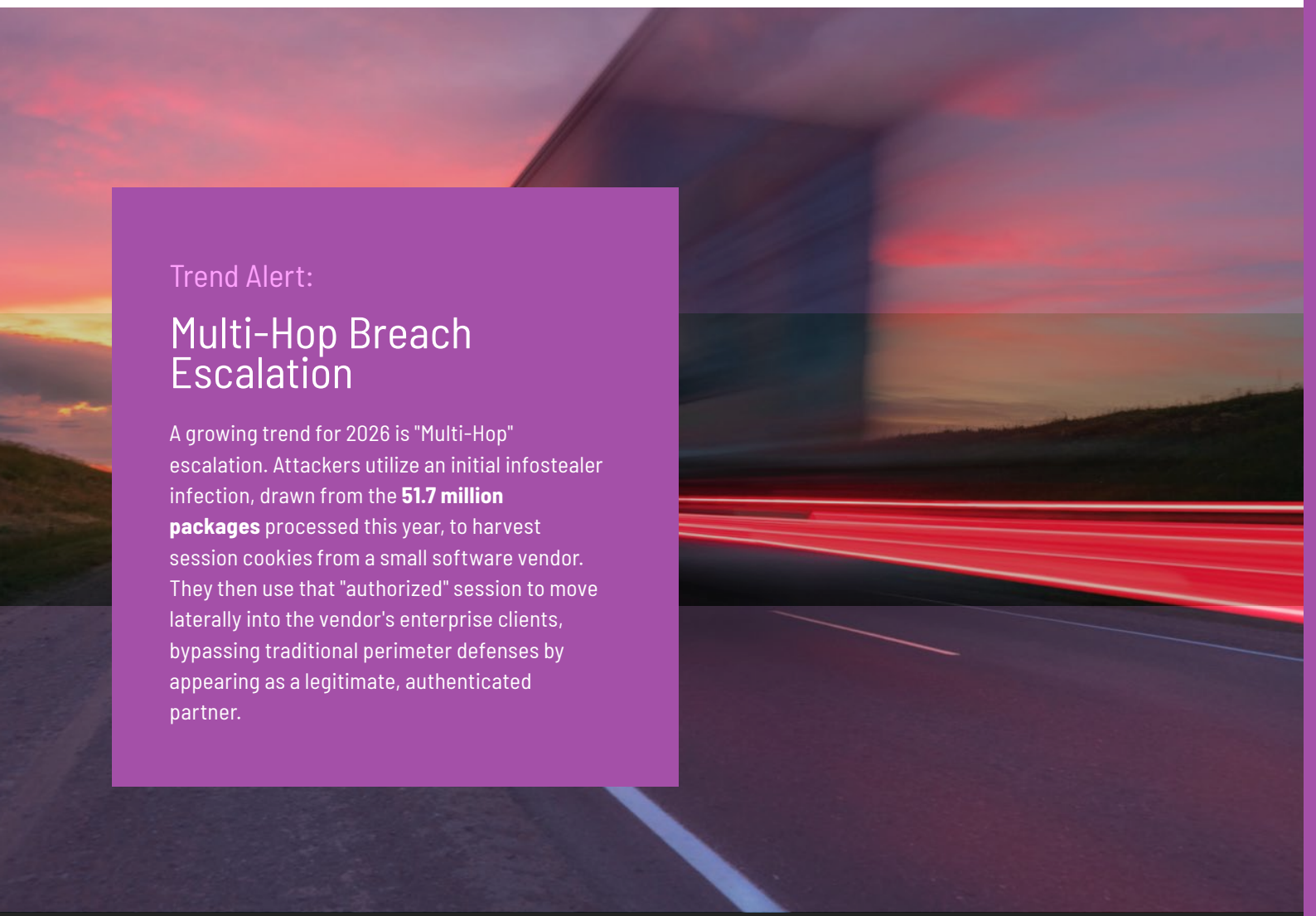
The 2025 data reveals that the most impactful breaches occurred within sectors that serve as critical infrastructure for other businesses, such as telecommunications, technology, and financial services.

- **Infrastructure Targets:** A major 2025 breach included **telcel.br (36.6M records)** which provide services integrated into thousands of other business workflows.
- **Sector Vulnerability:** The **Government sector** saw a **569% YoY increase** in breaches, while the **Services sector** grew by **238%**. These sectors are central nodes in the supply chain, where one compromised identity can grant access to multiple sensitive environments.
- **The Credential Bridge:** With **27.9 billion curated records ingested** in 2025, attackers have an exhaustive library of credentials to attempt across different platforms in the supply chain.

The Threat Perspective

"Supply chain attacks are no longer just about software vulnerabilities; they are about identity vulnerabilities. By compromising a single trusted identity within a vendor's ecosystem, an attacker can leapfrog over the most sophisticated defensive layers of an entire conglomerate."

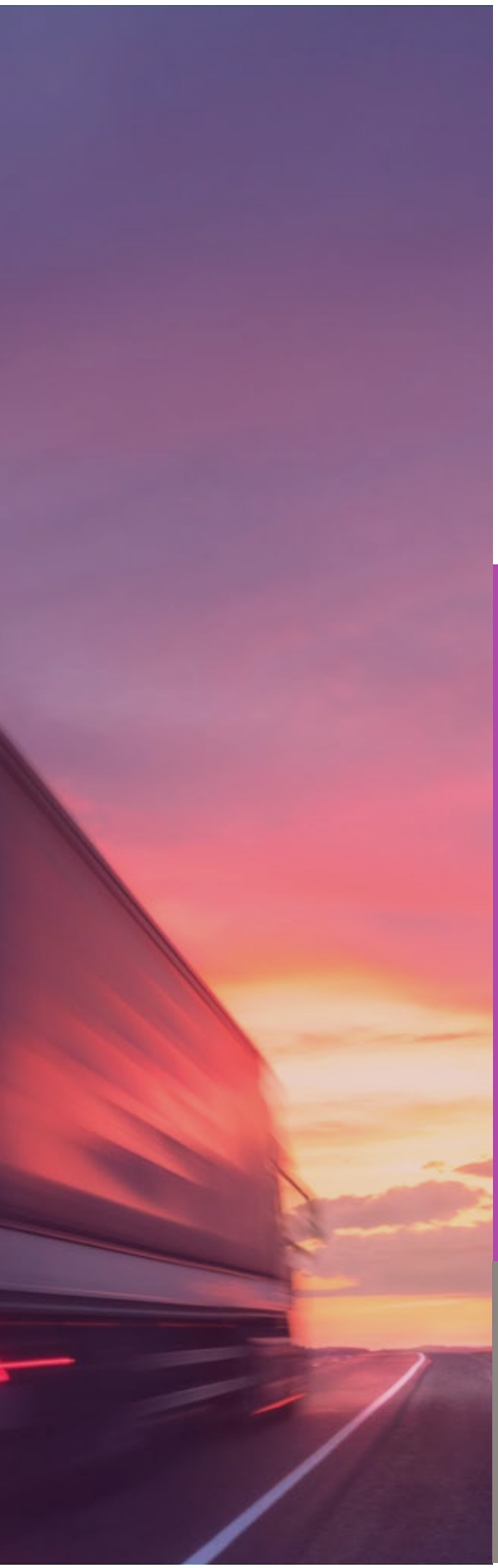
— European Union Agency for Cybersecurity (ENISA), 2025 Threat Landscape Report



Trend Alert:

Multi-Hop Breach Escalation

A growing trend for 2026 is "Multi-Hop" escalation. Attackers utilize an initial infostealer infection, drawn from the **51.7 million packages** processed this year, to harvest session cookies from a small software vendor. They then use that "authorized" session to move laterally into the vendor's enterprise clients, bypassing traditional perimeter defenses by appearing as a legitimate, authenticated partner.



Key Recommendations

Implement Tiered Identity Verification: Apply more stringent authentication requirements for users accessing your environment from partner domains or third-party service provider networks.

Continuous Vendor Exposure Monitoring: Don't just monitor your own domains; track the digital exposure of your critical supply chain partners to identify leaked credentials before they can be used against your organization.

Zero-Trust Session Management: Move beyond persistent logins. Implement short-lived session tokens and continuous re-authentication for high-privileged supply chain integrations to mitigate the risk of session hijacking.

The Identity Domino Effect:

Cascading Supply Chain

Summary

The **+661% surge** in breaches containing PII in 2025 confirms that attackers are successfully targeting primary data sources. When these sources sit within your supply chain, their identity risk becomes your identity risk. To stop the "Domino Effect," organizations must treat third-party identities with the same level of scrutiny as internal ones, using machine-scale intelligence to detect cascading exposures in real-time.

When these sources sit within your supply chain, their identity risk becomes your identity risk.

07

Machine-Scale Warfare:

The Rise of Autonomous Identity Exploitation

The transition from 2025 to 2026 marks the era of the "AI Economy," where the global digital landscape has shifted from AI-assisted to **AI-native.** For the first time, we are seeing the widespread deployment of **Agentic AI**: autonomous systems capable of reasoning, planning, and executing multi-step operations without human intervention. While this has provided defenders with a critical force multiplier, it has simultaneously gifted adversaries with a toolset for machine-scale, hyper-personalized warfare.

The Scale of the AI Revolution

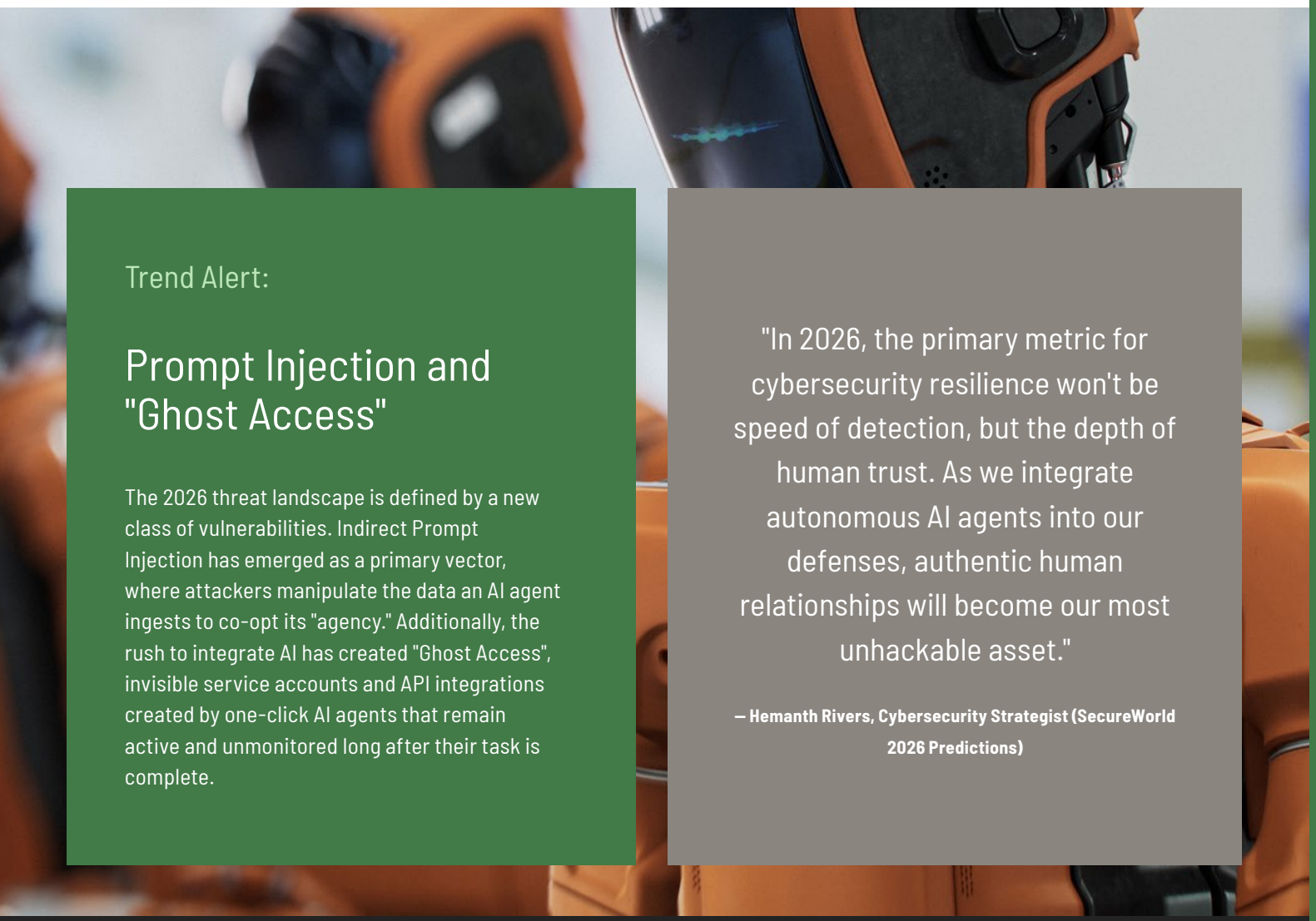
In 2025, Constella leveraged these very technologies to achieve a **159% increase in hunted breaches**, proving that automation is the only way to navigate an ocean of **429 billion attributes**. However, the same efficiency is being mirrored by threat actors.

- **Automation in the Kill Chain:** AI agents are now capable of managing the entire attack lifecycle, from autonomous reconnaissance using leaked PII to the execution of "vibe-coded" malware that morphs in real-time to evade detection.
- **The Non-Human Identity (NHI) Surge:** By the end of 2026, it is predicted that **40% of enterprise applications** will feature task-specific AI agents. In many environments, these non-human identities already outnumber human employees by a ratio of **82 to 1**, creating a massive, unmanaged shadow identity surface.
- **Speed of Response:** Mature security teams using agentic AI for triage and enrichment have seen a **30% to 50% reduction in Mean Time to Respond (MTTR)**, a necessary evolution to keep pace with "Instant-ATO" attacks.

The Threat Perspective

"2025 was the year threat intelligence reached industrial scale. By deploying Agentic AI to expand our detection by 159%, we've proved that the only way to defend against an automated adversary is with a machine-scale defense that curates billions of attributes in real-time."

— Andres Andreu, CEO, Constella Intelligence



Trend Alert:

Prompt Injection and "Ghost Access"

The 2026 threat landscape is defined by a new class of vulnerabilities. Indirect Prompt Injection has emerged as a primary vector, where attackers manipulate the data an AI agent ingests to co-opt its "agency." Additionally, the rush to integrate AI has created "Ghost Access", invisible service accounts and API integrations created by one-click AI agents that remain active and unmonitored long after their task is complete.

"In 2026, the primary metric for cybersecurity resilience won't be speed of detection, but the depth of human trust. As we integrate autonomous AI agents into our defenses, authentic human relationships will become our most unhackable asset."

— Hemanth Rivers, Cybersecurity Strategist (SecureWorld 2026 Predictions)

AI Simplified:

The Four Levels of AI in Cybersecurity

Understanding AI in 2026 is about the shift from systems that follow instructions to systems that pursue goals.

Traditional AI & Machine Learning

(The Pattern Finder)

- **What it is:** Uses history and statistics to find "known bad" patterns.
- **Cyber Role:** Identifying anomalies in massive data sets, like Constella's **54.6B+ record Data Lake.**
- **Bottom Line:** Reactive. It tells you what has happened.

AI Agents

(The Task Completer)

- **What it is:** Task-driven systems that execute specific steps, like scheduling or data retrieval, based on set rules.
- **Cyber Role:** Automating repetitive attacks, such as validating stolen credentials at scale or solving CAPTCHAs.
- **Bottom Line:** Efficient. They handle the "grunt work" of a breach.

Generative AI

(The Content Creator)

- **What it is:** Uses Large Language Models (LLMs) to create new text, images, or code.
- **Cyber Role:** Powering hyper-realistic phishing emails and deepfake videos using just seconds of stolen audio.
- **Bottom Line:** Influential. It excels at convincing human victims.

Agentic AI

(The Autonomous Strategist)

- **What it is:** Advanced systems that can **reason, plan, and act independently** to achieve a high-level goal.
- **Cyber Role:** Managing an entire attack lifecycle—from initial reconnaissance to active exploitation—with minimal human oversight.
- **Bottom Line:** Proactive. It moves at machine speed to outmaneuver traditional defenses.

Key Recommendations

Implement AI Governance & Circuit Breakers: Treat AI agents as high-privilege "digital employees." Deploy runtime "AI firewalls" to detect and block malicious prompt injections and tool misuse before the agent can execute a harmful command.

Audit Non-Human Identities (NHIs): Use automated discovery tools to map the sprawl of AI-driven service accounts and API keys. Apply the same Zero-Trust rigors, least privilege and continuous monitoring, to agents as you do to human users.

Prepare for "Q-Day" and Cryptographic Agility: With quantum computing advancements threatening traditional encryption, prioritize the migration of AI data pipelines to post-quantum standards to ensure the long-term integrity of the models.

The Identity Domino Effect:

Cascading Supply Chain

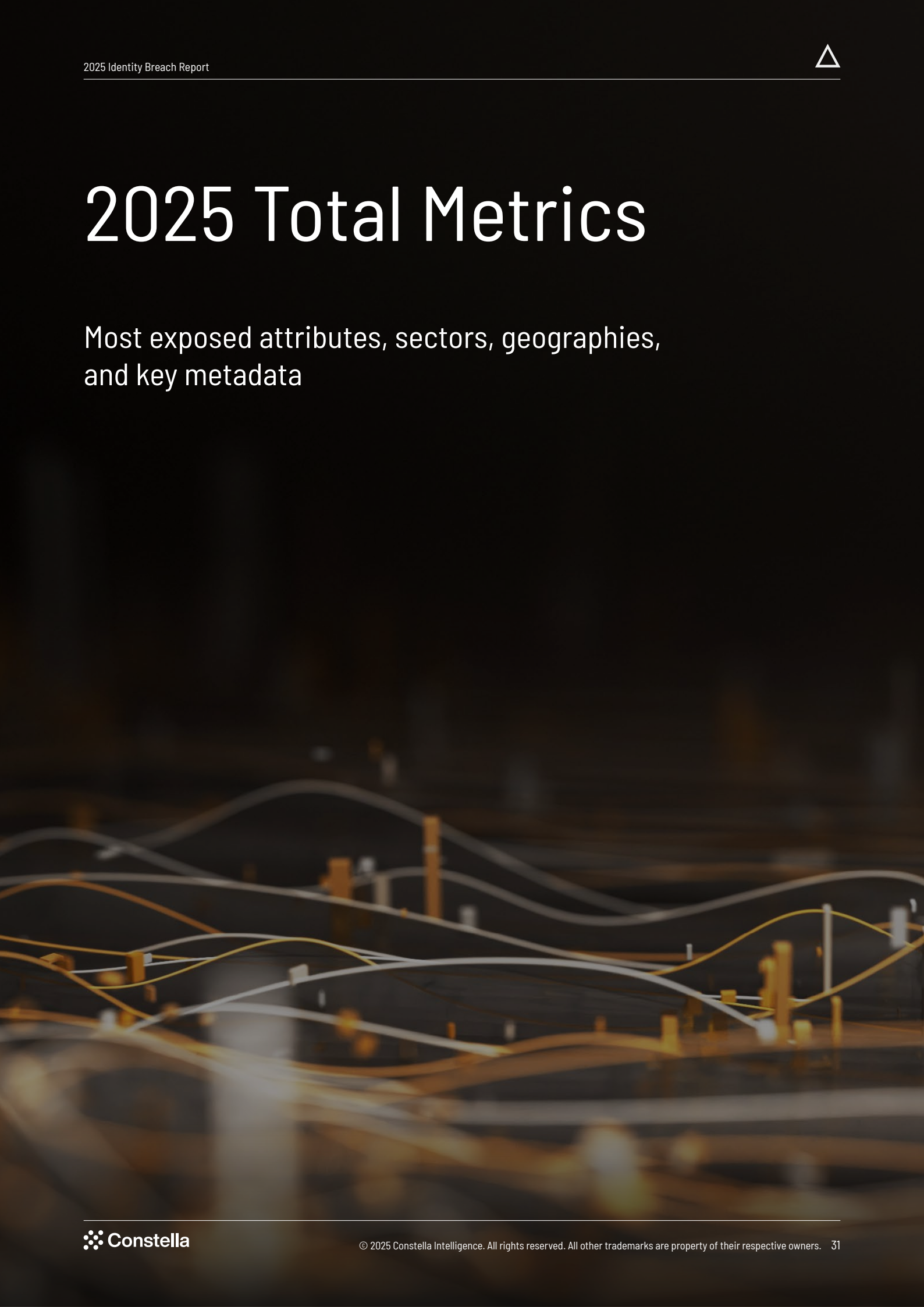
Summary

Agentic AI is the defining force of 2026, representing both the greatest opportunity and the most significant risk to identity security. As autonomous agents become the primary way we interact with data, the line between innovation and intrusion has blurred. Organizations that balance autonomy with rigorous identity-centric control tower oversight will thrive; those that gamble on unsecured automation will face the reality of an adversary that moves faster than human thought.

In 2025, Constella utilized this technology to hunt **159% more breaches** than the previous year. The challenge for 2026 is that adversaries are now using the same **Agentic AI** to automate the exploitation of the **68.89% of credentials stored in plaintext.**

2025 Total Metrics

Most exposed attributes, sectors, geographies, and key metadata



2025 Total Breached Identity Metrics

The industrialization of cybercrime reached peak velocity in 2025. Leveraging Agentic AI to automate the discovery and harvesting of exposed data, Constella identified and curated a record-breaking volume of identity intelligence. This year's metrics reflect a fundamental shift: a move away from recycled "combo" data toward high-fidelity, net-new PII exposures.

Ingested Breaches

8,460

A 16% decrease due to consolidation of Combo and URL Logs

Ingested Records

27.9 B

27,903,740,719 (up 135% YoY)

Attributes Inserted

141.2 B

141,218,292,500

Total Breaches Hunted

567,061

+159% Y/Y

Total Records Curated

27.9 B

+135% Y/Y

Breaches Containing PII

8,460

+661% Y/Y

Unique Emails Ingested

1.35 B

+11% Y/Y

Key Findings:

The Quality Pivot

- **Net-New Over Recycled:** The **66% decrease in Combo Breaches** is a direct result of our strategic focus on deduplication. By filtering out "zombie" data, we have prioritized high-fidelity, verified leakages.
- **The PII Surge:** The **661% increase in breaches containing PII** signals that threat actors are successfully targeting primary data sources, resulting in a surge of net-new PII (Names, Phone Numbers, and Physical Addresses) rather than just old credential lists.
- **Email Redundancy:** While curated records grew by **135%**, unique emails only grew by **11%.** This indicates that for every unique identity, we are now tracking approximately **six different points of exposure**, providing a 360-degree view of an individual's risk.

General Statistics:

Breaches, Combos & URL Logs

This section presents the core metrics from Constella Intelligence's threat intelligence operations during 2025. These statistics reflect the massive scale of the identity attack surface, contrasting the raw volume of threats identified (**Hunted**) with the refined, high-fidelity intelligence integrated into our **Data Lake (Ingested)**.

Data Lake Totals: **The Global Identity Repository**

The Constella Data Lake serves as the definitive source for identity risk, showing exponential growth as we index the "industrialized" leakages of 2025.

GENERAL STATISTICS	2025	2024	CHANGE Y/Y
Verified Breaches	1,365,152	1,071,817	⬆ +27%
Total Records	54,694,521,916	30,350,949,451	⬆ +80%
Identity Attributes	429,982,359,984	288,764,067,484	⬆ +49%

Data Lake Growth: The repository expanded significantly in 2025, with records increasing by **80%**. This reflects our continuous investment in automated collection and the reality of a threat landscape where a single breach can now expose hundreds of millions of PII-rich records.

Hunted Breaches 2025: **Raw Intelligence Volume**

"Hunted" data represents the raw intelligence identified through automated scanning, dark web monitoring, and intelligence partnerships prior to deduplication.

GENERAL STATISTICS	2025	2024	CHANGE Y/Y
Breaches Hunted	567,061	219,127	↑ +159%
Total Records Hunted	401,833,019,031	107,121,825,763	↑ +275%

Key Insight: The **159% increase** in hunted breaches is a direct result of deploying **Agentic AI** to scan deeper into the "unstructured" dark web. By automating the discovery phase, we are identifying nearly 3x the raw data volume compared to 2024.

Ingested Breaches 2025: **Verified & Curated Intelligence**

"Ingested" data is the refined output—verified, deduplicated, and processed through ML-based validation to ensure only actionable intelligence reaches our customers.

GENERAL STATISTICS	2025	2024	CHANGE Y/Y
Breaches Ingested	8,460	10,108	↓ -16%
Records Ingested	27,903,740,719	11,888,105,348	↑ +135%
Attributes Ingested	141,218,292,500	51,743,761,035	↑ +173%

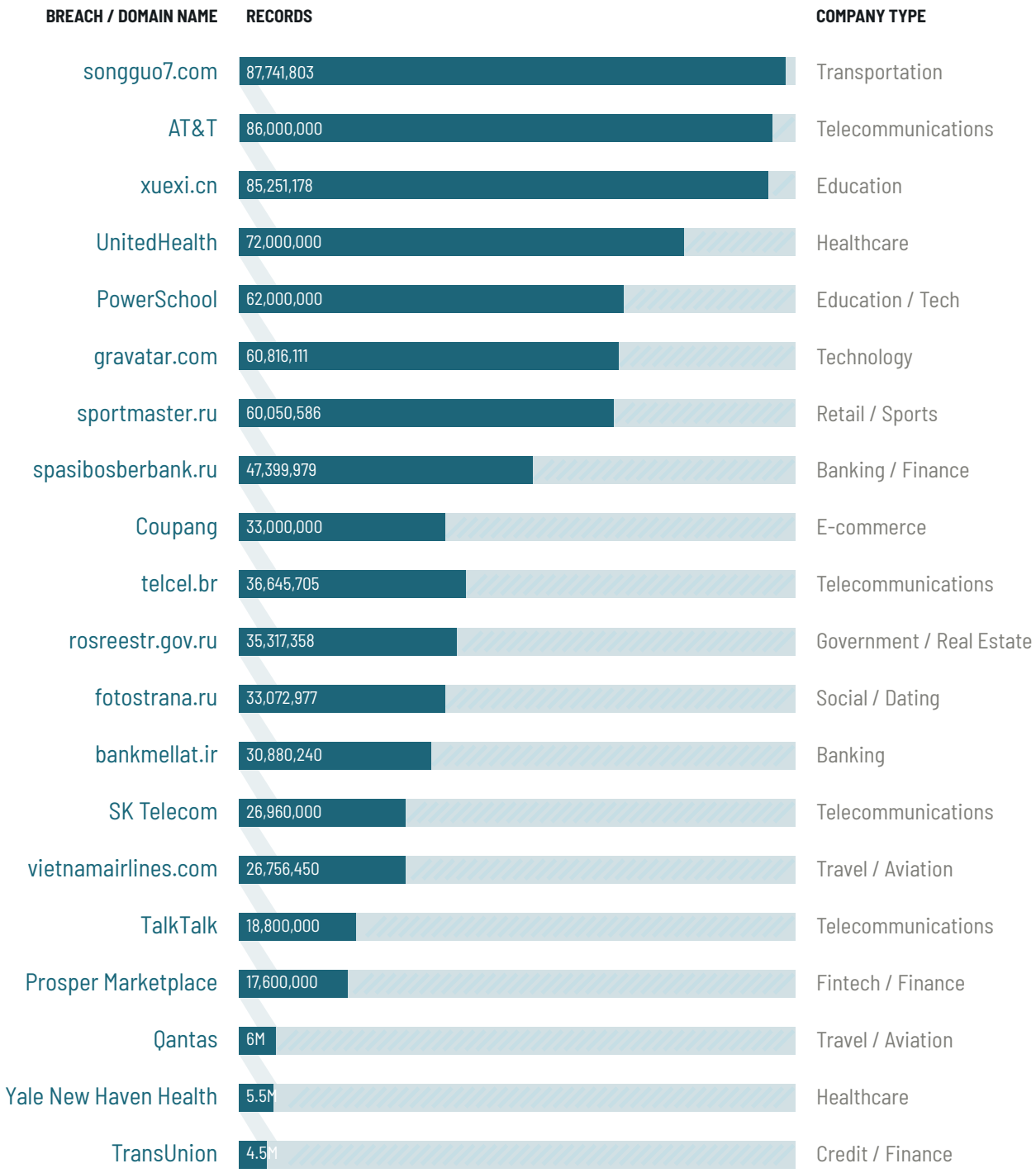
Key Insight: While the number of individual breach events inserted decreased by **16%**, the volume of **records increased by 135%**. This shift highlights the "Mega-Breach" trend of 2025, where fewer, larger incidents (like **songguo7.com** with 87M records) are being weaponized. The **173% surge in attributes** indicates that each ingested record is now richer in PII than ever before.

Summary

The 2025 data tells a story of **Quality over Quantity**. By leveraging AI-driven deduplication, we have reduced "noise" (evidenced by the 66% drop in Combo Breaches) while simultaneously increasing the volume of high-fidelity, verified records by **135%.** In an era of **429 billion attributes**, curation is the only path to effective defense.

Top 20 Breaches & Leakages

The following chart presents the 20 largest verified breaches and identity leakages affecting global domains and organizations as ingested by Constella Intelligence during 2025. These major incidents represent high-velocity exposure events that security teams should prioritize for immediate credential monitoring, proactive password resets, and session invalidation.



Impact Assessment: **The Ripple Effect**

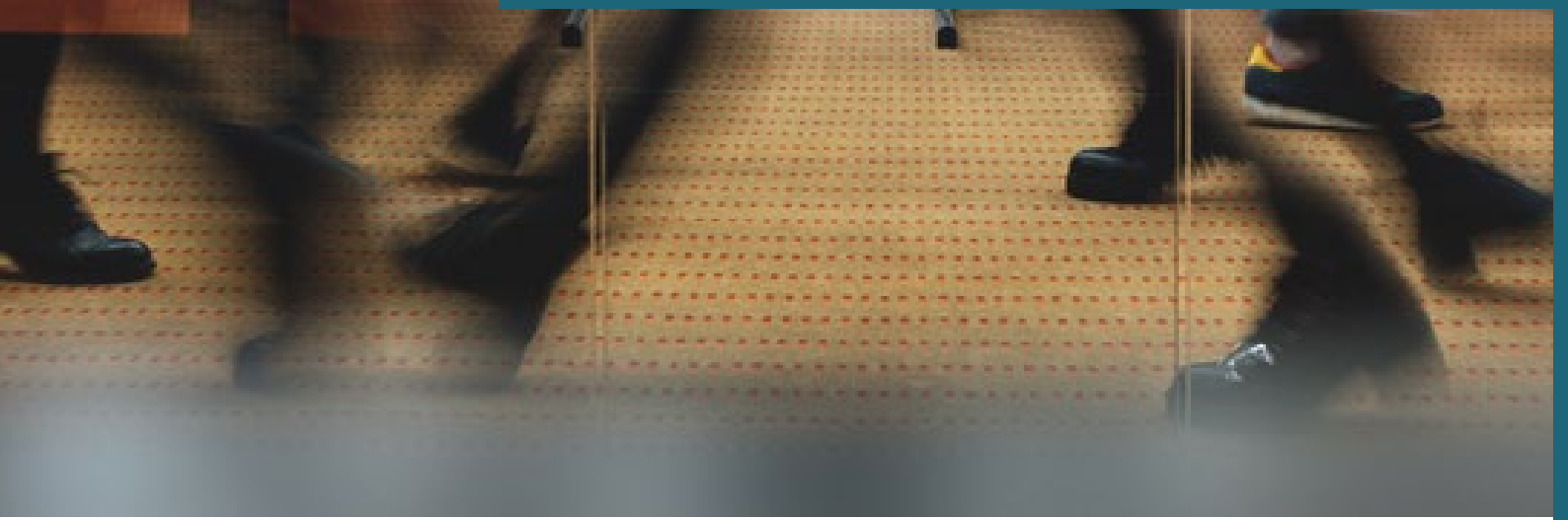
The "Identity Domino Effect" is most visible in these high-volume breaches. Because individuals frequently reuse passwords across professional and personal platforms, a single exposure in a **Retail (Sportmaster)** or **Education (PowerSchool)** breach often serves as the "master key" to an employee's corporate identity.

Critical Warning for 2026: With **68.89%** of credentials now appearing in **plaintext**, attackers no longer require time-consuming decryption phases. They are moving directly from ingestion to automated Account Takeover (ATO) across your supply chain.

Recommendation: Organizations should cross-reference these specific breach names against their employee and customer email domains. Large-scale breaches often contain "trusted" credentials that can be weaponized for lateral movement within your network.

Summary

The top 20 breaches of 2025 account for nearly one billion high-fidelity records. The diversity ofcompany types—from **Aviation** to **Government**—proves that no sector is immune to the industrialization of identity risk.



Three Year Comparison:

2025 vs 2024 vs 2023

Three-year trend analysis provides context for understanding threat landscape evolution and Constella's expanding intelligence capabilities.

GENERAL STATISTICS	2025	2024	2023	CHANGE Y/Y
Breaches Hunted	567,061	219,127	151,000	↑ +159%
Breaches Ingested	8,460	10,108	3,227	↓ -16%
Records Hunted	401,833,019,031	107,121,825,763	39,242,598,524	↑ +275%
Records Inserted	27,903,740,719	11,888,105,348	4,134,655,739	↑ +135%
Unique Emails Exposed	4,726,061,704	4,273,033,947	1,860,050,000	↑ +11%
Attributes Ingested	141,218,292,500	51,743,761,035	33,223,519,063	↑ +173%

Trend Analysis: **Precision at Scale: Solving the Identity Signal-to-Noise Problem**

The divergence between hunted **(+159%)** and ingested **(-16%)** breaches reflects improved quality filtering in Combo Breaches and URL Logs daily and yearly compilations. More threats are identified, but stricter deduplication ensures only unique, high-value data enters the Data Lake.

DATA COMPOSITION TRENDS	2025	2024	2023	CHANGE Y/Y
Breaches Containing PII	5,712	750	623	↑ +661%
Combo Breaches	2,055	6,054	2,604	↓ -66%
URL Logs	693	3,304	-	↓ -79%
TOTAL	**8,460**	**10,108**	**3,227**	↓ **-16%**

RECORDS BY TYPE TRENDS	2025	2024	2023	CHANGE Y/Y
Breaches Containing PII	12,331,228,218	10,306,192,452	2,704,020,713	↑ +20%
Combo Breaches	9,666,775,886	5,768,928,978	1,430,748,569	↑ +68%
URL Logs	5,905,736,615	896,897,750	-	↑ +558%
TOTAL	**27,903,740,719**	**16,972,019,180**	**4,134,769,282**	↑ **+64%**

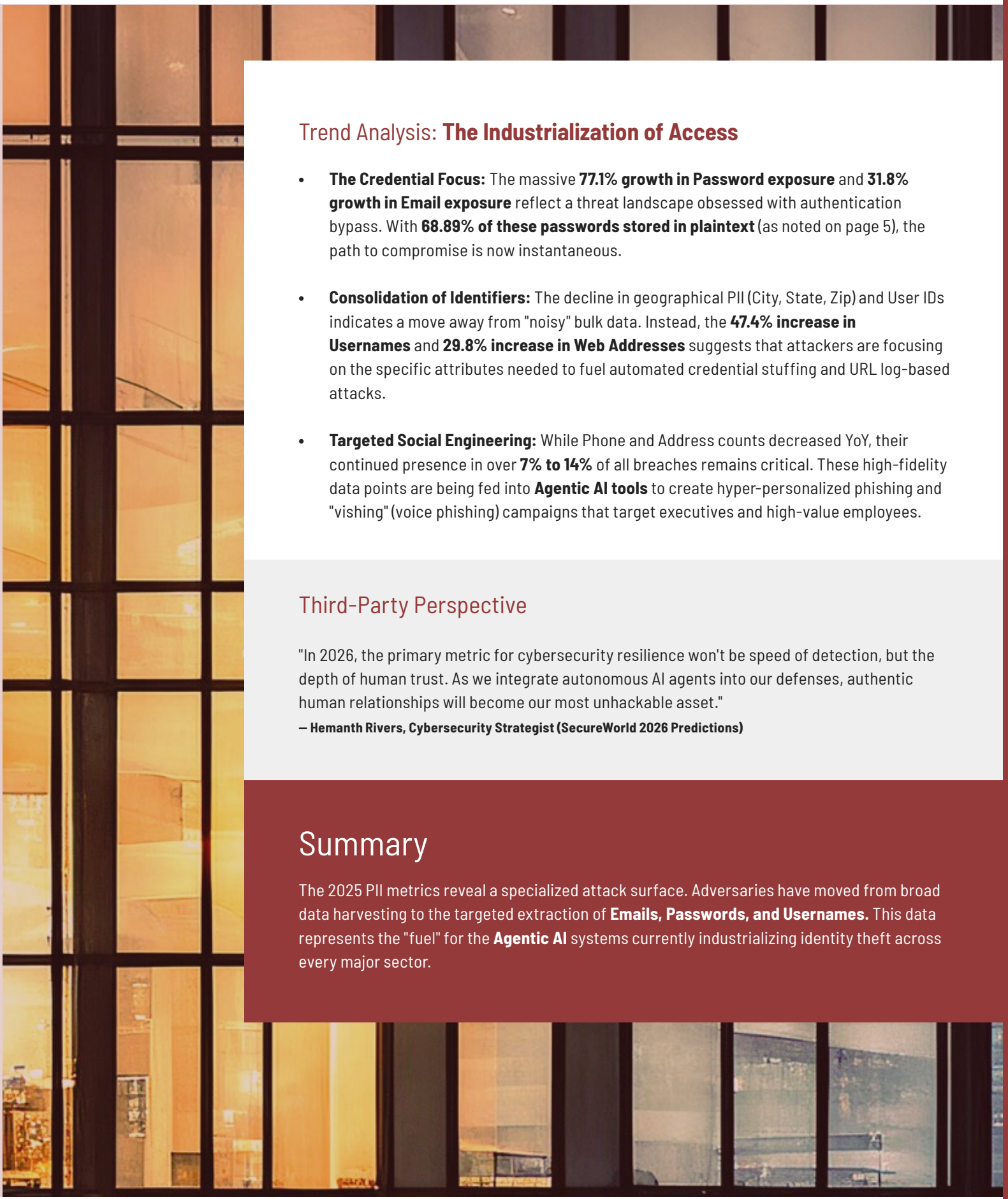
Top PII Exposed from Breaches

The prevalence of specific personally identifiable information (PII) types in 2025 highlights a clear tactical shift among threat actors. By leveraging **Agentic AI** to hunt and filter data, adversaries are prioritizing "high-utility" attributes that enable immediate account takeover and sophisticated impersonation.

Top PII Attributes Exposed

The following table details the prevalence of PII types across all verified breaches and identity leakages in 2025 compared to 2024.

RECORDS BY TYPE	2025	2024	Y/Y CHANGE
Email	71.38%	54.16%	+31.8%
Password	68.87%	38.89%	+77.1%
Username	34.76%	23.58%	+47.4%
Web Address	26.98%	20.79%	+29.8%
Phone	14.57%	35.33%	-58.8%
Full Name	8.98%	16.08%	-44.2%
First Name	7.43%	16.08%	-53.8%
Address	7.16%	13.87%	-48.4%
Last Name	6.79%	14.85%	-54.3%
Birthdate	5.42%	8.86%	-38.8%
City	5.37%	11.30%	-52.5%
Zip Code	5.18%	10.28%	-49.6%
State	3.60%	9.94%	-63.8%
User ID	2.89%	21.16%	-86.3%



Trend Analysis: **The Industrialization of Access**

- **The Credential Focus:** The massive **77.1% growth in Password exposure** and **31.8% growth in Email exposure** reflect a threat landscape obsessed with authentication bypass. With **68.89% of these passwords stored in plaintext** (as noted on page 5), the path to compromise is now instantaneous.
- **Consolidation of Identifiers:** The decline in geographical PII (City, State, Zip) and User IDs indicates a move away from "noisy" bulk data. Instead, the **47.4% increase in Usernames** and **29.8% increase in Web Addresses** suggests that attackers are focusing on the specific attributes needed to fuel automated credential stuffing and URL log-based attacks.
- **Targeted Social Engineering:** While Phone and Address counts decreased YoY, their continued presence in over **7% to 14%** of all breaches remains critical. These high-fidelity data points are being fed into **Agentic AI tools** to create hyper-personalized phishing and "vishing" (voice phishing) campaigns that target executives and high-value employees.

Third-Party Perspective

"In 2026, the primary metric for cybersecurity resilience won't be speed of detection, but the depth of human trust. As we integrate autonomous AI agents into our defenses, authentic human relationships will become our most unhackable asset."

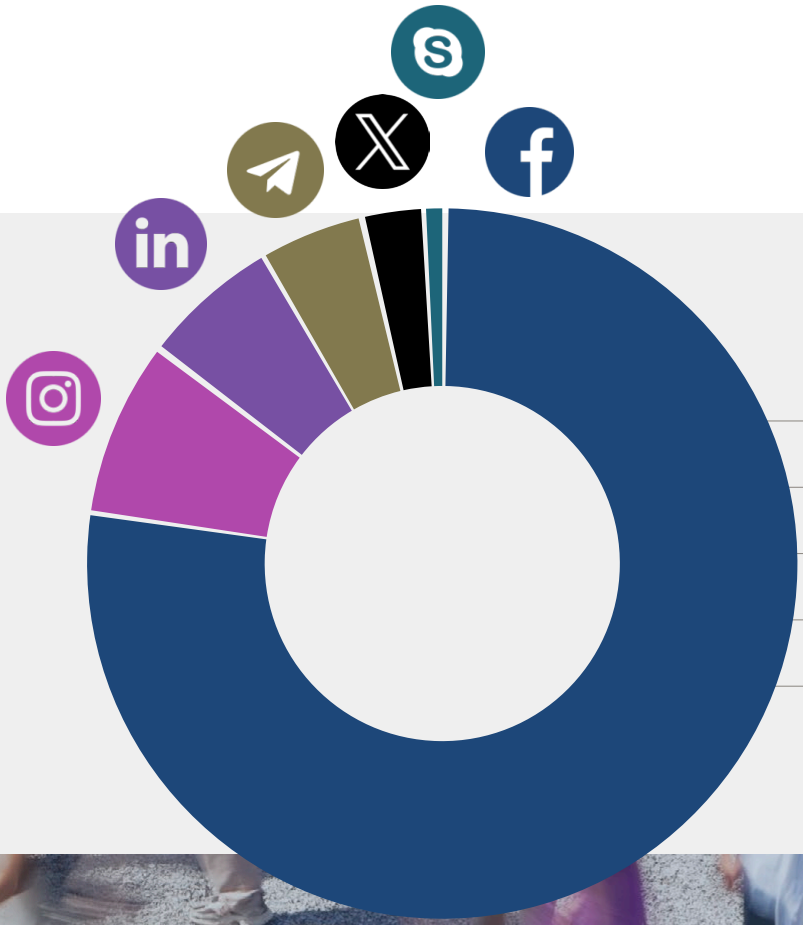
— Hemanth Rivers, Cybersecurity Strategist (SecureWorld 2026 Predictions)

Summary

The 2025 PII metrics reveal a specialized attack surface. Adversaries have moved from broad data harvesting to the targeted extraction of **Emails, Passwords, and Usernames.** This data represents the "fuel" for the **Agentic AI** systems currently industrializing identity theft across every major sector.

Social Media Profile Exposure

Social media identifiers are the connective tissue of modern digital identity. In 2026, these profiles are no longer just social outlets; they serve as the primary reconnaissance source for **Agentic AI** to correlate personal and professional data. By harvesting these identifiers, threat actors can build 360-degree profiles that facilitate hyper-personalized phishing, deepfake impersonation, and multi-channel social engineering attacks. The following chart and table details the volume and prevalence of specific social media profile identifiers identified across all 2025 breaches.



SOCIAL PROFILE	EXPOSED	SHARE
Facebook	89,552,374	77.25%
Instagram	9,224,856	7.96%
LinkedIn	7,330,544	6.32%
Telegram	5,517,446	4.76%
Twitter	3,390,104	2.92%
Skype	850,528	0.73%

Key Recommendations

- **Implement Social Media Shadow Monitoring:** Security teams should monitor for the exposure of executive and high-value employee social profiles to detect impersonation attempts before they reach the internal network.
- **Mandate Multi-Channel Verification:** Establish a strict policy requiring a secondary, out-of-band "liveness challenge" (such as a voice call to a known number) for any urgent request received via social or collaboration channels.
- **Zero-Trust for Social Integrations:** Treat social media logins (OAuth) as a high-risk vector. Limit the use of "Login with Facebook/LinkedIn" for corporate applications to prevent lateral movement following a social profile compromise.

Third-Party Threat Perspective

"In 2026, we are seeing a shift in how attackers target people rather than systems. Criminals now link campaigns to news events and use social media profiles to identify exactly who is connected to those events. With AI, a few public photos and a LinkedIn title are all that's needed to build a convincing deepfake that can swindle millions from an unsuspecting employee.

— Michael Covington, VP of Strategy, Jamf (2026 Outlook)

Summary

Social media exposure is the "fuel" for the 2026 impersonation crisis. With over **115 million total profiles** identified in 2025 breaches, adversaries have an exhaustive library of human behavior and corporate structure. Organizations must move beyond protecting the "official" brand page and begin protecting the **digital footprints** of their employees to disrupt the AI-driven attack lifecycle.

Trend Alert:

The Rise of AI-Native Impersonation

The most significant shift in 2026 is the automation of the "Social-to-BEC" pipeline. Attackers are using **Agentic AI** to scrape the 89.5M Facebook and **7.3M LinkedIn** profiles identified this year to generate "vibe-coded" deepfakes.

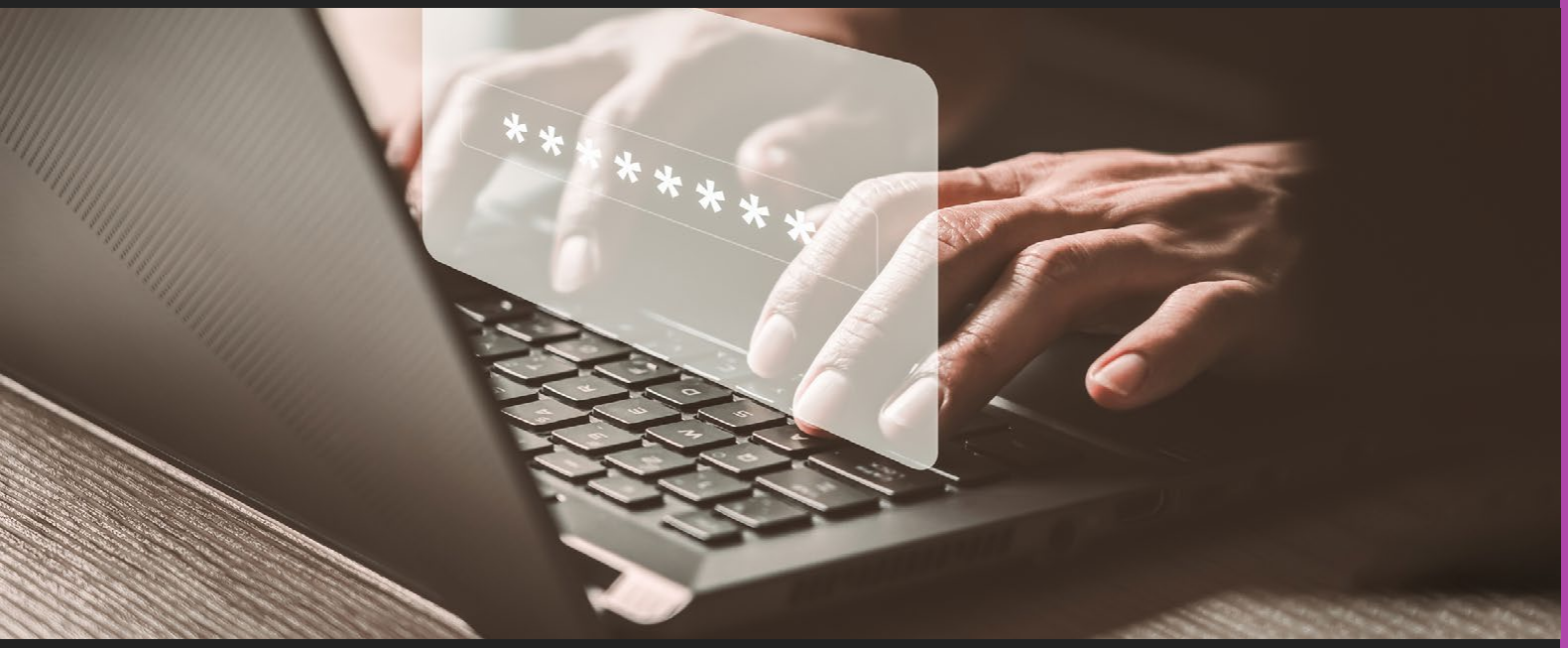
The Facebook "Master Key": Dominating exposure with **77.25%**, Facebook profiles provide the personal PII (pet names, family check-ins, hobbies) that AI agents use to bypass security questions and craft emotionally resonant lures.

LinkedIn as a Corporate Map: LinkedIn profiles remain the highest-value targets for **Business Email Compromise (BEC)**. Threat actors use these 7.3M profiles to map corporate hierarchies and identify "power users" in finance or HR for targeted exploitation.

Telegram as a Data Hub: The exposure of **5.5M Telegram profiles** is particularly concerning, as Telegram has become the primary infrastructure for "Malware-as-a-Service" and real-time credential distribution.

Breach Metadata Information

This section provides a deep-dive metadata analysis of ingested breaches, revealing a fundamental shift in adversarial tradecraft. As threat actors industrialize their operations, they are increasingly successful at stripping away security layers to recover plaintext passwords and bypass weak hashing. This surge in clear-text availability, driven by sophisticated exfiltration and high-velocity cracking, has become the primary engine for automated, high-velocity Account Takeover (ATO) campaigns.



Password Storage Distribution

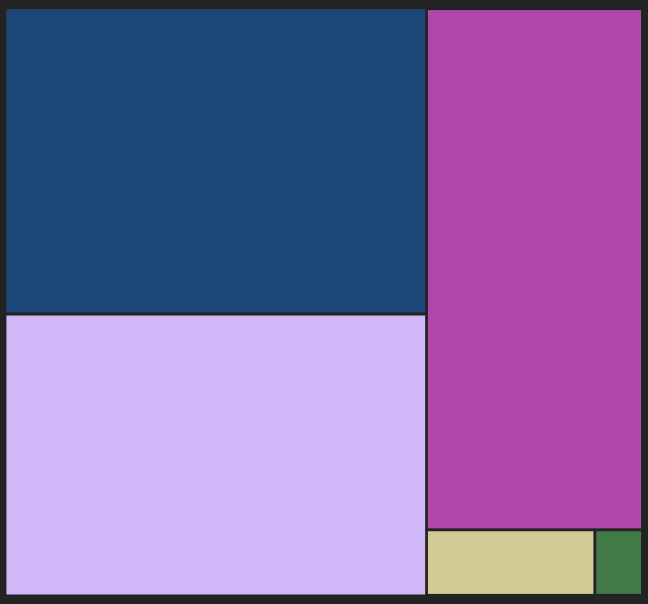
The 2025 data reveals a alarming shift toward unprotected credentials. While organizations prioritize complex policies, the underlying storage methods are failing to provide even basic cryptographic protection.

GENERAL STATISTICS	2025	SHARE	2024	SHARE	CHANGE Y/Y
Plaintext Passwords	15,154,181,157	↑ 68.89%	4,196,264,283	↑ 35.30%	↑ +261%
Hashed Passwords	1,158,193,010	↑ 5.26%	426,048,845	↑ 3.58%	↑ +172%
No Passwords	5,685,629,937	↑ 25.85%	7,265,792,220	↑ 61.12%	↓ -22%

Security Alert: Plaintext passwords now dominate the landscape at **68.89% of all records**. This **261% YoY explosion** indicates that the majority of modern breaches bypass the "cracking" phase entirely, granting adversaries immediate, clear-text access to sensitive accounts.

Password Hashing Algorithms

Even when data is hashed, it is rarely secure. Analysis of **15.7 billion detections** shows that the "protection" used by most breached entities is functionally obsolete.



MD5

4,402,094,546 Detections
27.99% of Total
Risk Level: **Broken**

MySQL323

3,872,958,353 Detections
24.63% of Total
Risk Level: **Broken**

Other Algorithms

**Private Corporate Algorithms*
4,042,232,956 Detections
25.7% of Total
Risk Level: **High**

SHA-1

373,481,054 Detections
2.37% of Total
Risk Level: **Deprecated**

Bcrypt

101,735,032 Detections
0.65% of Total
Risk Level: **Secure**

Security Insight

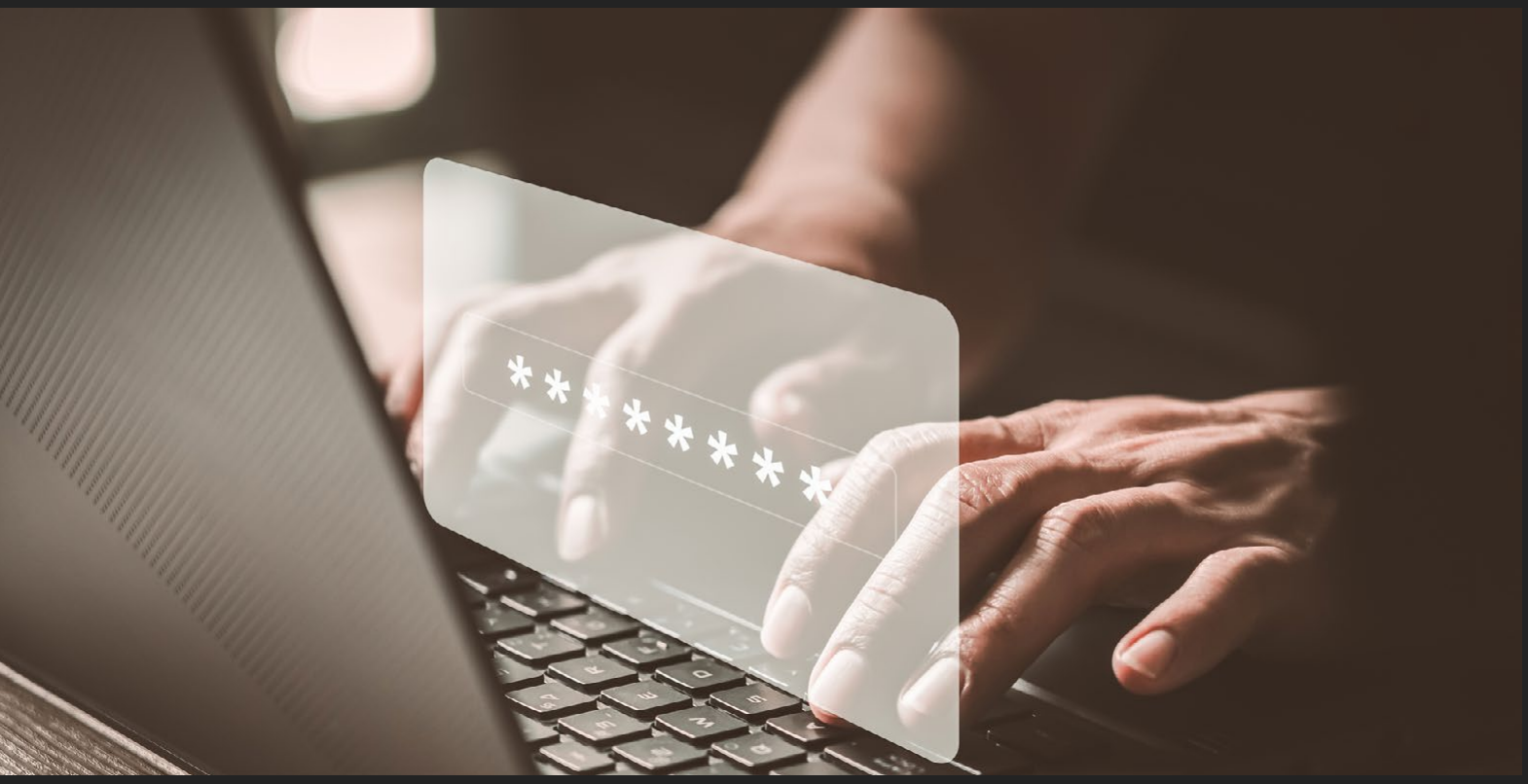
Weak and deprecated algorithms, specifically **MD5** and **MySQL323**, account for over **52% of all detections**. Modern, secure standards like **Bcrypt** protect a negligible **0.65%** of the year's hashed records. This "hashing gap" allows threat actors to reverse passwords in milliseconds using off-the-shelf hardware.

Strategic Recommendations

- **Eliminate Plaintext Storage:** Boardrooms must treat plaintext storage as a primary fiduciary liability. Enforce a zero-tolerance policy for unencrypted credential storage across all internal and vendor-managed databases.
- **Adopt Adaptive Hashing:** Standardize on **Argon2id** (the current gold standard) or **Bcrypt** with a high work factor. These "slow" algorithms are specifically designed to resist the high-speed brute-force attacks enabled by modern GPUs.
- **Implement NIST 800-63B Guidelines:** Move away from arbitrary complexity rules that lead to predictable patterns (e.g., "Password123!"). Instead, prioritize **password length** and check new credentials against **known breached password lists** in real-time.

Summary

Address the Plaintext Pipeline: The 261% surge in plaintext credentials represents a critical escalation in identity risk. In 2026, the challenge isn't simply a lack of protection at the source; rather, it's that threat actors have industrialized the recovery of plaintext passwords. By leveraging high-velocity cracking farms and infostealer exfiltration, adversaries are converting compromised datasets into clear-text weapon libraries, rendering legacy hashing obsolete.



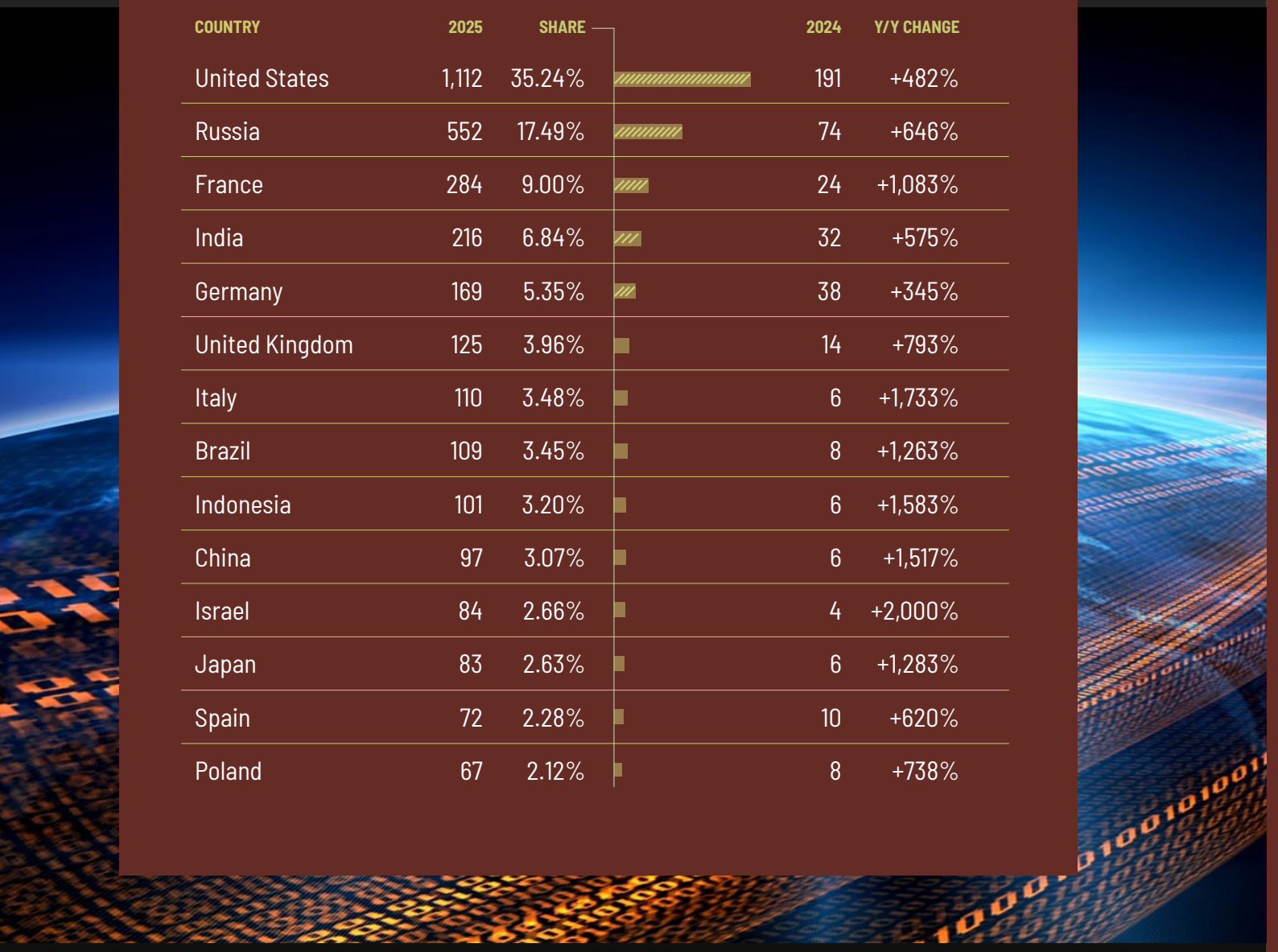
Geographic Distribution of Breaches

In 2025, the identity threat landscape became a mirror for global geopolitical volatility. As adversaries leveraged **Agentic AI** to automate reconnaissance, the volume of breaches surged across every major economy. This geographic expansion reflects a shift from opportunistic localized attacks to industrialized, cross-border campaigns targeting high-GDP nations and critical infrastructure.in Government and Finance sectors indicates a move toward high-impact, geopolitically motivated, and high-value extortion attacks.

Top 15 Countries by Breach Volume

The following metrics represent the verified breaches impacting organizations headquartered in these regions, extracted from Constella's 2025 metadata.

COUNTRY	2025	SHARE	2024	Y/Y CHANGE
United States	1,112	35.24%	191	+482%
Russia	552	17.49%	74	+646%
France	284	9.00%	24	+1,083%
India	216	6.84%	32	+575%
Germany	169	5.35%	38	+345%
United Kingdom	125	3.96%	14	+793%
Italy	110	3.48%	6	+1,733%
Brazil	109	3.45%	8	+1,263%
Indonesia	101	3.20%	6	+1,583%
China	97	3.07%	6	+1,517%
Israel	84	2.66%	4	+2,000%
Japan	83	2.63%	6	+1,283%
Spain	72	2.28%	10	+620%
Poland	67	2.12%	8	+738%



Key Recommendations

- **Localize Threat Intelligence:** Organizations with multi-national footprints must monitor for region-specific breach trends. An identity compromised in an Indonesian breach (+1,583%) can be the entry point for a global corporate network.
- **Harden Supply Chain Nodes:** Focus security audits on vendors in high-growth breach regions like Italy and India, where the rapid industrialization of business has outpaced the implementation of modern identity controls.
- **Prepare for State-Sponsored Impersonation:** In regions of high conflict (e.g., Israel/Russia), be hyper-vigilant against AI-driven impersonation of local government or corporate authorities.

Third-Party Threat Perspective

"Cybersecurity in 2026 is accelerating amid growing threats and geopolitical fragmentation. Disruptions now move swiftly across borders, and 64% of organizations are now accounting for geopolitically motivated cyberattacks—such as the disruption of critical infrastructure or espionage—in their primary risk strategies."

— World Economic Forum, Global Cybersecurity Outlook 2026

Geographic Trends:

The Geopolitics of Identity

The Sovereign Dilemma: The unprecedented **2,000% surge in Israel** and **646% increase in Russia** are direct results of "hacktivism" and state-sponsored cyber warfare. Identity data is no longer just stolen for profit; it is weaponized for espionage and national disruption.

Western Europe's Critical Pivot: France (**+1,083%**) and Italy (**+1,733%**) experienced explosive growth in breaches targeting the services and manufacturing sectors. This aligns with the 2026 trend of adversaries targeting "closed-source" commercial software in Western supply chains.

The U.S. Dominance: The United States continues to lead the world in breach volume. With an average breach cost reaching a record **$10.22 million** in 2025 [Source: IBM], the financial incentives for targeting U.S. organizations remain the primary driver for global cybercrime syndicates.

Summary

The **+482% YoY increase** in the United States and the triple-digit growth across nearly every major economy proves that identity risk has become systemic. In 2026, a breach in one corner of the globe is no longer an isolated incident; it is a signal of a broader, AI-driven campaign targeting the global economy's interconnected supply chain.

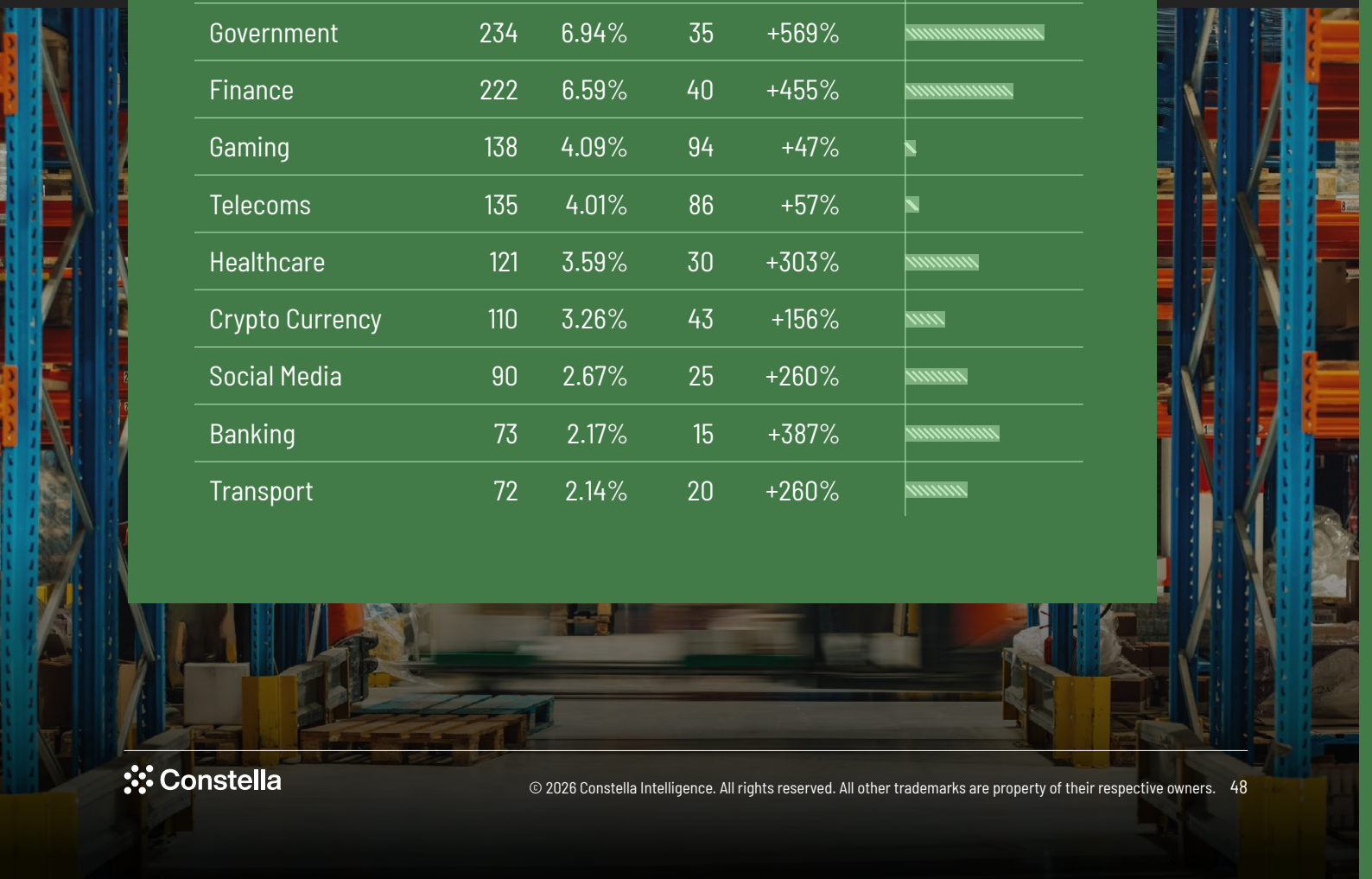
Industry Sector Analysis

The 2025 sector analysis reveals a landscape where adversaries are no longer just "casting a wide net" but are strategically targeting the foundational pillars of the digital economy. While **E-Commerce** and **Services** remain high-volume targets due to the sheer density of consumer data, the dramatic surge in **Government** and **Finance** sectors indicates a move toward high-impact, geopolitically motivated, and high-value extortion attacks.

2025 Breach Distribution by Sector

The following data is derived from breach source analysis, affected domain categorization, and high-fidelity threat intelligence enrichment.

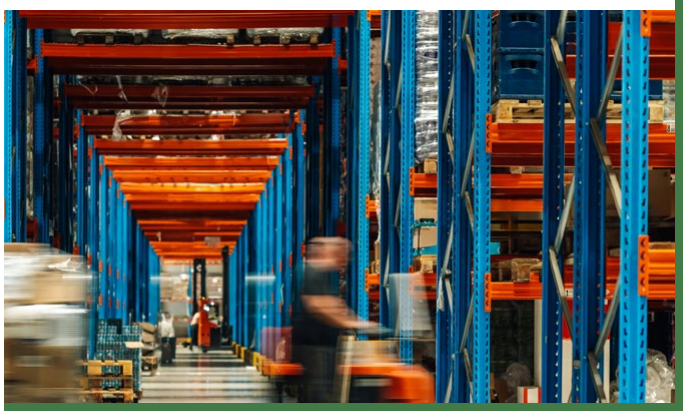
SECTOR	2025	SHARE	2024	Y/Y CHANGE
E-Commerce / Retail	603	17.89%	178	+239%
Services	510	15.13%	151	+238%
Education	342	10.15%	78	+338%
Technology	327	9.70%	87	+276%
Email Providers	309	9.17%	1,509	-80%
Government	234	6.94%	35	+569%
Finance	222	6.59%	40	+455%
Gaming	138	4.09%	94	+47%
Telecoms	135	4.01%	86	+57%
Healthcare	121	3.59%	30	+303%
Crypto Currency	110	3.26%	43	+156%
Social Media	90	2.67%	25	+260%
Banking	73	2.17%	15	+387%
Transport	72	2.14%	20	+260%



Sector Trends:

The Industrialized Attack Surface

- **Government at the Tipping Point:** The **569% increase** in Government breaches reflects a "perfect storm" of state-sponsored espionage and the exploitation of legacy systems. Adversaries are targeting municipal and federal infrastructure to disrupt essential services and harvest PII for future influence operations.
- **The E-Commerce "Gold Mine":** With a **239% increase,** Retail remains the primary target for Infostealers (see page 2). Attackers are leveraging the 51.7 million packages processed this year to harvest customer payment data and session cookies, enabling "Instant Account Takeover" (ATO).
- **Education's Growing Risk:** A **338% surge** in the Education sector highlights the vulnerability of remote learning environments and decentralized student data systems, which often lack the enterprise-grade security of the Finance or Tech sectors.
- **The Email Provider "Anomaly":** The **80% decrease** in Email Provider breaches does not signal a drop in risk, but a shift in tactics. Attackers have moved away from compromising providers directly and are instead using **Agentic AI** to bypass MFA on individual accounts through session hijacking.



Strategic Recommendations:

Cross-Sector Identity Monitoring: Organizations in the Services and Tech sectors must monitor for employee exposures in E-Commerce breaches, as credential reuse is the most common bridge for lateral movement into corporate environments.

Zero-Trust for Government Entities: Given the 569% surge, government agencies must accelerate the transition to phishing-resistant MFA (FIDO2) and mandate strict third-party risk assessments for all contractors.

Harden Retail API Endpoints: With the rise of automated ATO, Retailers should deploy AI-driven behavioral analysis to detect non-human login patterns that signal bot-driven credential stuffing.

Third-Party Threat Perspective

"In 2026, the foundational infrastructure we use to innovate—like cloud, AI, and OT/IT convergence—is rapidly expanding the attack surface. More attackers will simply 'walk in' using valid credentials and trusted AI agents to blend into normal activity, aiming to disrupt real-world services in the Government and Finance sectors."

— Nextgov/FCW, 2026 Cyber Expert Forecast

Summary

The **+569% spike** in Government breaches and triple-digit growth across **Finance, Banking, and Healthcare** confirms that no pillar of society is safe from industrialized identity theft. As the attack surface continues to expand, organizations must recognize that their risk is inherently linked to the security posture of the broader ecosystem.

Data Composition:

Strategic Pivot to High-Fidelity Intelligence

In 2026, the volume of raw data available on the dark web has reached a saturation point. As adversaries use **Agentic Alto** flood the market with recycled credential sets, the primary challenge for security operations has shifted from data collection to **intelligence synthesis**. This section breaks down the composition of ingested breach data, illustrating our transition toward high-value, verified breaches containing PII sources over lower-fidelity bulk compilations.

Breaches by Type

Classification of breaches by source type. Breaches with PII contain personally identifiable information from direct compromises while Combo Breaches aggregate credential pairs from multiple sources. The **+661% surge** in breaches

BREACHES COMPOSITION	2025	SHARE	2024	CHANGE Y/Y
Breaches / Leakages	5,712	↑ 67.52%	750	↑ +661%
Combo Breaches	2,055	↑ 24.29%	6,054	↓ -66%
URL Logs	693	↑ 8.19%	3,304	↓ -79%

Breaches with PII Breakdown

Breakdown of breaches with PII by data richness. HL with PII contain multiple personally identifiable attributes, while breaches with only eamil contain only a list of emails without passwords nor additional PII.

BREACHES COMPOSITION	2025	SHARE	2024	CHANGE Y/Y
Breaches with PII	5,406	↑ 94.64%	745	↑ +626%
Breaches with only email	233	↑ 4.08%	5	+4,560%

Records Composition: **The Identity Correlation Gap**

The presence of a verified email address is the "anchor" of identity risk. Records containing emails allow for immediate cross-referencing against corporate domains and are the primary fuel for **Business Email Compromise (BEC).**

Trend Analysis: **The 2026 Identity Density Metric**

- **The Unique Identity Ratio:** While the volume of total records increased by **135%**, unique emails grew by only **11%**. This indicates a massive increase in **data density per identity.** Attackers now have multiple passwords, physical addresses, and social identifiers for the same individual, enabling the "composite identity" attacks described on page 34.
- **The BEC Weaponization:** In 2025, 1 in 6 breaches involved AI-driven orchestration [Source: IBM 2025]. With **17.2 billion email-linked records** now available, AI agents can map a target's communication style and "vibe" with unprecedented accuracy.
- **Non-Human Identity Risks:** The growth in records without emails (**+95%**) often points to machine-to-machine (M2M) credentials, API keys, and session tokens—vectors that traditional identity monitoring often ignores but which account for the costliest breaches in the 2026 landscape.

RECORDS COMPOSITION	2025	SHARE	2024	CHANGE Y/Y
Records with Email	17,254,138,405	↑ 61.84%	6,438,442,813	↑ +168%
Records without Email	10,649,602,314	↑ 38.16%	5,449,662,535	↓ +95%
Unique Emails Ingested	4,726,061,704	↑ 27.39%	4,273,033,947	↓ +11%

Summary: The data composition of 2025 marks the end of the "Bulk Data" era. With **94.6% of HL breaches containing deep PII**, the risk is no longer just a stolen password—it is a stolen persona. Stricter deduplication has allowed us to increase ingested records by **168%** while maintaining a lean, actionable intelligence feed that prioritizes the most dangerous exposures.

General Statistics:

The Infostealer Pandemic

Infostealer malware has solidified its position as the primary engine of the global threat economy. In 2025, these lightweight, high-impact programs evolved from simple credential harvesters into sophisticated data-exfiltration hubs, feeding **Agentic AI** systems the raw session tokens and cookies required to bypass modern MFA.

Constella's **Net New deduplication pipelin**e ensures that we filter out the noise of recycled logs, providing a 100% unique view of the 2025-2026 infostealer landscape.

Infostealer Packages Inserted
51.7 Million

Infostealer Records Inserted
2.3 Billion

Infected Devices (2025 Date)
24.8 Million

Passwords Extracted
2.3 Billion
98.6% of Records

Web Addresses (URLs) Extracted
2.3 Billion
99.54% of Records

Email Addresses Extracted
1.1 Billion
48.56% of Records

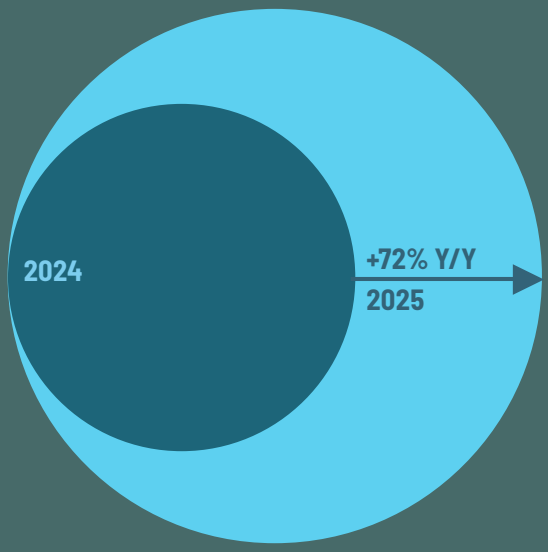
Usernames Extracted
1.1 Billion
48.80% of Records

Hardware IDs (HWIDs) Extracted
434.3 Million
18.51% of Records

Key Insight:

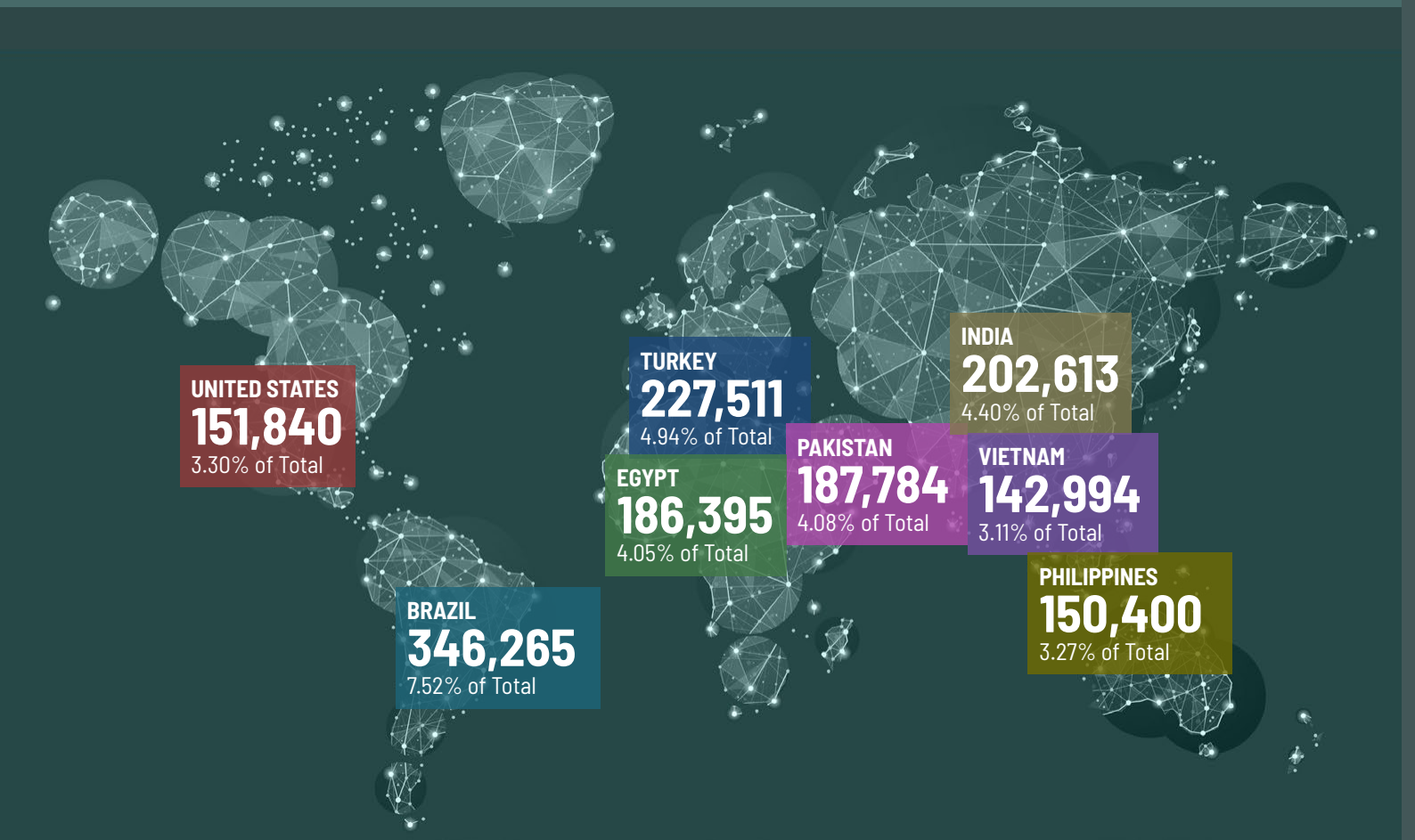
In 2025, over **51.7 million** packages were processed—a **72% increase,** YoY. Most alarming is that **98.6%** of these logs contained active passwords, and nearly **100%** included the specific URLs where those credentials were used, providing a direct roadmap for automated attacks.

51.7 Million Packages Processed



Geographic Distribution of Infections

Infostealer activity is highly concentrated in regions with high digital growth and evolving cybersecurity regulations. Geographic attribution is derived from IP geolocation and system locale data.



Trend Alert:

Brazil, Turkey, and India remain the "epicenters" of infostealer distribution. The high volume in these regions is driven primarily by the proliferation of cracked software, pirated media, and sophisticated local "social engineering" rings that tailor malware delivery to regional events.

Comparative Analysis: 2025 vs. 2024

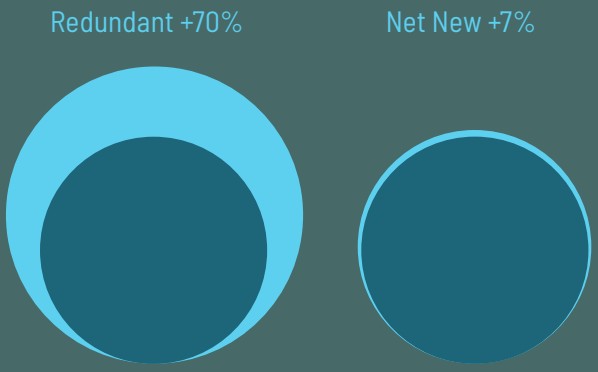
The divergence between processed packages and inserted records highlights the critical importance of deduplication in maintaining threat intelligence integrity.

CORE METRICS	2025	2024	CHANGE Y/Y
Packages Processed	51,731,126	30,000,000	↑ +72%
Unique Devices Identified	24,804,145	14,000,000	↑ +77%
Deduplicated Records	2,345,568,361	2,198,546,806	↑ +7%

Critical Analysis

While the volume of packages and infected devices exploded by over 70%, the number of net-new records inserted into the Data Lake grew by only 7%. This disparity proves that threat actors are repeatedly compromising the same high-value targets.

Without Constella's NetNew pipeline, security teams would be overwhelmed by a 70% increase in redundant, "noisy" alerts.



Expert Recommendation

- **Session Token Invalidation:** Infostealers don't just steal passwords; they steal active sessions. Organizations must implement "Session Zero-Trust," where a change in IP or device hardware ID (HWID) immediately triggers a re-authentication challenge.
- **Audit HWID Extraction:** With 434 million HWIDs now in the hands of attackers, "Device Trust" based on static hardware identifiers is no longer sufficient. Move toward dynamic, certificate-based device identification.

Summary

The **77% surge in unique infected devices** signifies that the perimeter has moved to the employee's personal machine. In 2026, infostealer logs are the "unstructured fuel" that enables **Agentic AI** to simulate human logins with terrifying accuracy.

Appendix and Data Sources

Data Sources

Constella Intelligence continuously monitors the Tactics, Techniques, and Procedures (TTPs) of global threat actors to identify the breaches and leakages that occurred between January and December 2025. In addition to breaches formally reported in the media, Constella utilizes proprietary technology to detect information within transient data dumps and industrial-scale harvests across the surface, deep, and dark web.

Our automated crawlers and subject matter experts leverage a diverse array of sources to capture and verify identity data, including:

- **Industrialized Infostealer Logs:** 51.7 million packages processed from infected devices worldwide.
- **Underground Communities and Forums:** High-activity nodes where breaches containing PII data is traded.
- **Black Markets and Dark Web Repositories:** Monitoring the sale of fresh PII and validated credential lists.
- **Agentic AI-Driven Discovery:** Using autonomous agents to hunt and identify transient leaks that evade traditional indexing.

Constella analyzes, verifies, cleans, and attributes this data to determine the severity of risk facing consumers and enterprises. We then alert impacted parties to mitigate risk based on factors such as attribute sensitivity, data authenticity, the number of individuals impacted, and the recency of the exposure.

Data Verification / Methodology

The integrity of the 2026 Identity Breach Report is maintained through a rigorous, multi-stage verification process designed to handle the 135% year-over-year increase in curated records.

1. **Collection & Deduplication:** After raw data collection, Constella employs machine learning algorithms and daily credential compilations to remove "zombie" data and duplicate records. This strategic "Combo Consolidation" ensures that we focus on net-new unique records, resulting in the -66% decrease in Combo Breaches observed this year.
2. **AI-Enhanced Validation:** Our team utilizes **Agentic AI automation** to hunt 159% more breaches than in previous years. These agents perform preliminary triage, flagging sensitive information and validating domain authenticity at machine speed.
3. **Expert Attribution:** Subject matter experts use investigative methods to ensure that domains and breach details are valid. Each breach is attributed to a specific source and normalized for integration into the **54.6B+ record Data Lake.**
4. **Risk Scoring:** Once verified, the Constella platform calculates a risk score for each identity exposure based on variables such as password strength (with a focus on the 68.89% of credentials currently in plaintext) and the richness of associated PII attributes.

By applying this methodology, Constella converts raw underground data into the high-fidelity Identity Risk Intelligence required for proactive, boardroom-level defense.

About Constella Intelligence

Constella is a global leader in Identity Risk Intelligence. Powered by the world's largest breach and infostealer data lake, spanning over one trillion attributes across 125+ countries and 50+ languages, Constella empowers organizations to detect, investigate, and respond to threats linked to exposed personal data. Enterprises, managed service providers, and law enforcement agencies worldwide rely on Constella to strengthen identity posture, fuel threat intelligence, and defeat digital risk.

Constella

Where Data & Identity Intersect

Powered by

THE WORLD'S LARGEST VERIFIED IDENTITY DATA LAKE

