# AGS-INTEL: Authentic and Granular Source for Data Breach Intelligence

**Anil Parthasarathi, Sean Cho and Shreyas Kumar**
Texas A&M University, College Station, TX, USA

anilparthasarathi@tamu.edu
donghatamu@tamu.edu
shreyas.kumar@tamu.edu

**Abstract:** As artificial intelligence reshapes the cybersecurity landscape, the demand for a trustworthy, real-time intelligence platform to track security incidents has become mission-critical. This paper proposes AGS-INTEL, an AI-driven platform designed to revolutionize data breach intelligence by providing a credible, real-time repository that consolidates, verifies, and contextualizes global security incidents. Unlike traditional databases, AGS-INTEL employs a validated scoring algorithm and enriched metadata to capture breach dimensions (legal, technical, sectoral, geopolitical), drawing from GDPR/HIPAA disclosures, threat intelligence, dark web forums, and academic reports, among other sources. Utilizing NLP and agentic AI, it extracts structured metadata from unstructured narratives while integrating ethical data scraping, regulatory compliance, and cross-jurisdictional filtering to ensure high fidelity. A visual analytics dashboard empowers stakeholders, including regulators, policymakers, cybersecurity professionals, and journalists, to analyze breach trends by industry, geography, and threat modality, enhancing transparency and risk governance. By delivering authenticated, actionable data, AGS-INTEL addresses critical gaps in existing tools, setting a new standard for ethical AI in breach intelligence and strengthening societal resilience against escalating cyber threats.

**Keywords**: Data breaches, Agentic AI, Cybersecurity, Threat intelligence, Web scraping, Ethical AI

## 1. Introduction

The exponential rise in data breaches has made the accurate tracking and analysis of these incidents a global imperative. From multinational corporations to critical infrastructure providers, no entity is immune to the consequences of data loss, credential theft, and system compromise. Kumar et al. highlight proactive defenses against cyber threats as vital to effective governance in the current age (Kumar 2025b). Yet despite the abundance of breach reporting platforms, the world lacks a single, universally trusted, and methodologically rigorous source of breach data. Current databases often suffer from limited verification, incomplete metadata, geographic bias, or inadequate context for regulatory and research use. This paper presents AGS-INTEL (the Authentic & Granular Source for Data Breach Intelligence), a framework designed to serve as an authentic and comprehensive source of data breach intelligence that leverages the power of LLMs. Devised to fill critical gaps in data breach reporting, AGS-INTEL consolidates breach events across industries, continents, and attack vectors with emphasis on authenticity, completeness, and usability. Throughout this paper, we outline the motivation behind the design of AGS-INTEL, addressing the limitations of existing platforms and highlighting the growing need for reliable breach reporting to inform policy, threat intelligence, insurance modeling, and academic research. By adopting a structured approach grounded in information forensics, machine learning, and legal data mining, the system is positioned to function as a globally authoritative breach intelligence repository. The remainder of the paper describes the system's architecture, data sources, verification methodology, and implications for cybersecurity governance and resilience.

## 2. Related Work

The landscape of data breach repositories has evolved significantly amid the rising frequency and impact of cybersecurity incidents. As new threats continue to emerge, assessing risk has become increasingly challenging, hindering initiatives such as the growth of cyber insurance (Kumar 2025a). In this context, breach repositories have become more critical than ever toward making sense of the current paradigm. This section reviews existing breach repositories, highlighting their methodologies, strengths, and limitations, as well as academic efforts that can help to improve breach intelligence. These insights inform the design of AGS-INTEL, which is structured to address critical gaps in authenticity, granularity, and global coverage.

### 2.1 Currently Existing Repositories

Several widely used platforms currently provide data breach intelligence, each with distinct approaches to data collection, verification, and user interaction. Below, we analyze a sample of key repositories to understand what makes them stand out.

*Privacy Rights Clearinghouse (PRC)*

The Privacy Rights Clearinghouse (PRC), established in 1992, aggregates breach notifications from 15 U.S. government agencies, including the U.S. Department of Health and Human Services and state Attorneys General. PRC employs AI to normalize scraped text and extract contextual metadata, such as organization type and breach details, with multiple automated validation checks to ensure accuracy. Its strengths include broad U.S. public breach coverage and user-backed accuracy validation. However, it is limited to publicly reported U.S. incidents, missing non-disclosed or international ones, with occasional duplicates persisting despite normalization. While PRC covers the major states that comprise most reported data breaches, it does not take into account every state, leaving potential for some breaches to fall through the cracks.

"Have I Been Pwned" (HIBP)

Launched in 2013, "Have I Been Pwned" (HIBP) aggregates data from public breach dumps and security agency feeds, such as the FBI, offering k-anonymity-based password and email checks. Its API supports integrations with tools like Firefox and 1Password, and its "Dump Monitor" bot detects new leaks in near real-time. HIBP's strengths lie in its broad sourcing and privacy-focused querying, but it is constrained by its reliance on public data, which excludes private breaches, and the potential for unverified dumps to include inaccurate information.

*XposedOrNot*

XposedOrNot, established in 2017, is a community-driven, open-source platform that aggregates public breach data and provides domain-level breach analysis. It offers a free API, risk scores for compromised emails, and industry-based breach classifications. Furthermore, it also has mechanisms in place to manually rate the credibility of breach reports. It boasts features to help ordinary people stay on top of breaches such as an alert system capable of notifying users of domain-specific breaches and a function where users can input an email to see its risk rating depending on how compromised it is. However, it lacks advanced predictive analytics, and its manual processing/verification slows operations and risks errors.

*Leak-Lookup*

Since 2016, Leak-Lookup has provided a user-friendly interface for searching against billions of breached records, with a pay-per-lookup model and a limited free API. It focuses on real-time breach updates but lacks transparency about its data sources. Furthermore, premium features also restrict accessibility for non-paying users.

*Identity Theft Resource Center (ITRC)*

The ITRC, founded in 1999, compiles credible data from "government agencies, news media outlets, company news releases, filings, and publications, industry and trade websites, and direct notices provided by consumers or companies" (ITRC n.d.). It offers an interactive dashboard and victim support services but lacks a public API. Its Breach Alert platform provides metadata like business sector, affected individuals, data types, and attack vector. Updates occur weekdays (or weekends for significant incidents). As a whole, ITRC emphasizes victim support and education over technical data aggregation. The ITRC is considered a credible source, but its lack of an API and gaps in coverage limit its effectiveness on a larger scale.

*Verizon Data Breach Investigations Report (DBIR)*

The DBIR, published annually since 2008 by Verizon Business, analyzes thousands of incidents (e.g., over 22,000 in 2025) from global contributors. Verizon's sources include international law enforcement agencies, forensic firms, law firms, cyber insurance agencies, cybersecurity industry sharing groups, and their own Verizon Threat Research Advisory Center. DBIR offers detailed attack vector and industry trend insights but lacks real-time updates and a public API. Its confirmed-breach focus and annual cycles limit timeliness and granularity. However, its worldwide focus, high credibility, and unique sourcing methods make it a vital resource for breach intelligence despite the limitations.

*SpyCloud*

SpyCloud specializes in account takeover prevention, aggregating breach data from the dark web and other closed sources in the criminal underground. Their team analyzes more than 25 billion data breach assets per month. SpyCloud niches in retrieving data from non-public, unsavory outlets (vs. official government sources), though credibility is a concern. One other factor that holds SpyCloud back is that it is a paid service, which limits its reach.

*DLA Piper GDPR Fines and Data Breach Survey*

DLA Piper's annual survey, started in 2019, aggregates GDPR breach notifications from 31 European jurisdictions. Its NOTIFY tool aids client compliance but lacks a public API. This survey appears to be a valuable resource due to its European focused intelligence, but it lacks important metadata for the breaches it reports, which limits applicability (Neto 2021).

*ENISA Threat Landscape Reports*

The European Union Agency for Cybersecurity (ENISA) publishes annual Threat Landscape Reports, aggregating breach and incident data across Europe. While not a searchable database, its European focus, similar to DLA Piper, could also provide a valuable alternative perspective compared to America focused breach repositories. However, it also comes up short in its non-real-time reporting which limits timeliness.

*Trend Micro Zero Day Initiative (ZDI)*

Established in 2005, ZDI focuses on vulnerability intelligence rather than breaches, purchasing and verifying zero-day vulnerabilities before public disclosure. It is currently considered "the world's largest vendor-agnostic bug bounty program" (Trend Micro n.d.). The program sources vulnerability intelligence and then in-house researchers will rigorously verify the credibility before reporting it to the affected company. Public reports will generally only be made once a vulnerability is patched, although in the meantime, Trend Micro will use the intelligence to update the protection filters on its products. While not a breach database, ZDI data could enhance breach repositories by helping to link vulnerabilities to incidents. This could allow for deeper, more informative, breach analysis.

*MITRE CVE Feeds*

The MITRE Common Vulnerabilities and Exposures (CVE) database, established in 1999, provides standardized identifiers and details for publicly disclosed vulnerabilities, including CVSS scores. Widely used across the cybersecurity industry, it supports vulnerability tracking and risk assessment. While not a breach repository, its integration with breach data enables linking incidents to exploited vulnerabilities, unlike most breach-focused platforms. Alongside ZDI, integrating MITRE vulnerability details could greatly improve the analysis of breach data.

### 2.2 Relevant Academic Work

Recent academic studies highlight challenges that AGS-INTEL is designed to address and novel techniques relevant to its implementation. Pursuing a similar mission to AGS-INTEL, Neto et al. discussed the development of a comprehensive global breach database. Their framework sourced breach details from government entities, security research groups, research entities, and media reports. The group constrained their sourcing to only these types of public sources to ascertain credibility. The result was a database engineered to chronicle incidents that took place between 2018 and 2019. Key difficulties faced included lack of standardization in reporting and the difficulty in sourcing intelligence from outside of the US and Europe (Neto 2021). Issues such as obstacles in determining credibility and lack of standardized reporting could be mitigated by recent advances in artificial intelligence. Ayuso et al. have discussed the emerging potential for AI integration into web scraping to greatly improve performance in data collection. Bots can be created to collect links from the web to establish sources, extract data from these sources, and then save that data to a database (Ayuso 2024). With AI, this process can be accomplished much more effectively by immediately adapting to the unique structure of dynamic websites. No matter the way the information is formatted, AI can effortlessly parse through it and return all the desired info in a standardized manner. This technique can be applied towards sourcing information concerning data breaches from a wide variety of reporting outlets. Meanwhile, Noor et. al have introduced a novel machine learning framework that takes the data presented in threat intelligence repositories to semantically analyze the root cause of the incident and predict future occurrences. This system is capable of pinpointing a threat incident classification with high reliability (Noor 2019). Such approaches strengthen breach databases by providing more relevant insights and can be leveraged to assess the credibility of cyber threat intelligence.

### 2.3 Gaps and Opportunities

Existing repositories excel in areas like PRC's regulatory depth, HIBP's accessibility, and SpyCloud's dark web insights, but face limitations: incomprehensive public data reliance, inconsistent coverage, lacking credibility appraisal, manual faults, and limited real-time/predictive capabilities. In the modern age, agentic AI enables novel advancements in metadata extraction and credibility scoring to boost the range and reliability of breach

databases. AGS-INTEL consolidates the strongest aspects of existing solutions, creating a comprehensive source of intelligence on known breaches. Furthermore, our solution integrates modern LLM-based enhancements designed to advance the worldwide understanding and analysis of breach intelligence.

## 3. Proposed Solution

Our proposed solution comprises a large-scale, comprehensive database that charts all known data breaches across the world with meaningful metadata and analysis. This design integrates data from all major pre-existing repositories to minimize gaps in coverage. The framework encompasses data from all major states, European jurisdictions under the GDPR, threat intelligence feeds, company disclosures, dark web forums, user submissions, and academic reports. A key issue that currently plagues data breach repositories is the non-standard reporting across the board. Such a monumental obstacle is overcome through the use of agentic AI powered data collection and parsing. Regardless of reporting format, LLM agents are designed to extract useful information and reformat it into standardized database entries.

To address credibility concerns that plague current repositories, AGS-INTEL incorporates AI-driven trust ratings. Each reported breach is assigned a credibility rating, and entries below a defined threshold are excluded from publication. The framework includes a transparent citation trail for each breach that reveals where the details were sourced from. These measures help ensure that large volumes of data can be ingested while maintaining authenticity and reliability.

To further strengthen the depth of breach analysis, AGS-INTEL integrates intelligence on known vulnerabilities and links them to breaches. By pulling from sources such as ZDI and MITRE CVE, the system accumulates vulnerability intelligence and then employs an algorithm to attach vulnerabilities to breaches where relevant matches exist. This information supports a deeper understanding of breaches and how they came about.

To enhance the user experience, the system incorporates several widely adopted features that other data breach services offer. This includes a clean, easy to use, front-end user interface with visual analytics to boost comprehension. The framework also provides a credential-based query service enabling users to verify if their data has been compromised. A notification system is integrated to help users stay up-to-date on breaches that may be relevant to them. Finally, AGS-INTEL is envisioned as a publicly accessible, free-to-use platform that promotes open data safety for all.

## 4. System Architecture

The architecture of AGS-INTEL encompasses a modular, scalable system engineered to collect, process, verify, and present global data breach intelligence in real time. Comprising four core layers: Data Ingestion, Verification and Normalization, Core Database, and User Interface, this framework ensures comprehensive coverage, high trustworthiness, and actionable insights. Each layer leverages modern artificial intelligence and data science techniques to address the limitations of existing breach repositories. The system is designed to handle the complexity of diverse data sources while maintaining regulatory compliance, ethical standards, and user accessibility.

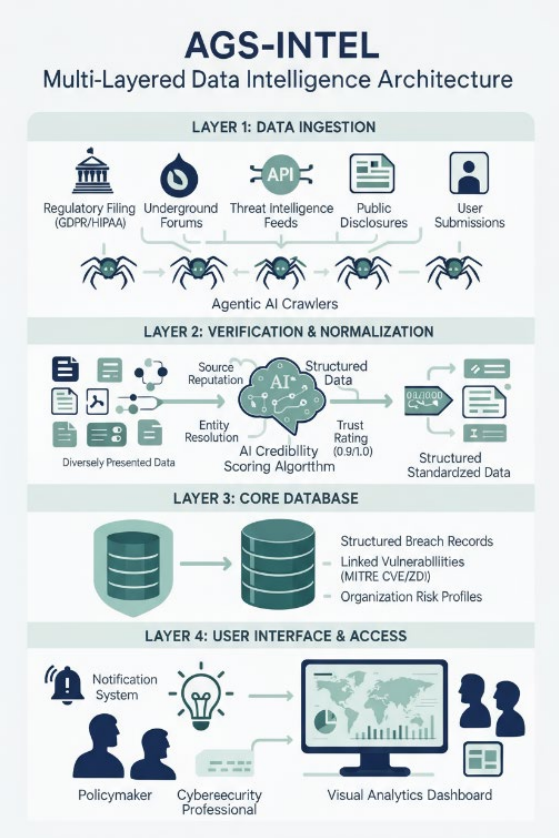


**Figure 1: AGS-INTEL System Architecture Flow**

### 4.1 Data Ingestion Layer

The Data Ingestion Layer is the foundational pillar of AGS-INTEL, designed to aggregate breach-related data from an exhaustive array of sources, including all major pre-existing repositories, regulatory bodies, and emerging channels. Building on insights from Related Work regarding global sourcing challenges, this layer addresses gaps like geographic biases and incomplete coverage by collecting raw data from diverse sources. The layer employs AI-driven ingestion pipelines, inspired by Ayuso et al. (2024), to efficiently capture high-volume, heterogeneous data.

AGS-INTEL ingests data from eight distinct source categories: (1) existing breach repositories (e.g. PRC, HIBP, etc.), (2) U.S. state and federal breach notification registries, (3) international regulatory filings (4) company press releases and investor disclosures (5) threat intelligence feeds and public breach dumps, (6) dark web and underground forums, (7) vulnerability intelligence reports (ZDI), and (8) user submissions. This combination enables comprehensive coverage by ensuring both depth through verified incidents and breadth through early-stage leaks and global occurrences.

The ingestion layer employs a hybrid pipeline to handle diverse data formats and high ingestion velocities, focusing on raw data collection without standardization (reserved for the next layer). Methods are optimized for adaptability, leveraging agentic AI systems capable of reasoning, planning multi-step actions, and adapting dynamically. These agents operate with minimal human intervention while maintaining oversight through automated audit logging and periodic human review. This design enables efficient coverage of all data source categories by orchestrating tools, handling anomalies, and ensuring resilience. Below, each method is detailed with a breakdown of its mechanisms.

- **AI-Powered Web Crawlers**: These autonomous crawlers, built on agentic AI models, scrape unstructured or semi-structured content from websites. Agents start with seed URLs, explore linked pages, and use NLP to identify breach-relevant sections. For dynamic sites, agents simulate user interactions (e.g., scrolling or form submissions) while respecting ethical boundaries. Agentic operations include decision-making on crawl depth and adaptive retrying if sites change layouts,

addressing the challenges in existing solutions regarding non-standard formats. Crawls run on schedules with fault tolerance via distributed agents.

- **API Integrations**: Direct, structured integrations pull data via RESTful or GraphQL APIs. This ensures low-latency, high-volume ingestion without web overhead.
- **Real-Time Monitoring and Feeds**: Agentic AI agents subscribe to streaming feeds and monitor channels using semantic search and anomaly detection, processing incoming data in real time. Agents filter noise and aggregate from multiple streams. For dark web sources, agents use Tor-enabled passive listeners to detect leaks without engagement (Xu 2021). Agentic features include autonomous escalation and integration with keyword-based filters, ensuring timeliness that exceeds current solutions.
- **User Submission Portal**: A secure, web-based interface allows manual uploads, with AI validating and enriching submissions in real time. Agents scan uploaded files for relevance using NLP, cross-reference against known incidents, and flag for human review if needed. This pipeline extends coverage to non-public or emerging breaches.

### 4.2 Verification and Normalization Layer

Given the heterogeneity and varying reliability of breach data sources, the Verification and Normalization Layer plays a crucial role in transforming raw ingested data into structured, trustworthy data for analytical and predictive tasks. This layer ensures the consistency, completeness, and authenticity of each breach record before it enters the core AGS-INTEL database.

- **Data Normalization**: Raw breach records often vary widely in structure and granularity, and observed datasets reflect detection/disclosure biases and uneven reporting across incident types and jurisdictions (Romanosky 2016). To enable downstream comparison and analysis, all records are normalized to a standardized schema that includes breach date, organization, industry sector, data types exposed (e.g., email, SSN), breach vector, affected geography, and source. LLMs adaptively handle variations in format, such as converting GDPR notifications from DLA Piper sources or U.S. state registries into standardized entries, ensuring global comparability.
- **Deduplication and Entity Resolution**: Multiple sources often report the same breach with slight variations. To prevent inflating statistics and introducing inconsistencies, AGS-INTEL applies probabilistic matching algorithms (e.g., based on Jaccard similarity for entity names, timelines, and impact metrics) to identify and merge overlapping reports. The system automatically updates existing records when secondary sources contribute additional details, preserving data integrity and avoiding duplication.
- **Credibility Scoring**: To mitigate reliability gaps in sources like dark web forums or unverified dumps, each breach receives an AI-generated trust rating on a scale of 0-1, based on factors including source reputation, cross-corroboration across multiple sources, metadata completeness, and ML confidence in extraction. Entries scoring below a 0.5 threshold are excluded from the core database but logged for potential future validation; included entries display the rating transparently, allowing users to gauge reliability.
- **Breach Classification and Enrichment**: Using a multi-label classification model fine-tuned on historical breach data, breaches are classified by type (e.g., ransomware, phishing) and mapped to the MITRE ATT&CK framework for threat actor tactics. Additionally, records are assigned CVSS scores for severity and linked to relevant vulnerabilities from integrated sources like ZDI and MITRE CVE feeds based on similarity matching. An algorithmic matching process correlates breach details (e.g., exploit patterns) with vulnerability intelligence to infer root causes and present them to users, enhancing analytical depth beyond current repositories.
- **Transparency and Oversight**: Each record maintains a transparent citation trail linking back to the original sources. When a breach entry is created or updated, the source is logged in the database for accountability. Furthermore, entries include AI-generated confidence levels, enabling expert users to distinguish between high-certainty and tentative classifications. Low-confidence records are flagged for manual review by analysts.

### 4.3 Core Database Layer

The Core Database Layer acts as the centralized repository for AGS-INTEL, storing normalized and verified breach intelligence in a scalable, query-optimized structure. It accommodates the vision of a comprehensive global database with standardized entries, vulnerability integrations, and organization trust ratings, while overcoming

limitations in current solutions like incomplete metadata or lack of real-time access. This layer supports advanced querying, trend analysis, and risk forecasting.

The database schema is designed to support multidimensional breach intelligence. Each entry includes fields such as breach date, affected organization, breach type, exposed data types, source citation, credibility score, geographic scope, and linked vulnerabilities. Records are cross-referenced with organization profiles, which contain metadata like industry, size, location, historical breach count, and trustworthiness score. Vulnerability data includes CVE IDs, CVSS scores, exploit status, and MITRE ATT&CK mappings.

The tech stack for the database layer involves development tools that are well suited to handling large collections of data. AGS-INTEL incorporates a hybrid storage architecture leveraging a relational database built with PostgreSQL. This environment houses structured breach and vulnerability data and supports operations such as JOINs, aggregations, and filtering. Elasticsearch powers full-text and fuzzy search capabilities. Redis serves as a caching layer to accelerate high-frequency lookups (e.g., email and domain breach checks). Finally, TimescaleDB extensions support time-series analytics on breach timelines and trend assessments.

AGS-INTEL supports the following advanced features in its database implementation:

- **Indexing and Optimization**: To support low-latency queries over a large volume of breach records, the system applies aggressive indexing on critical fields, including breach type, domain, industry, and CVE IDs. Materialized views are used to precompute aggregated statistics for dashboard reporting. Partitioning by breach date and geography enhances scalability for regional queries.
- **Data Integrity and Versioning**: Breach records are treated as immutable snapshots to preserve auditability. If updated or corrected, a new version is stored with a reference to the original, and changes are logged. Each field is citation-tracked, maintaining a transparent record of the sources used to populate it.
- **Security and Privacy**: To ensure data privacy and regulatory compliance, AGS-INTEL employs a multi-layered security framework. Sensitive personal information, such as full email-password pairs, is never stored in plaintext. Inspired by *Have I Been Pwned*, passwords are processed using a k-anonymity model, where only the first six characters of a hash are stored to prevent direct re-identification (Hunt 2022). All data transmissions between system components and external interfaces are encrypted using TLS. Internally, strict access control policies govern write operations, restricting modification privileges to authenticated analysts and automated system processes. These safeguards collectively ensure data confidentiality, integrity, and auditability in alignment with GDPR and CCPA standards.

### 4.4 User Interface and Access Layer

The User Interface and Access Layer of AGS-INTEL enables stakeholders, including cybersecurity analysts, researchers, policy makers, and the general public, to interact with breach intelligence through an intuitive and transparent platform that supports diverse access modes, such as real-time dashboard visualizations and programmatic API integration. Users can query breach records in real time using multiple input modalities, such as email address, domain, company name, breach vector, CVE ID, or geographic region, with the lookup interface emphasizing performance and privacy. The interactive front-end dashboard allows users to filter breaches using a variety of characteristics, offering easily understood visual analytics. A customizable alerting system allows real-time notifications for relevant breaches. These features align with AGS-INTEL's mission to expand public breach knowledge.

## 5. Limitations, Risks, and Evaluation

While AGS-INTEL aims to deliver a robust and comprehensive platform for global data breach intelligence, certain limitations must be acknowledged. These challenges stem from the complexity of aggregating diverse data sources, ensuring reliability, and maintaining usability while adhering to ethical and legal standards. Below, we outline these limitations and describe how our system’s architecture incorporates mitigation strategies. We then present a framework for evaluating these mitigations post development to ensure the system performs effectively and evolves with emerging needs.

### 5.1 Data Ingestion Challenges

Collecting data from volatile sources, such as dark web forums that frequently change or go offline, risks incomplete coverage. AI-driven crawlers may also introduce biases, such as favoring English-language content or struggling with anti-scraping measures. AGS-INTEL mitigates these by using multilingual NLP models

supporting over 20 languages and adaptive crawlers that retrain on failed attempts. Redundant ingestion pipelines, including API integrations and user submissions, provide fallback options, while a source health monitoring system dynamically adjusts to unreliable feeds, ensuring broad and resilient data collection. Finally, since generative AI is not entirely resistant to hallucination, AGS-INTEL's verification layer cross-validates reports during credibility scoring to ensure the authenticity of ingested data.

### 5.2 Database and System Performance

Real-time ingestion and large-scale storage create performance challenges. Immutable versioning of breach records increases storage overhead, and balancing low-latency querying with resource-intensive analytics may result in trade-offs between speed and depth. Ensuring fault tolerance and scalability under surging data volumes remains a nontrivial engineering task that AGS-INTEL must address.

### 5.3 Security and Privacy

AGS-INTEL processes highly sensitive breach intelligence, necessitating robust safeguards for confidentiality, integrity, and availability. Despite protections like k-anonymity lookups and TLS encryption, risks remain such as re-identification from partial hashes, insider misuse, and adversarial manipulation of AI components. To mitigate these, AGS-INTEL adopts a multilayered defense architecture incorporating secure hashing algorithms, strict role-based access control, web application firewalls, input sanitization, and rate-limiting to resist denial-of-service and injection attacks. These technical controls are designed to ensure compliance with GDPR and CCPA principles of data minimization and pseudonymization.

Inspired by established platforms like Have I Been Pwned, AGS-INTEL's credential lookup system avoids storing personally identifiable information. Only compromised emails and passwords are retained, stored separately in isolated databases with no linking mechanisms to prevent reconstruction of user profiles. Users cannot browse the entire catalog; instead, they must query specific credentials, and these queries are not logged to preserve privacy. For general breach reports, only non-sensitive aggregate statistics are collected, such as the quantity of affected records, data categories, and broad impact descriptors (e.g., employees, customers, third-party collaborators), without retaining any raw leaked data.

Given AGS-INTEL's reliance on agentic AI for ingestion, verification, and enrichment, AI-specific vulnerabilities such as data poisoning are paramount concerns. Adversaries may attempt to exploit model outputs or introduce malicious inputs, potentially resulting in false breach entries or inflated credibility scores. To counter these threats, AGS-INTEL utilizes multi-source cross-validation, ML-driven anomaly detection (such as isolation forests), and mandatory expert review of low-confidence AI classifications. Periodic retraining on sanitized datasets further limits the propagation of adversarial data.

Security risks for AGS-INTEL are systematically assessed using a hybrid threat model (Table 1) that integrates the STRIDE framework (Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege). Threats are analyzed across all system layers—Ingestion, Verification, Database, and User Interface—to identify vulnerabilities, estimate impact, and match mitigations.

**Table 1: AGS-INTEL Threat Model**

| Category | Example Threats | Impact | Mitigations |
|---|---|---|---|
| **Spoofing (Authenticity)** | Impersonation of source feed providers to inject false data | Compromised data provenance, loss of trust | Signed API requests, mutual TLS authentication, source allow-lists, credibility verification algorithm |
| **Tampering (Integrity)** | Insertion of poisoned or adversarial data into ingestion or retraining datasets; manipulation of AI-generated metadata; tampering with in-transit database entries | Corrupted intelligence dataset, inaccurate public outputs | Data validation pipelines, anomaly detection (isolation forests), sanitized model retraining, manual review of anomalies, immutable versioning, input validation, WAF filtering |
| **Repudiation (Accountability)** | Insider actor hides activity or modifies logs | Reduced auditability and non-repudiation | Immutable, timestamped audit trails; cryptographic signing of logs |

| Category | Example Threats | Impact | Mitigations |
|---|---|---|---|
| **Information Disclosure (Confidentiality)** | Model inference or metadata leakage through analytical outputs | Privacy breach | Differential privacy in AI outputs, query rate-limiting, TLS 1.3 encryption |
| **Denial of Service (Availability)** | Overloading query APIs through automation | Disruption of alerts and data feeds | Redis caching, autoscaling clusters, CAPTCHA on interactive endpoints |
| **Elevation of Privilege (Authorization)** | Abuse of access token or privilege escalation by compromised analyst account | Full database exfiltration or alteration | RBAC with least privilege, zero-trust continuous authentication, token expiration policies |

### 5.4 Ethical and Legal Considerations

Collecting and disseminating breach intelligence raises complex ethical and legal challenges. Scraping and dark web monitoring may conflict with national laws, while storing and analyzing leaked personal data, even in hashed form, poses compliance risks under the GDPR, CCPA, and other regulations. It is important to consider applicable laws and address how AGS-INTEL maintains compliance.

Sourcing breach intelligence from dark web forums and public breach dumps involves acquiring and possessing stolen data, which risks violating the CFAA if handled improperly. According to the U.S. Department of Justice (2020), access must be obtained using legitimate credentials while adhering to all forum policies. Furthermore, it is critical to avoid any exchange of information on these forums. AGS-INTEL will limit activities to passively scraping publicly available discourse without engaging in communication with forum users, thereby minimizing legal risks. Additionally, purchasing stolen data poses significant concerns; to ensure ethical and legal compliance, AGS-INTEL will refrain from any dark web data purchases, sourcing only free, publicly released information.

Under the GDPR, hashed personal data, even when salted, is generally categorized as pseudonymized personal data, not fully anonymized data (GDPR, Art. 4(5)). As a result, it is subject to specific compliance obligations such as the right to access and the right of erasure (GDPR, Art. 15, 17). To ensure full compliance with the GDPR, AGS-INTEL will adhere to its foundational principles by processing any such data under a lawful basis of legitimate interests (GDPR, Art. 6(1)(f)), specifically the pursuit of enhanced cybersecurity resilience and public awareness of breaches, which would outweigh potential risks to data subjects given the platform's protective safeguards. Data minimization (GDPR, Art. 5(1)(c)) will be strictly enforced by aggregating only non-sensitive metadata from public sources, such as breach statistics, affected sectors, and vulnerability links, while avoiding the collection or storage of directly identifiable personal information beyond what is necessary for verification. Transparency and accountability (GDPR, Art. 5(1)(a) and (2)) will be upheld through detailed documentation of data processing activities, including source citation trails and audit logs. The right to access and erasure will be upheld through AGS-INTEL's freely available credential lookup system alongside the ability to request removal of credentials. Integrity and confidentiality (GDPR, Art. 5(1)(f)) will be maintained via the robust security measures described in the previous section. These measures not only mitigate compliance obligations for pseudonymized data but also position AGS-INTEL as a responsible actor that works to prioritize safety and privacy above all.

### 5.5 Evaluation Framework

To validate these mitigations and ensure AGS-INTEL's effectiveness after development, we will implement an evaluation plan with iterative six-month cycles combining quantitative metrics, qualitative feedback, and real-world testing. For verification and normalization, we will measure deduplication and classification accuracy against benchmark datasets. Our target is 99% deduplication accuracy, ensuring that duplicate entries remain rare occurrences manageable through periodic manual review. For classification, we aim for at least 92% accuracy, following the benchmark established by Noor et al. (2019) for machine-learning–based cyber threat analysis. Data ingestion coverage will be evaluated relative to leading repositories (e.g., HIBP, ITRC, PRC), with a target of $\geq$ 20% more global incidents captured. System performance will be evaluated through latency tests (targeting under 500ms for queries) and scalability under simulated high loads. Security will be tested via vulnerability scans and adversarial simulations, with a goal of zero critical findings in annual audits. Usability will be gauged through user surveys aimed at assessing ease of understanding, perceived usefulness, and willingness to recommend to others. Ethical and legal compliance will be verified through independent third-party reviews

to ensure continued alignment with data-protection regulations. A staged beta with controlled users will monitor live metrics like alert accuracy and system uptime (targeting 99.9%), benchmarks aligned with commercial-grade reliability expectations for real-time CTI services. As iterative testing progresses, user feedback will guide continuous improvements, while transparent publication of evaluation outcomes will foster accountability and stakeholder trust.

## 6. Conclusion

The conception of AGS-INTEL represents not only a technical contribution but also a rethinking of how cybersecurity knowledge can be governed in the age of AI. Rather than accepting fragmented repositories and uneven reporting as inevitable, this work envisions breach intelligence as a shared global resource that is transparent, verifiable, and continuously improving. The real test for AGS-INTEL lies not in its algorithms alone but in how well it earns trust across stakeholders: regulators who depend on credible oversight, organizations that weigh disclosure against reputation, laypeople seeking clarity, and researchers who require reliable data to advance the field. Moving forward, AGS-INTEL's framework provides a foundation for practical implementation and benchmarking against leading repositories. Its evaluation methodology enables systematic assessment of credibility scoring, multilingual coverage, and enrichment accuracy at scale. By maintaining an iterative cycle of testing and refinement, the system is positioned to adapt alongside the evolving threat landscape. Through its emphasis on authenticity, accessibility, and analytical depth, AGS-INTEL exemplifies how AI-driven intelligence platforms can redefine the future of breach analysis and digital trust.

**Ethics Declaration:** This research did not involve any activity that required ethical clearance.

**AI Declaration:** Artificial intelligence assisted in drafting and editing this paper. All core ideas, research, and design originate solely from the authors. The content has been rigorously validated for accuracy and fully reflects the authors' intent.

## References

Ayuso, E., Dumfeh Brogya, M.S., Kumar Ahlawat, V. and Sain, M. (2024) 'From manual to machine: How AI is redefining web scraping for superior efficiency: A literature review', in 2024 International Conference on Communication, Control, and Intelligent Systems (CCIS). IEEE, pp. 1–9. Available at: https://doi.org/10.1109/CCIS63231.2024.10931912 (Accessed: 7 September 2025).

DLA Piper (2024) DLA Piper GDPR and data breach report. Available at: https://www.dlapiper.com/en/news/2024/03/dla-piper-gdpr-and-data-breach-report (Accessed: 25 July 2025).

ENISA (2024) ENISA threat landscape 2024. Available at: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024 (Accessed: 25 July 2025).

European Parliament and Council of the European Union (2016) Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union, L119, 1–88. Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj (Accessed: 22 October 2025).

Have I Been Pwned (n.d.) About. Available at: https://haveibeenpwned.com/About (Accessed: 25 July 2025).

Hunt, T. (2022) Understanding Have I Been Pwned's use of SHA-1 and k-Anonymity. Troy Hunt [blog]. 30 June. Available at: https://www.troyhunt.com/understanding-have-i-been-pwneds-use-of-sha-1-and-k-anonymity/ (Accessed: 7 September 2025).

Identity Theft Resource Center (n.d.) Breach Alert. Available at: https://www.idtheftcenter.org/breach-alert/# (Accessed: 25 July 2025).

Kumar, S., deWitte, P. and Gu, G. (2025a) 'Incentivizing Security Excellence in Cyber Liability Insurance', in 2025 IEEE 10th European Symposium on Security and Privacy (EuroS&P), Venice, Italy, 2025, pp. 251-267. Available at: https://doi.org/10.1109/EuroSP63326.2025.00023 (Accessed: 15 September 2025).

Kumar, S., Garg, A. and Niranjan, M. (2025b) 'Enhancing Government Efficiency Through Cybersecurity Hardening', Conference on Digital Government Research, 26. Available at: https://doi.org/10.59490/dgo.2025.1047 (Accessed: 15 September 2025).

Leak-Lookup (n.d.) Dashboard. Available at: https://leak-lookup.com (Accessed: 25 July 2025).

MITRE (n.d.) About. Common Vulnerabilities and Exposures. Available at: https://www.cve.org/About/Overview (Accessed: 25 July 2025).

Noor, U., Anwar, Z., Malik, A.W., Khan, S. and Saleem, S. (2019) 'A machine learning framework for investigating data breaches based on semantic analysis of adversary's attack patterns in threat intelligence repositories', Future Generation Computer Systems, 95, pp. 467–487. Available at: https://doi.org/10.1016/j.future.2019.01.022 (Accessed: 7 September 2025).

Novaes Neto, N., Madnick, S., De Paula, A.M.G. and Borges, N.M. (2021) 'Developing a global data breach database and the challenges encountered', Journal of Data and Information Quality, 13(1), Article 3. Available at: https://doi.org/10.1145/3439873 (Accessed: 7 September 2025).

Privacy Rights Clearinghouse (n.d.) Data breach chronology. Available at: https://privacyrights.org/data-breaches (Accessed: 25 July 2025).

Romanosky, S. (2016) 'Examining the costs and causes of cyber incidents', Journal of Cybersecurity, 2(2), pp. 121–135. Available at: https://doi.org/10.1093/cybsec/tyw001 (Accessed: 7 September 2025).

SpyCloud (n.d.) Frequently asked questions. Available at: https://spycloud.com/faqs/ (Accessed: 25 July 2025).

Trend Micro (n.d.) About ZDI. Available at: https://www.zerodayinitiative.com/about/ (Accessed: 25 July 2025).

U.S. Department of Justice (2020) 'Legal considerations when gathering online cyber threat intelligence and purchasing data from illicit sources'. Available at: https://www.justice.gov/criminal/criminal-ccips/page/file/1252341/dl?inline (Accessed: 22 October 2025).

Verizon Business (2025) 2025 Data Breach Investigations Report. Available at: https://www.verizon.com/business/resources/Tea/reports/2025-dbir-data-breach-investigations-report.pdf (Accessed: 25 July 2025).

XposedOrNot (n.d.) Frequently asked questions. Available at: https://xposedornot.com/faq (Accessed: 25 July 2025).

Xu, Y., Chen, G., Wu, J., Xu, W. and Liu, Q. (2021) 'Research on dark web monitoring crawler based on TOR', in 2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA). IEEE, pp. 197–202. Available at: https://doi.org/10.1109/ICIBA52610.2021.9687954 (Accessed: 7 September 2025).