



Automated Data Security Posture Management in the Age of AI



Automated DSPM in the Age of AI

Introduction: Traditional vs. AI-Powered DSPM

Traditional Data Security Posture Management (DSPM) is a comprehensive cybersecurity approach to identifying, assessing, and managing enterprise-wide security risks across an organization's data ecosystem. However, traditional DSPM platforms have been operating for 30+ years with little innovation, except minor automation of highly manual DSPM processes including data discovery and classification, and IT workflows. Traditional as well as contemporary DSPM teams still heavily rely on conventional processes in their data security operations to protect (mostly) known data types in structured repositories - which falls far short of the modern capabilities required to continuously observe, assess and reduce vulnerabilities to comprehensively protect all enterprise data.

Fast forward to today, enterprises of every size globally are scrambling to integrate AI into their business functions to reduce costs, increase revenue or gain market share. Legacy companies are also beginning to compete with a new breed of automated enterprises in which AI is native to the core of their business operations. Think of an AI-native business as one in which AI is a core business infrastructure like cloud-native businesses in which cloud infrastructure is core to operations. In both legacy and AI-native enterprises, data flowing 24x7 is the lifeblood of enterprise value, offering the information and knowledge required for millions of better business decisions faster to quickly respond to evolving market needs and capitalize on new value-opportunities.

In the longer run, traditional DSPM operators without AI will not be able to keep up and compete. And current DSPM operators touting AI/ML by baking it into legacy DSPM operations still struggle to meet the needs of business enterprises in the Age of AI. Like the new breed of AI-native-core enterprises, truly modern DSPM platforms require AI-powered capabilities at their core (AI-native DSPM platform solutions) including the automated and continuous monitoring of data and knowledge assets, access controls, policies and configurations.

With all of this said, a real solution to today's DSPM gaps does exist. Modern, AI-native DSPM that expands into Knowledge Security Posture Management (KSPM) offers fully automated data Observability, Assess-ability and Enforceability that ensure adequate cybersecurity measures for continuously securing and protecting data assets 24x7 and throughout the asset lifecycle.

Legacy DSPM Problems in the Age of AI

As we enter the Age of AI, data security remains one of the least well-defined issues, even as enterprises are evaluating AI strategies to replace or transform business functions through intelligent automation. With vast amounts of data already being ingested to train LLMs, the issues associated with finding incremental sources of training data are already on the table including ways to create synthetic data sources, ingesting recent or real-time data, (data recency) and ways to train models using feedback loops of enterprise decisioned-data. To summarize – data is vastly important to the valuations of today's LLMs and the automated enterprise services being developed on top of them.

With enterprise data now flowing 24x7, it is quite surprising, if not completely shocking, how little dialogue is happening about securing this data flow into / out of today's AI-enabled enterprises. Even more so given the risks posed by unsecured data or weak enterprise security postures. If data is the lifeblood of the digital enterprise, the

incremental value created by AI is dependent on protecting the static and dynamic data used to train the ML models that intelligently automate business functions. Unknown data movement is one source of enterprise risk. But new types of AI data security risk are emerging such as data poisoning. The statistics are very alarming:

According to IBM's 2024 Securing Generative AI Report in collaboration with AWS¹ - of the 82% of respondents who say, "secure and trustworthy AI is essential to the success of their business", nearly 70% say "innovation takes precedence over security." **In fact, only 24% of enterprise AI projects are being secured today**¹.

Given the sense of urgency, priority and valuation placed on today's enterprise AI projects, AI data security must be one of the top priorities for DSPM's teams. **However most traditional DSPM tools are no longer meaningful or valuable to enterprise teams wanting to secure their AI projects or the data that feeds their AI algorithms** – while the capabilities of both legacy and contemporary DSPM tools fall far short of understanding and protecting enterprise data. The biggest data security problems facing digital enterprises today challenge the limitations of current DSPM offerings in the Age of AI:

1. **Observability/Visibility Crisis from Data Sprawl** Enterprises have lost track of where sensitive data resides across multi-cloud environments, SaaS applications, hybrid infrastructures and on-prem unstructured repositories. Most enterprise CISOs and CIOs can't answer basic questions like "what sensitive customer data do we have and where is it located?" Shadow data proliferates as business units create databases, file shares, and cloud storage without IT oversight, creating massive blind spots that regulatory auditors, security teams and DSPM platforms can't adequately assess.
2. **Data Discovery Gaps, Classification and Governance Failures** Manual data classification processes simply can't scale with modern data volumes in the Age of AI. Enterprises struggle to accurately find, identify and classify sensitive data, information and content in real-time as data growth and sprawl rise exponentially. Classification is often a side IT function of DSPM teams. Furthermore, sophisticated classification criteria required by AI-powered business applications is not even supported by the traditional DSPM classification paradigms. Most enterprises lack automated data classification tools, resulting in inconsistent protection levels and policy enforcement.

Without understanding context of sensitive data, security teams apply blanket controls that either over-restrict business operations or under-protect critical assets - a constant DSPM struggle. Furthermore, traditional classification paradigms cannot support more sophisticated classification criteria as required by the more demanding business applications.

3. **Compliance and Regulatory Pressure Regulations** like GDPR, CCPA, and HIPAA demand precise data lineage and access controls that current security tools and DSPM providers can't possibly provide at the scale of data volume in the Age of AI. Organizations face severe financial penalties for non-compliance while struggling to maintain continuous compliance across distributed data environments.

New AI regulations are underway that will bring even greater regulatory pressure to global digital companies.

4. **Lack of Scalability** Manual compliance in today's DSPM processes can't scale with modern data velocity, creating constant audit exposure and regulatory risk. Data professionals are seeing data volumes growing by an average of 63% per month - and nearly six in ten organizations say they can't keep up².

¹ Securing Generative AI by IBM's Institute for Business Value in Collaboration with Amazon AWS, 2024

² [Dataversity, 8/14/2023](https://www.dataversity.net/ai-data-growth/)

5. **Cloud Security Blind Spots** Traditional perimeter-based security fails in cloud-native environments where data flows dynamically between services. Organizations can't monitor or control these data movements effectively. As enterprises migrate more data to the cloud, sensitive data scatters across multiple cloud providers, geographic regions, and services.
6. **Excessive Access and Privilege Sprawl** Over-privileged users and applications create massive data attack surfaces. Organizations lack visibility into who has access to what data and whether permissions align with security policies. Traditional perimeter-based security models fail in cloud-native architectures, leaving data exposed through misconfigurations, over-privileged access, and inadequate encryption.
7. **Insider Threats and Over-Privileged Access Employees** often retain access to sensitive data long after their roles change, creating significant insider risk. Legacy access management systems can't provide granular, context-aware controls needed for modern zero-trust architectures, leading to data breaches from both malicious and accidental insider actions.

The Financial Impact of DSPM Failures

Cybersecurity Ventures projects that cybercrime damages will reach \$10.5 trillion annually by 2025³, with data-related incidents representing approximately 60% of total cyber losses— translating to over \$6 trillion in annual global data security risk. Legacy DSPM solutions attempt to address this risk but are losing ground in the Age of AI. Increasing losses from regulatory fines, breach costs, and business disruption compel the need for modern DSPM solutions that are scalable, comprehensive and effective for modern enterprise data security strategies. Many types of losses are cited by leading cybersecurity research firms:

1. **Data Breach Costs** According to IBM's 2024 Cost of a Data Breach Report, the global average cost of a data breach reached \$4.88 million, with healthcare breaches averaging \$11.05 million. For enterprises with over 50,000 employees, average breach costs exceed \$7 million. With over 3,200 publicly disclosed breaches in 2023 affecting billions of records, the aggregate annual impact exceeds \$15 billion globally. Of course, these are just the “hard costs”. Breaches result in significant costs from reputational loss and business disruption which can be equally large.
2. **Regulatory Compliance Penalties** Gartner research indicates that regulatory fines for data protection violations reached \$1.6 billion in 2023, with GDPR fines alone totaling over \$2.9 billion since their release.
3. **Cloud Security Misconfigurations** According to Forrester's 2024 State of Cloud Security report, 83% of organizations experienced at least one cloud security incident in the past year, with average remediation costs of \$1.2 million per incident. Gartner estimates that in 2025, 95% of cloud security failures will be due to customer misconfigurations rather than provider vulnerabilities.
4. **Data Governance Failures** IDC research indicates that poor data governance costs large enterprises an average of \$15 million annually in operational inefficiencies, compliance gaps, and missed business opportunities. Forrester estimates that data quality issues alone cost the US economy \$3.1 trillion annually.

³ <https://cybersecurityventures.com/official-cybercrime-report-2025/>

Why Today's DSPM Capabilities Are Inadequate in the Age of AI

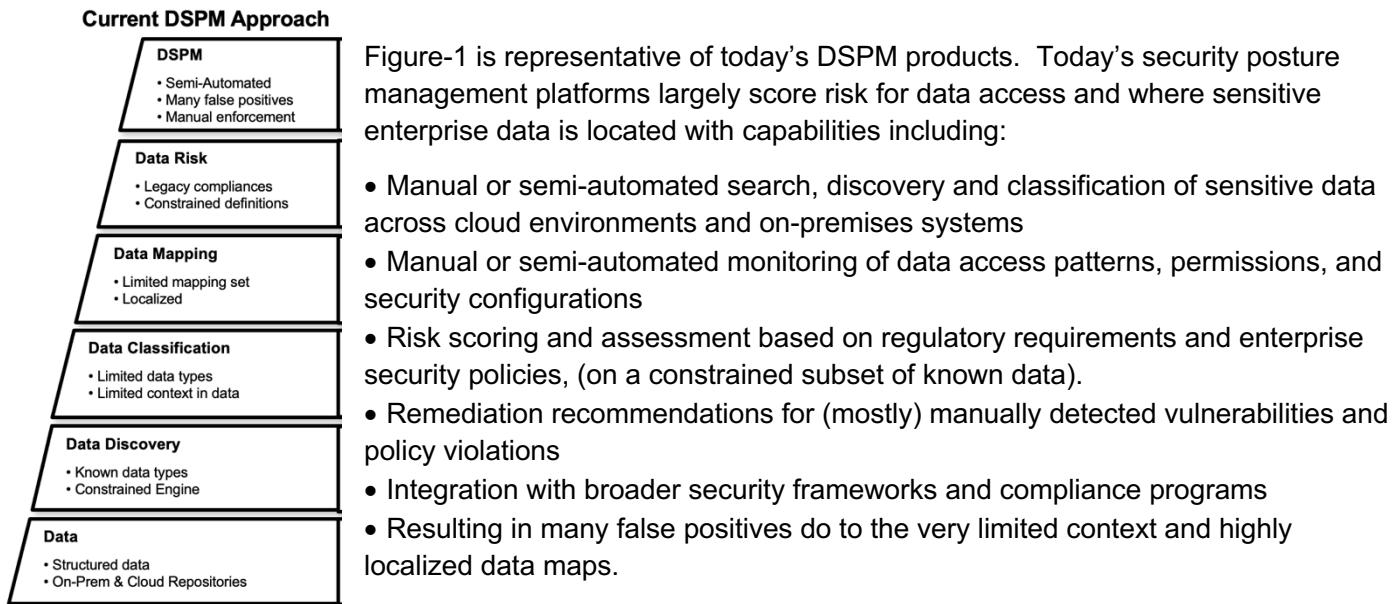


Figure 1.

Important DSPM Questions for the Age of AI:

In today's fast paced AI-powered world, manual or semi-manual searching and mapping of highly constrained, known data types (that is at best sampled during repository scanning), is no longer sufficient for modern data compliance and DSPM because of the massive volume, sprawl and complexity of structured and unstructured enterprise data flowing 24x7. How can any manual DSPM process or infrequent repository scan provide the recency required to provide an accurate, up-to-date data security posture score? How can the desirable accuracy be achieved when data, information and content flowing 24x7 can be of any known or unknown structure?

How can data sampling of mostly structured data (in today's AI/ML DSPM platforms) ever provide the necessary comprehensiveness to create a precise security posture score that reflects true data risk on today's largely unconstrained data files? How can this score be accurate for DSPM processes that only assess a subset of highly constrained, structured enterprise data - while neglecting unstructured data which makes up 80% of enterprise data today?

And as security policies are updated daily or multiple times per day, how can current DSPM platforms enforce data policies when unknown internal dependencies exist from recent policy changes? How can security teams take corrective action if malicious insiders change data tags, and security alerts are issued to IT teams based on outdated access controls that can no longer be trusted for these (changed) policies?

These questions and many more reflect critical, unmet DSPM needs and challenges of today's manual or semi-automated DSPM platforms resulting in very large gaps in enterprise security postures.

Why Deep-AI-powered DSPM is a Much-Needed Paradigm Shift

In the Age of AI, Deep-AI powered, comprehensive and fully automated DSPM is required to dramatically expand data Observability, Assessability and Enforceability to keep up with enterprise data flowing 24x7 while adequately protecting a compliant enterprise. The next-paradigm DSPM capabilities needed extend every layer of current DSPM operators (in Figure-1.) while **shifting the focus from protecting enterprise data, to protecting enterprise knowledge (that contains data)**.

To make this transition from data types to “knowledge object types” (that comprise data) we have the following transformed DSPM capabilities that include a knowledge representation layer (in black below). This Knowledge Object layer enables comprehensive knowledge discovery resulting in significantly improved capabilities, resulting in comprehensive enterprise DSPM:

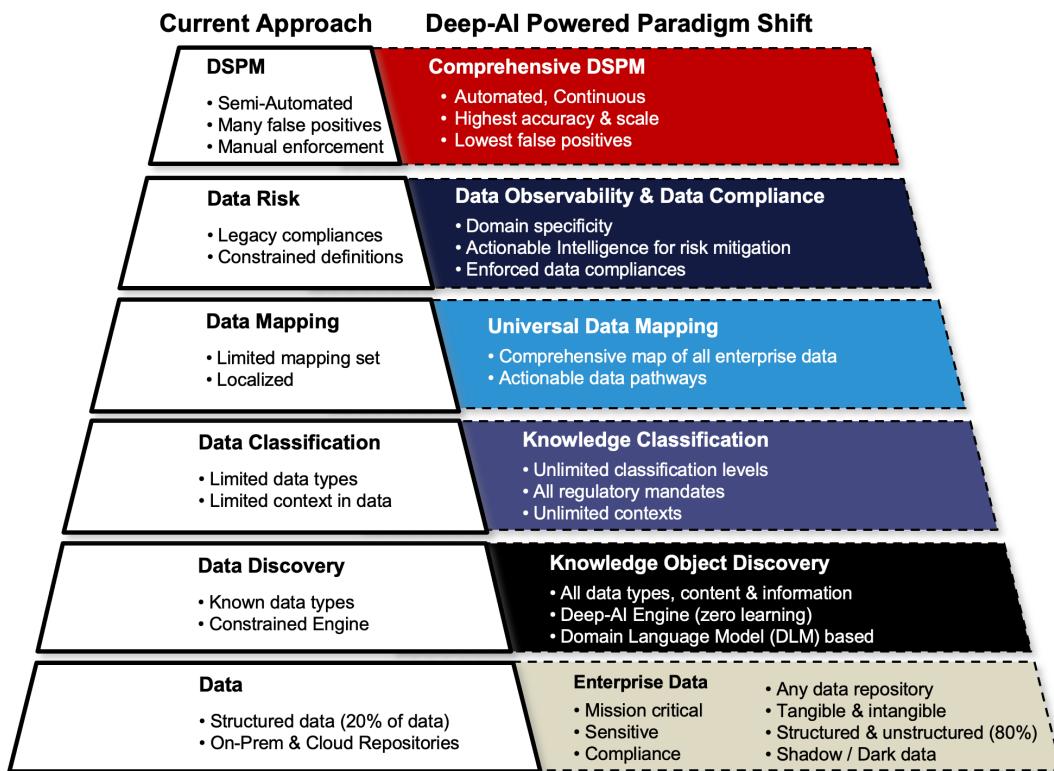


Figure 2.

Next-Paradigm Knowledge Object Layer

By introducing knowledge objects as an abstraction layer above the enterprise data, AI-powered knowledge discovery algorithms can find and classify any type of data without constraints. The introduction of knowledge objects into DSPM not only enables AI-powered automation at every DSPM capability above this layer – but also expands the outcomes and value provided by each DSPM layer in Figure-2. AI-powered DSPM with next-paradigm capabilities provide robust data observation, assessment, enforcement

and risk remediation for any kind of data repository, on-prem or multi-cloud, across any kind of data environment or structure, at ultra-high scale, while dramatically reducing regulatory compliance risk with a single scan.

Significantly Improving Data Context and Structure

Shifting to a knowledge-object-based DSPM paradigm dramatically increases data context to improve discovery accuracy and reduce false positives. For example, today's AI/ML offers "context-aware data classification" by seeking helpful text labels or other tokens close to the sensitive data in question. For example, finding the text "SSN" by a customer's Social Security Number. But if this SSN alpha-text is not present in the data or sufficiently close to the actual number then the needed context is not present for the system to correctly classify the Social Security Number as such. Knowledge discovery objects eliminate this issue.

In a knowledge-object paradigm, when the algorithms discover a knowledge object such as "driver's license" or "passport", the context of the data contained in the object is known precisely because the system "knows" the "object" ahead of discovery, (since the object's canonical data is pre-defined and represented). For example, if the knowledge object is a US state driver's license, the AI system will know the meaning of all the data contained in the driver's license, because a "driver's license" knowledge object is represented in the system ontologically and epistemologically.

The system therefore expects to see Name, Address, State, DOB and driver's license number. **When knowledge objects are discovered in unstructured data repositories, the meaning of the unstructured data within the knowledge object is identified easily and quickly, with classification tags written back to provide valuable structure in the unstructured repository.** Writing classification tags back to data repositories enables dramatically improved and continuous understanding of unstructured data, with much more comprehensive security risk posture management because 80% of today's enterprise data is unstructured.

Conventional DSPM teams searching for patterns in regular expressions with automated scripts cannot compete with comprehensive AI-powered algorithms based on knowledge objects. Current search methods can only find data types known in advance of the search. Knowledge-object based discovery finds both known and unknown data types without constraint. Customized knowledge objects can also be created by DSPM teams to find and classify important enterprise-specific data fast, such as the data in a custom medical form for an insurance provider or other kinds of specialized enterprise content and information.

Last, in a knowledge-based paradigm, the AI algorithms in the Deep-AI engine implement non-ML disciplines of knowledge object representation, inference calculus and planning. No ML means no expensive ML overhead to train ML models. Knowledge-based AI identifies and classifies data, model dependencies and policies in real-time such that exposed vulnerabilities can be automatically remediated as well as data subject access requests (DSARs submitted by customers). **Deep-AI-powered DSPM is the next generation of security posture management offering unparalleled DSPM Observability, Assess-ability, and Enforceability.**

Expanding DSPM Observability

Enterprise "observability" of enterprise data is a more expansive, all-encompassing capability that comprises requires multiple essential features:

1. **Data Discovery:** Inherent to comprehensive “Data Observability” is the assumption that the enterprise is 100% capable of discovering known and unknown enterprise data. You cannot observe data you don’t know you have or data you can’t identify. The types of data include:

- 1. Mission critical data
- 2. Confidential and sensitive data
- 3. Compliance-dependent data
- 4. Shadow / Dark data

In any digital organization this type of data is a “canonical” or most elemental form of knowledge. This data/knowledge is stored in an enterprise such as knowledge of the customer (name, street address, phone and email are contact object with contact data), knowledge of their legal driving status (driver’s license object with identity data), knowledge of their international travel status (Passport object with passport data) or knowledge of their medical condition (Medical diagnosis object with medical data).

2. **Knowledge Object Discovery:** Inherent to comprehensive “Data Observability” is the assumption that the enterprise is 100% capable of discovering known and unknown enterprise data types. You cannot observe data you don’t know you have or data you cannot identify. Knowledge objects are also derived from both tangible and intangible data and serve as a store of value in an enterprise – with tangible data like “social security number” or “street address” but also intangible data of unknown structure that is conceptual in nature, or custom to an enterprise (Driver’s License, Passport, Contact or Medical Record. Observability at the knowledge object and data levels is more valuable to an enterprise than simply searching for and finding individual (canonical) data elements. Knowledge objects are much higher-value forms of data value and dramatically improve the enterprise’s observability of known and unknown data. Examples follow:

E-commerce Conceptual Objects

- Purchase Order
- Sales Receipt
- Travel itinerary
- Invoice
- Billing or Delivery Address
- Identity proof like driver’s license or passport

Medical Conceptual Objects

- Patient’s Health Examination Report
- Blood Test or Lab Test Results

Financial Services Conceptual Objects

- Credit Report
- Business Entity Type
- Payment or Spending Record
- Spending Record

3. **Classification:** A next-gen DSPM platform can discover, identify and classify knowledge objects and data using unlimited levels of classification along topical, security and geographic dimensions with algorithms that can perform cross-domain classifications in real-time as data and knowledge objects are being scanned. Classification tags or metadata should be written back to the repository. In this manner a single repository scan can discover knowledge objects and data across regulatory mandates providing unmatched contextual accuracy to correctly identify and classify any data type or knowledge objects for any regulatory mandate. In real-time tags are written back to the repository to provide actionable pathways to easily locate the knowledge objects and data and use AI to automatically perform functions on this classified data.

4. **Mapping:** DSPM Mapping must be suitable to not just discover and map sensitive data and provide its location - but also to provide actionable compliance intelligence with domain-level insights continuously. Through intelligent automation, actionable domain-level intelligence enables DSPM

teams to take immediate action and remediate vulnerabilities - continuously improving the enterprise's security posture across all forms and domains of sensitive data.

DSPM Assessability

DSPM risk assessment is highly dependent on overall data observability. If data isn't observable (like the 80% of enterprise data that is unstructured data), then its contribution to the overall security posture can't be assessed, and policy violations for this data cannot be observed or enforced either. Maximum DSPM "Assessability" requires comprehensive data Observability. The more data observed, identified and classified, the more data can be assessed and policy violations surfaced and repaired. "Assessability" includes:

1. **Real-Time Data & Mapping Intelligence:** Data intelligence means enterprises have up-to-date, precise and reliable information plus actionable intelligence about their data, security controls, compliance status, and risk posture. Real-time intelligence for comprehensive data and knowledge objects requires maintaining current and accurate data mapping containing on any data repository that features actionable intelligence in rapidly changing digital environments – which requires continuous discovery and classification as new data is created or modified.
2. **Dynamic Risk Scoring based** on data access controls, data sensitivities, and current regulatory landscapes with scoring for each repository should be aggregated and visualized in different ways. For example. the overall DSPM score for a repository vs. a group of repositories or the enterprise score. The regulatory compliance risk score should include components for compliance status monitoring for one or more regulatory requirements.
3. **Accurate and Dynamic Risk Assessment:** Most current DSPM platforms score risk mostly for data access and where sensitive enterprise data is located. AI-based DSPM provides precise compliance risk assessment with improved recency for both structured and unstructured data repositories which account for 20% and 80% of all enterprise data respectively. But automated and continuous risk assessments of Deep-AI-powered DSPM platforms transcend current risk scores, by including in risk scoring signals for access controls to repositories.
4. **Up-to-Date Policy Compliance:** The intelligently automated, real-time monitoring and assessment of regulatory compliance policy violations and compliance drift is a must-have for modern DSPM. Robust DSPM policy compliance platforms will integrate to enterprise DLP security systems such that current access privileges can be validated and with recent certification status. Automated compliance checks also require the measurement of security control effectiveness.

DSPM Control and Enforceability

DSPM Enforceability and control represents an organization's ability to actively manage, govern, and control data access, movement, and protection through automated policies, technical controls, and responsive DSPM and security measures. This dimension focuses on the automated remediation ("doing") aspect of DSPM - taking action to secure data assets and to increase the enterprise's data security posture score.

Modern enforceability requires automated, accurate and efficient determination of access control and use of data. This requires automatically determining or “synthesizing” pertinent access policies for data and knowledge objects that balance access constraints with permissions. Too little access and the enterprise productivity is impacted. Too much access does not address critical security and privacy concerns.

With intelligently automated access policies for both knowledge objects and data in place, modern DSPM enforceability and control enables the automatic enforcement of policies. By enforcing controllability at the knowledge object level, the data comprising the objects can inherit the enforceability benefits as well. Critical enforceability features include:

- Extracted and displayed accessibility contours and visualizations so that enterprise DSPM teams and users can easily visualize controls and act or setup automated actions for access violations.
- Deep-AI powered algorithms used to derive the accessibility and use the extracted contours to evaluate and surface compliance policy violations.

Next-Paradigm DSPM Capabilities Driving Next Level Outcomes

A knowledge object-based paradigm of a modern DSPM compliance platform addresses all the problems of legacy DSPM solutions while future proofing the enterprise in the Age of AI. By inserting a data abstraction layer to represent enterprise knowledge, the Deep-AI engine (shown in Figure 3 below) empowers enterprise DSPM teams with fully automated data discovery, identification, classification and risk assessment for any data type in real-time, across any repository, at ultra-high speed. As DSPM teams learn to compose more advanced knowledge data objects, the AI-powered data mapping creates more user-friendly visualizations for enterprise and end users to become aware of strong dependencies between valuable enterprise knowledge objects.

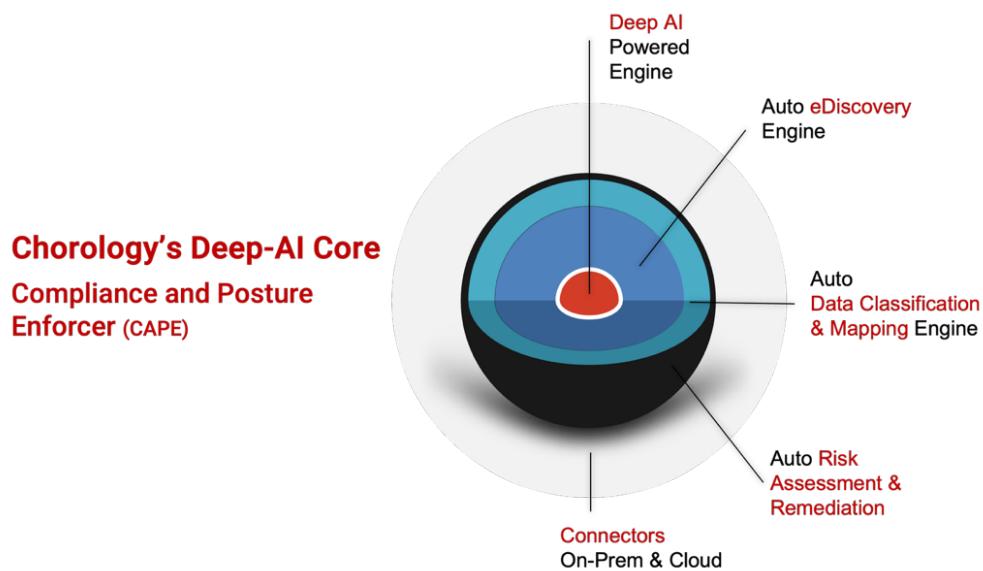


Figure 3.

The Deep-AI-powered compliance core ingests any type of data including canonical data types that together represent a composite knowledge object. This Deep-AI core enables real-time comprehensive data observability with dramatically improved accuracy and currency of sensitive knowledge objects and enterprise data. With

dramatically improved data Observability and Assessability, DSPM teams can have better control and enforceability at a much lower total cost of ownership due to the automated compliance processes powered by a Deep-AI engine core.

CAPE: A New DSPM Paradigm in the Age of AI

As the Age of AI emerges in which AI assists or optimizes manual business functions, the new thinking required to address the security and protection of data flowing 24x7 and to automate comprehensive enterprise DSPM is available now and Chorology is the only next-generation DSPM platform providing it. Automated DSPM processes acting on knowledge objects with sensitive data enables a next-paradigm, Deep-AI DSPM engine to power a fundamentally new core that transforms the DSPM industry instead of “bolting AI on to” traditional DSPM solutions after the fact.

Figure-4 below gives a depiction of the next-generation CAPE platform.

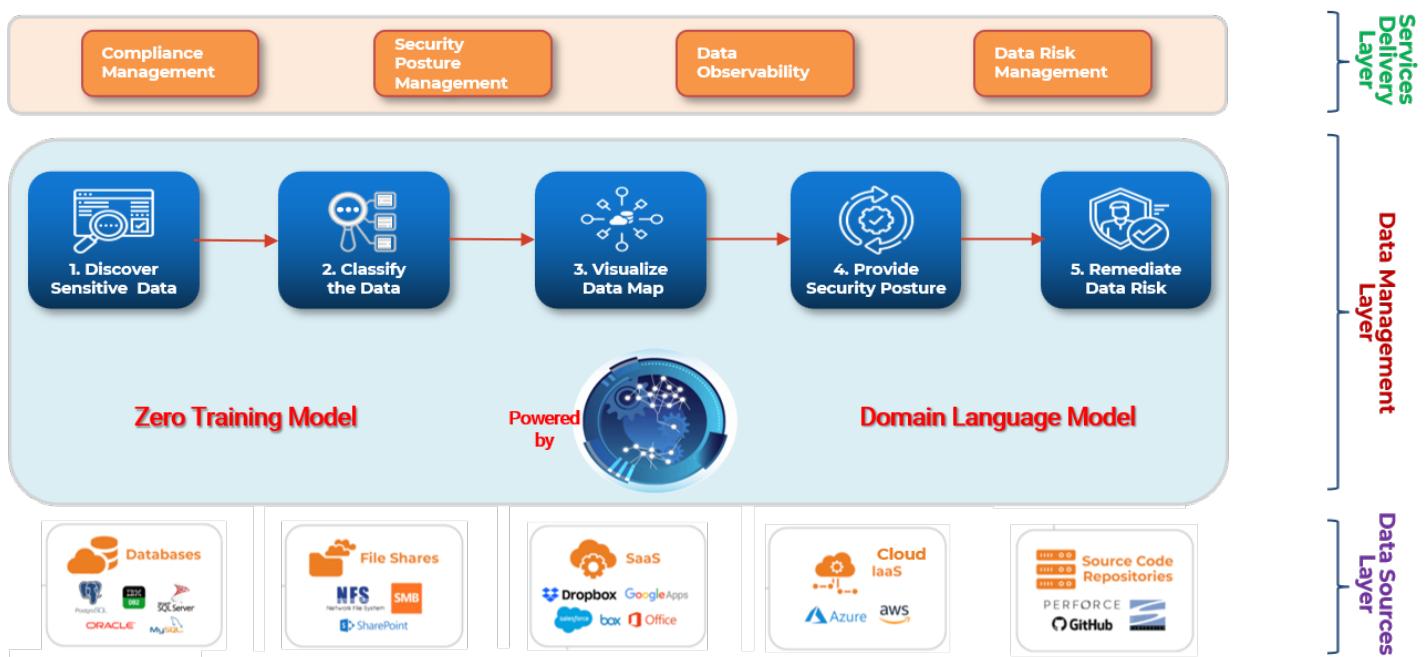


Figure 4.

CAPE introduces a knowledge and data object abstraction layer and Domain Language Modeling to make domain-specific, tangible and intangible knowledge objects discoverable. CAPE's Deep-AI engine can then intelligently automate the core DSPM processes of Data Discovery, Classification, Mapping and Posture Scoring based on deterministic AI disciplines of knowledge representation and inference calculus.

AI planning enables CAPE to also offer comprehensive, actionable alerts and insights that enable the automatic remediation of DSR's and the remediation activities required to elevate data security posture as well.

CAPE by Chorology.ai represents a new DSPM paradigm that automates DSPM processes, accounts for system and data dependencies, enforces continuously modified policies while dramatically increasing the speed, recency and scale of DSPM Observability, Assessability and Enforceability.

Through this Deep-AI based architecture, CAPE offers next-level DSPM outcomes including comprehensive enterprise knowledge and data understanding that continuously identifies, maps and visualizes data security postures with the highest accuracy and recency at ultra-high scale, with the lowest false negatives, at the lowest total cost.

In the Age of AI, CAPE by Chorology.ai is the only operator with a Deep-AI foundation offering a whole new paradigm of DSPM capabilities that are designed to keep pace with today's enterprise data growth and sprawl - with data flowing 24x7 - while future-proofing enterprises and organizations for new and more demanding regulatory environments including future AI mandates.

SUMMARY

Achieving and maintaining compliance and data security posture management in high-volume, sprawled data environments is a complex but essential task for organizations navigating today's regulatory landscape in the Age of AI. As highlighted in this whitepaper, the inherent Observability, Assessability, Enforceability, Flexibility and Scalability of the CAPE platform solves unique security challenges that demand a proactive and automated approach to protecting enterprise data and knowledge. CAPE's knowledge object framework powered by a Deep-AI engine based on the disciplines of knowledge representation, inference calculus and domain language modeling intelligently automate DSPM processes to comply with regional and global regulatory mandates including GDPR, CCPA/CPRA, HIPAA, and others. With CAPE, now organizations can establish robust and peta-byte scale compliance and DSPM strategies tailored to their operational needs to observe, assess and control their data security posture management. Automation of data and knowledge discovery, classification, mapping and posture management play a pivotal role in simplifying compliance efforts—essential tools for Data Security Posture Management and Knowledge Security Posture Management, that will significantly reduce manual overhead while significantly improving accuracy and recency.

Moreover, embracing best practices such as continuous monitoring, detailed repository and global DSPM reporting, and managed data security policies improves long-term compliance and reduces long-term risk of storing sensitive data in both structure and unstructured repositories. Ultimately, tackling data security compliance is not just about meeting regulatory requirements—it's about strengthening organizational security, minimizing risks, and fostering trust among customers and stakeholders. By addressing these challenges head on and leveraging the next paradigm in DSPM platforms outlined in this guide, enterprises can transform compliance from a challenge into a competitive advantage, and future-proof their organizations for future mandates in the Age of AI.

To learn more or for a free demonstration of the CHOROLOGY.ai platform, please visit <https://chorology.ai/>

About CHOROLOGY.ai

CHOROLOGY.ai is headquartered in San Jose, Silicon Valley, California with development offices in South Asia. Contact info@chorology.ai, (408) 713-3303, 2001 Gateway Place, Ste: 710 West Tower, San Jose, CA-95110, USA.

Learn more by visiting <https://www.chorology.ai/>.