Chorology.ai
KNOW YOUR DATA

White Paper

Chorology.ai
KNOW YOUR DATA

Introducing CAPE™
Compliance and Posture Enforcer

Intelligently Automated Data Discovery, Classification and Compliance Enforcement powered by Deep AI

Chorology.ai
KNOW YOUR DATA

1. Executive Overview

In recent years, Consumer Data Privacy and Protection has become one of the most significant security issues. The misuse of consumer data by eCommerce merchants and service providers in activities other than for use in the maintenance of customer information, has historically caused much concern amongst the general consumers of these services. Furthermore, the unauthorized disclosure, sale and sharing of the consumer information without their consent result in a major privacy exposure for the consumers, and has been a topic of increasing concern. Over the recent years, therefore, many regulations and compliance mandates have been sanctioned against such practices by the government and regulating authorities in many parts of the world.

Businesses are required to comply with these government regulations and mandates such as GDPR (General Data Privacy Regulation Act) in the European Union. In the United States similar mandates have been created. For example, CCPA (California Consumer Privacy Act) has been mandated in the state of California while several other states in the US are also formulating similar compliance acts. These Privacy laws are set up to protect personal information, financial information, consumer behavioral information and other such data from any wrongful perusal or for business/financial gains without the prior consent of the consumer.

MANY FORMS OF COMPLIANCE:

Corporate Data Compliance – First Generation Data Compliance and Privacy Mandates:
The First-Generation Data Privacy mandates were primarily focused on Corporate specific Mission Critical Data such as HR Data, Financial Data, SEC Data, Sales & Marketing Data, etc. These regulations were mostly enacted as a part of Sarbanes Oxley (SOX) Compliance mandate.

Industry Data Compliance – Second Generation Data Compliance and Privacy Mandates:
The Second-Generation Data Privacy mandates widened the scope to focus on protecting a broader range of Industry Specific data such as Payment Card Industry (PCI), Personally Identifiable Information (PII), Health Insurance Portability and Accountability Act (HIPAA), Gramm-Leach-Bliley Act (GLBA), etc.

Consumer Data Compliance – Third Generation Data Compliance and Privacy Mandates:
The Third-Generation Data Privacy mandates have further widened the scope of data compliance regulations to protecting even a broader range of consumer data such as a consumer's Personal Information (PII), Financial and Payment based information (PCI), On-line Behavioral Information (such as history of on-line purchases, financial transactions, personal preferences viz. a viz. merchandise, web-site browsing history, pattern of "likes" or "dislikes" of online services/web-sites/individuals/etc.), and any other information that represents any aspect of the **"*persona*"** of an individual or consumer.

Chorology.ai
KNOW YOUR DATA

➢ What are Consumer Data Privacy Regulation Mandates?

Several Consumer Privacy Mandates have been enforced in many parts of the world. As depicted below in Fig-1, the General Data Protection Regulation (GDPR) mandate in the European Union, has prescribed clear guidelines for consumer rights to demand access, deletion, privacy, non-disclosure, non-sale or non-use of their data by the entity that may hold any consumer data – not to mention the ***'Right to be Forgotten'***. Similarly, in addition to California Consumer Privacy Act (CCPA) in California, many other states in the US are also formulating their respective compliance mandates. Several other countries have respective mandates such as LGPD, NZPA, PDPA, POPIA, APPS, DIFC, PIPEDA, ADHICS, ECL, LPPD, DPA, PDPO, DPA, UAE, etc.

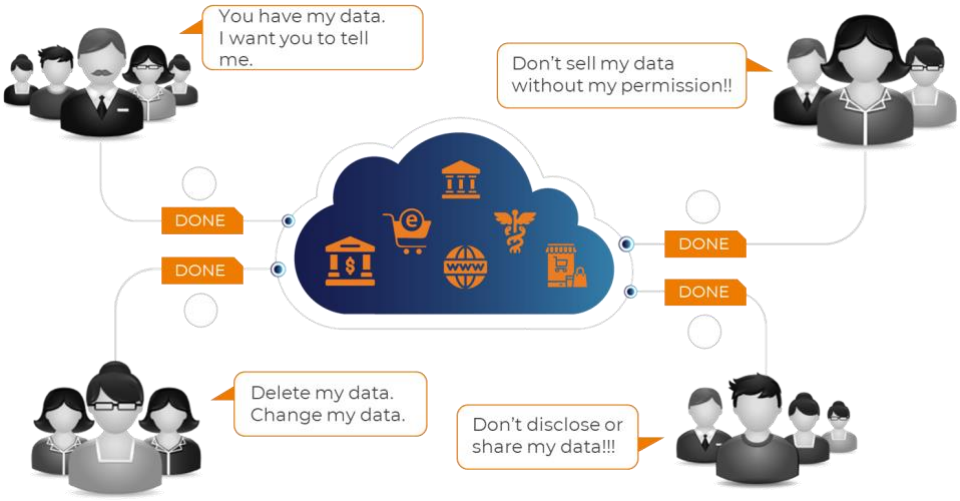


Fig-1: Consumer Data Privacy and Compliance

➢ Why are they important – Impact on businesses such as legal issues, penalties, etc.

Organizations face pervasive risk of serious fines as mandated by the Consumer Data Privacy Regulations. GDPR enforcements resulted in **€**56 Million in first year alone, and by January 2020 it had reached **€**114 Million in fines. CCPA in US cost US **$**55 Million in enforcement fines.

According to one survey (Ref: FTI Consulting Report 2020), 75% of the surveyed companies had made budgetary allocations to implement the Data Compliance and Privacy processes. It is projected that 97% of companies surveyed, will increase their data privacy budget by 50% (Ref: FTI Consulting Report 2020).

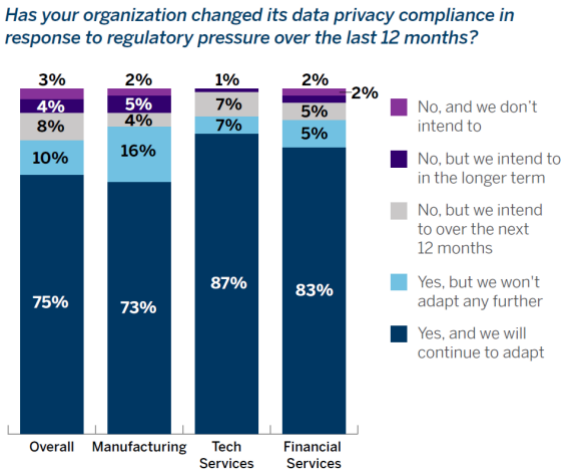
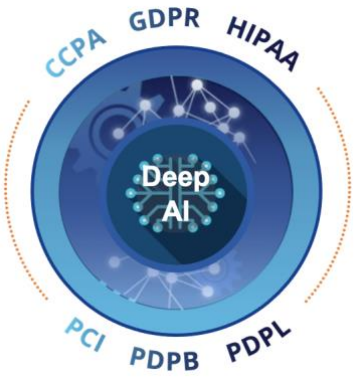


Fig-2: Adoption and Budgetary Allocations for Data Compliance and Privacy Mandates

Chorology.ai
KNOW YOUR DATA

2. Foundational Components of an AI Powered Intelligently Automated Data/Information Compliance and Posture Enforcement System

The most differentiating characteristic of a Data Compliance and Posture Enforcement System is its ability to 'Cost Effectively' implement the 'Comprehensive Set' of Data/Information Compliance Mandates via Intelligent Automation of Key Processes.



Data/Information Compliance and Posture Enforcement is a complex and daunting task that involves several complex processes and requires the ability to sieve through large volume of Data Corpuses (both on-prem and online) at a very high rate. A modern-day data compliance and privacy system, therefore, must embody a high level of ***"Intelligent Automation"*** for expeditious and error free performance of these complex tasks, while ***minimizing the cost*** of enforcement of these mandates.

As depicted in Fig-3, in order to achieve the ability to implement a comprehensive set of Data/Information Compliance and Privacy Mandates, it is imperative that the solution must incorporate the following foundational technologies and features;

1) Automated Data eDiscovery System
2) Automated Data Mapping System
3) Automated Data Service Request Handler
4) Large Volume Data Repository Connectors

Cohesive and ***seamless integration*** of these ***intelligently automated*** features is imperative inasmuch as manual dependency is prone to inadvertent human error and/or malicious circumvention of these processes. Furthermore, ***Intelligent Automation*** greatly reduces the Total Cost of Service (TCS) of the Consumer Data Compliance Service.

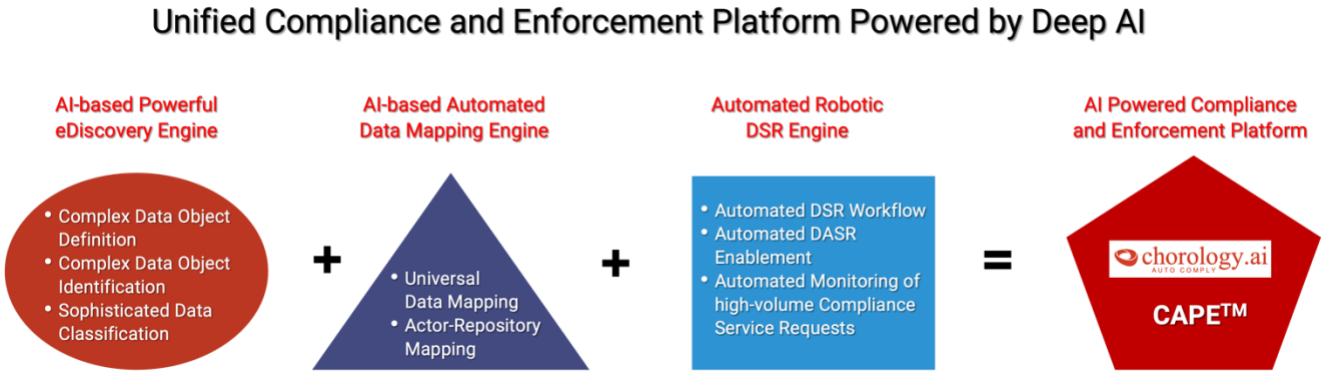


Fig-3: Key Components of a Modern Data/Information Compliance Enforcement Platform

1) Automated Data eDiscovery System

Foundational to any Data Compliance and Privacy Posture (especially, Consumer Data Compliance and Posture) enforcement is the ability to accurately ***Discover*** an increasingly more complex set of relevant data/information in a large corpus. To this effect, a more sophisticated eDiscovery Technology is essential that must provide for;

a. **Complex Data Object Types Definition –** The traditional eDiscovery technology has been confined to basic Lexical/Keyword match, Regular Expressions, Pattern Match, and Lexical Fingerprints. Due to the proliferation of a large number of Data/Information Compliance mandates (as more compliance mandates keep emerging frequently), a sophisticated eDiscovery System must provide the ability to define newer types of ***'Complex Data Objects'*** to support a wide variety of current and future Data Objects Discovery in an automated fashion.

b. **Automated Identification of Complex Data Objects –** The eDiscovery System must be able to perform Automatic Data Identification for virtually any type and modality of ***'Complex Data Objects'***. To this effect, sophisticated data identification techniques are imperative. Simple Keyword, Lexical Matches or Regular Expression based Pattern Matching techniques are NOT enough for automated identification of more ***'Complex Data Objects'*** as they have been proven to be error prone resulting in high False Positive rates.

c. **Automated Classification of Complex Data Objects –** The eDiscovery System must be able to recognize and Auto-Classify confidential and/or Compliance Mandated Data in any format without human intervention. Traditional Data Classification systems that require 'Manual Processing' are greatly ineffective as they are error-prone and unscalable.

Collectively, these advanced capabilities enable the AI Powered eDiscovery Engine as an ideal platform to provide high efficacy (low false-positive and zero false-negative) eDiscovery of Complex Data/Information Objects of any modality. This also results in a much lower TCO over the life cycle of the eDiscovery Process.

2) Automated Data Mapping System

Data Mapping, in the context of Data/Information Privacy, pertains to the process of creating an inventory of all relevant data/information that exists in an enterprise's corpus and mapping it out over the enterprise's data infrastructure. Automated Data Mapping System automatically creates a persistent map of the Data/Information Objects that exist in the enterprise corpuses. This is a crucial capability that greatly facilitates efficient navigation through large storage systems and corpuses following the ***'Lineage'***[1] of any given Data/Information Object of interest. Data/Information Object Map greatly facilitates the compliance enforcement process as it serves as a crucial input to the Compliance Enforcement Workflow generation process.

3) Automated Data Service Request Handler

Another foundational component of a modern-day Data Compliance and Posture Management (especially, Consumer Data Compliance and Posture) enforcement system is the ability to automatically handle the Data Service Requests in a timely and scalable fashion. To this effect, a sophisticated Data Service Request Handler must incorporate the following capabilities;

a. **Automatic Generation of DSR Workflow –** DSR Workflow creation is a critical and complex process that requires knowledge of (1) Data Map – distribution of Data Objects over the entire data corpus structure of the enterprise, (2) Accessibility Map – a jurisdiction layout for the ***'Actors'***[2] over the various Data Corpus and Repositories, and (3) Task Breakdown Structure – a deep knowledge of how a specific type of DSR can be broken down into a set of 'Primitive Tasks' required to complete the enforcement of a DSR.
b. **Automatic Enforcement of DSR Task Primitives –** For successful ***timely*** execution of a DSR, the system must be able to automatically implement all the constituent DSR Tasks within the prescribed time-frame. This requires ***'Intelligent Automation'*** of the DSR Task execution process. Traditional DSR Systems are typically limited to manual Intervention in the process which is not only tedious but also error prone.

[1] Lineage of a data object represents the point of its origin and any transformation during its flow through the corpus over time.

[2] Actors are Compliance Staff, IT Staff or Database Managers designated for the creation, maintenance or upkeep of a data object corpus, including databases and other data repositories.

3. Compliance and Posture Enforcer (CAPE™) – A Deep AI Powered Platform for Unified Data Discovery, Posture and Compliance Enforcement

Chorology's Compliance and Posture Enforcer (CAPE™) is a next generation Unified Data Compliance and Posture Enforcement solution that enables total enforcement of data/information security, privacy and compliance across all different types of compliance mandates (such as GDPR, CCPA, etc.).

As illustrated in Figure-4, the CAPE™ platform mitigates data/information security, privacy and compliance violation risks from a single centralized Compliance Administrator's console, using Chorology's patented advanced Deep Artificial Intelligence (AI) technologies.

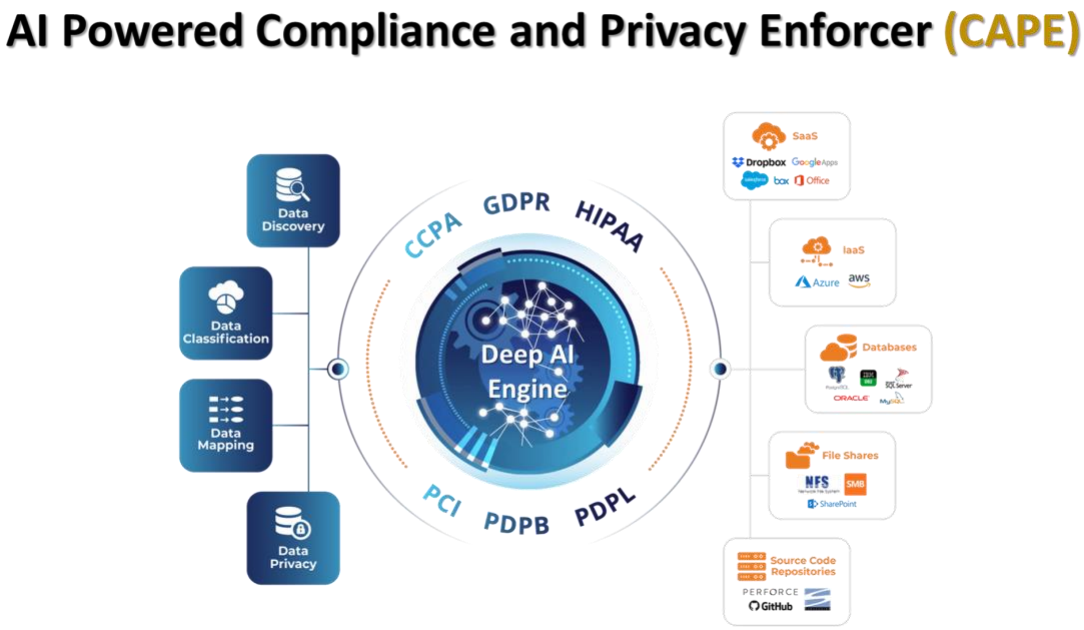


Fig-4: CAPE™ – Deep AI Platform for Full Featured Data /Information Posture and Compliance

1) AI Powered Data Discovery Engine

CAPE™ incorporates industry's most sophisticated Data Discovery Engine that has been built from ground up to address the deficiencies and constraints of previous generation Data Discovery solutions.

Previously, tools and methods were developed and deployed to enforce compliance measures such as SOX for corporate data compliance, PCI/PII for payment card industry vertical, HIPAA for healthcare industry, and several other mandates against theft and/or unauthorized disclosure of confidential

enterprise and individual data/information. With the emergence of new consumer data privacy compliance mandates (such as GDPR, CCPA, etc.) tools and processes are now required to enforce appropriate security and privacy measures against not only theft but also unauthorized disclosure or usage of confidential ***consumer's data/information***.

As depicted in Fig-5, the emergence of newer Data/Information Security and Compliance regulations mandates proper handling of increasingly more varied and complex data types and structures. These may involve simple Keywords (tags or labels) to complex Regular Expressions (RegEx.), as well as complex Composite Data Objects (comprised of more than one type of primitive data objects).

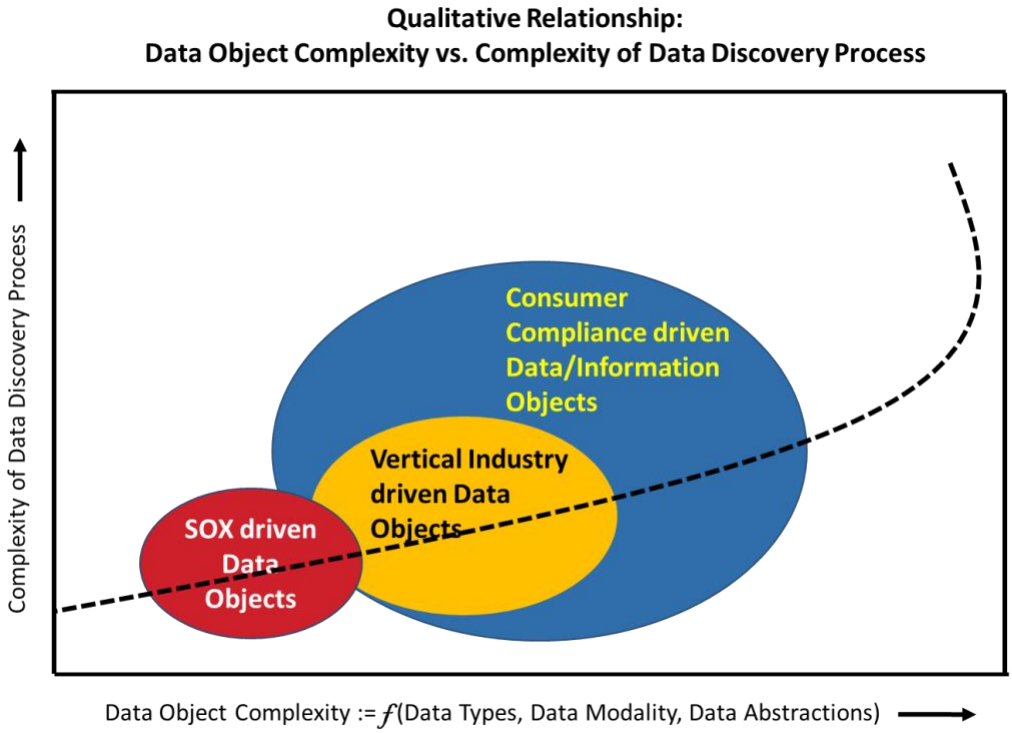


Fig-5: Complexity of Data /Information Objects as sanctioned by different Security, Privacy and Compliance Mandates

CAPE™ embodies unique patented technologies that greatly facilitate the definition and discovery process for complex Data/Information Objects embedded in large corpuses, namely;

- **Complex Data/Information Object Definition –** Based on sophisticated AI based paradigm, this unique patented technology enables definition of Complex Data/Information Object Types that can represent any type of concrete and abstract data or information objects of interest.

  Virtually any kind of data/information such as Structured, Unstructured, Semi-Structured or even Advanced Complex Data/Information Types such as Ordered / Unordered Set of Data and Data

Sequences can all be ***'modelled'*** and, Automatically Identified and Classified by CAPE's Automated Data/Information Discovery Engine.

This is a unique and powerful patented technology. It allows;

1. **BASIC Data Object Types** – Creation of new basic Canonical forms of Data Objects, representing very high granularity of individual data items.

   CAPE™ provides a default set of regular expressions and canonical classifications, used especially for PCI Compliance and PII (Personally Identifiable Information). Examples are credit card number formats, social security number formats, US zip codes, addresses, etc.
   Data Object definitions can further be refined. Regular expressions can be changed, added, used or not used for classification.

2. **COMPLEX Data Object Types** – Creation of very sophisticated Complex Data Objects by combining the Canonical Data Objects through,

   a. Unordered (non-sequenced) Data Aggregation – this is analogous to creation of SET of Data Objects that may comprise of any number of Canonical and/or Complex Data Objects.

   b. Ordered and Sequenced Data Aggregation – this is analogous to creation of ORDERED and SEQUENCED combination of Canonical and/or Complex Data Objects.

   This enables definition of complex structures that contain Structural constraints such as Sequence and Physical Relativity (such as Proximity or Distance) between the constituent Data Objects.

- **Automated Data Identification** – Chorology's CAPE™ incorporates sophisticated patented technology for Auto-Identification of high granularity Canonical as well as Complex Data Objects. Built on its AI powered Algorithms and constructs, Chorology's Data Identification Technology is deemed "BEST OF THE BREED" in the industry.

  The high efficacy of Chorology's data identification technology is of prime importance specially for Complex Data Types that involve components with cross modality, cross type, composite structured and unstructured, embedded or independent Canonical Data Types.

2) AI Powered Data Classification Engine

Foundational to any Consumer Data Privacy (or any Data Privacy) platform is the ability to accurately Classify the relevant Data in the corpus. To this effect superior Classification technology is of crucial import.

CAPE™ incorporates a unique Auto-Classification Engine that works with sophisticated *Data Object Ontologies* to Auto-Classify sensitive information. The Auto-Classification engine examines every Data/Information Object in the corpuses and using Chorology's patented algorithms automatically classifies it into one of the Classification Types defined in the Data Object Ontology. It can classify

sensitive information as granular as specific words and phrases. The Auto-Classification Engine completely replaces the requirement for manual tagging or fingerprinting of sensitive information. It can readily work out-of-the-box and does not require any pre-processing of data or a laborious training / learning process.

The ability to perform Automated Data Identification and Data Classification (regardless of its modality, i.e. Structured or Unstructured Data) results in the elimination of 'Manual Classification and Tagging' process hence enabling a more sophisticated and less error-prone Auto-Classification paradigm. This also eliminates the possibility of ***"Purposeful Misclassification"*** of data/content thus reducing the risk of ***"Malicious Acts"***.

3) AI Powered Automated Data Mapping Engine

CAPE™ incorporates a sophisticated Data Mapping Engine that automatically creates a persistent Universal Data Map (UDM) for the Data/Information Objects that exist in the enterprise corpuses. This is a crucial capability that greatly facilitates efficient navigation through large storage systems and corpuses following the ***'Lineage'*** of any given Data/Information Object of interest. The UDM created by the Data Mapping Engine is used effectively by the CAPE's Posture Enforcement Workflow Engine to automatically generate the DSR and DSAR service Workflow.

4) AI Powered Automated Privacy Request Enforcement Engine

CAPE™ incorporates a unique fully automated Role-based (a) Data Subject Access Request - DSAR, and (b) Data Subject Request – DSR, Enforcement Engine.

Using the patented AI Algorithms, the engine can automatically discretize the incoming DSAR or DSR jobs into corresponding sets of 'primitive' (or atomic) tasks. The 'primitive' tasks are then automatically 'serialized' into a Task Sequence using the Logical and Precedence Dependencies between these tasks.

Note – In any enterprise, an Actor is usually limited by authorization to certain Repository and types of data as per the Actor's Job Entitlement. Furthermore, Actors are also usually limited in the 'Operations' that they can perform. Different Operations such as 'Delete', 'Obfuscate', 'Mark', etc. are limited for each Actor. For example, an ERP or CRM Database Administrator can make changes to corresponding information, but may not have authorization to make changes to any other data repository content such as an HR Database. This is referred to as the principle of 'Segmentation of Duty (SoD)'. Segmentation of Duty policies use business logic to associate Data/Information Objects with the relevant Actors (i.e. custodians of the data/information objects).

CAPE™ incorporates an Actor Repository Map (ARM) that contains a correlation between Actors and the corresponding Repositories based upon the 'Segmentation of Duty' policies.

Using Sophisticated AI Algorithms, CAPE's Posture Enforcement Engine utilizes its Universal Data Map (UDM) and Actor Repository Map (ARM) to correlate Actors (i.e. custodians of data repositories), specific set of repositories over which a given Actor has jurisdiction/authorization and the corresponding Operations that they are authorized to perform on the specific data repositories. The result of the correlation is primitive Task Dispatch Policy for enforcement of the incoming DSRs and DSARs.

As shown in Fig-5, a detailed 'Workflow' is then automatically created which dispatches these 'primitive' tasks to the correspondingly relevant compliance or IT staff (or 'Actors') for fulfillment of the DSAR or DSR. The Task Dispatch is done automatically using policies based on elaborate Governance and Regulatory Compliance principles such as 'Segmentation of Duty (SoD)'.

Fig-5: CAPE – DSR/DSAR Enforcement Workflow

The 'Workflow' engine is equipped with a built-in mechanism to monitor and report the status of the DSAR or DSR fulfillment process, and raise Alerts, Alarms or other Notifications as appropriate during the fulfillment process.

Chorology.ai
KNOW YOUR DATA

4. Summary

Ever since the inception of the various Consumer Data Privacy Compliance Acts, the need for intelligently automated solutions for the enforcement of these mandates have become increasingly unequivocal. However, intelligent automation of several of the tasks involved in this process is quite challenging.

Traditional solutions seriously lack in their ability to enable intelligent automation of these tasks. A new set of technological innovations is essential to meet the objective of providing intelligent automation of these complex tasks in a scalable and error free manner,

Chorology's unique (patented and patent-pending) Deep AI based technologies enable a highly automated compliance and privacy enforcement platform. It eliminates the limitations in the traditional Data/Information Discovery, while providing superior accuracy, performance and maintainability.

Chorology's CAPE™ product, which has a highly scalable architecture, can be quickly deployed across an enterprise as well as in the multi-cloud environments. The ability to automatically identify content, automatically classify it, automatically generate privacy enforcement policies in real-time without requiring constant tedious manual intervention, and provide real-time enforcement of privacy mandates is unmatched.

The CAPE™ platform eliminates Manual Pre-processing Cost and enables protection against human error and malicious acts.

In summary, CAPE™ is a full stack Privacy Compliance Enforcement Platform that incorporates advanced features to deliver reliable efficacy in automatic performance of Data Discovery, Data Classification, Data Mapping and Consumer Privacy Compliance Enforcement tasks in real-life Consumer Data Posture Enforcement scenarios.

Chorology is the leader in the emerging next generation enterprise information security & access management market

Contact Us

For copies of this white paper, please email sales@chorology.ai
For further information please contact: info@chorology.ai
To schedule a demo submit a request at: https://www.chorology.ai/

Chorology, Inc.
2001 Gateway Place, Suite 710, West Tower,
San Jose, CA. 95110
Main: (408) 713-3303