Chorology.ai

Separating AI Hype from Reality in Data Security and Compliance

A Strategic AI Overview for Cybersecurity Practitioners

Table of Contents

1. Background: On the Status of Artificial Intelligence

Over the last several years, the hype around the fourth wave of AI and ML is reaching fever pitch[1]. The noise level of this AI wave reminds us of previous AI hype cycles which faded into history. Most of the ideas of the three previous waves in AI[2] seemed promising, but never achieved ROI fruition despite their grandiose, premature and hopeful claims funded with billions of investment dollars.

The current AI hype cycle[3] also seems destined for a similar lackluster ROI performance as AI fatigue is already surfacing across industries and C-Suites. As is evident by early market indicators and recent comments from across high-tech pundits, ROI success stories where machine learning/AI has generated any significant return on the excessive amount of investment is still hard to come by with some exceptions such as machine-learned image recognition. In certain industries such as medical imaging, specific high-value medical applications of AI/ML for pre-screening and early disease detection are commanding significant revenue and ROI.

Massive AI investments such as Open AI's Chat GPT are on track to lose $5B this year[4] with overwhelming enthusiasm from large language models (LLMs) for services that seemingly provide instant access to knowledge and answers. Recent investment conferences yield overhyped customer expectations such as replacing 30% to 50% of a workforce with AI[5]. Many practical LLM-enabled product features are useful such as auto-correcting grammar or auto-generating logos for marketing professionals, or auto-generating resumes or product descriptions for eBay sellers. However, prompting gimmicks such as text-to-image, text-to-software and text-to-video from Chat GPT and similar LLMs, as fascinating as the potential for these capabilities may seem, are still far from yielding impactful outcomes or Wall Street sized ROI.

Many large industry observers have already noticed the limitations of rudimentary LLM-based functionalities, both in terms of technical malfunctions (issues such as hallucinations, inability to reproduce results, etc..) and inability to facilitate "original creativity" (write an original novel) not to mention the exorbitant cost associated with creating and training such model capabilities. Based on the first three AI hype cycles, when the smoke screen clears, the current 4th AI/ML hype cycle will likely culminate into;

a. 80% wasted investment
b. 20% real progress in the ML/AI technology and its applications.

A large portion of this 20% of real progress will be attributed to either more refined machine learning algos to intelligently automate some of the important yet simplistic functions and processes or ML hardware enhancements. Any new advancement in foundational AI will be scarce to find, such as first principles-based constructs and innovations that can break through the obstacles in the way of more generalized AI paradigms.

[1] Hype Cycle for Artificial Intelligence, Gartner 2024
[2] AI Tech Will Arrive in Three Waves, Futurism, 051317
[3] Hype Cycle for Emerging Technologies, Gartner 2024
[4] Open AI Could be on Track to lose $5B this Year, Yahoo Finance, July 29, 2024
[5] Jason Lempkin, SaaStr Fall Silicon Valley Conference 2024

2. Purpose of this Paper

Welcome to “Separating AI Hype from Reality in Data Compliance and Security”, a strategic guide for data security professionals seeking to develop a more holistic view of AI and its application in today’s complex data security and governance environments. In this introductory AI data security whitepaper, we will dive into AI technologies and their use in the intelligent automation of data security, compliance, and governance processes and workflows.

We created this paper to help industry practitioners understand the real-world practicality of AI in Data Compliance and Governance applications beyond the hype of today’s business headlines. The 80/20 waste-to-progress rule cited in Section 1 also applies for the marketing opportunities from practical business uses of AI in the Data Security, Compliance and Governance industries.

Data breaches and security threats remain a serious problem for organizations worldwide today with losses due to cybercrime estimated to exceed $10T in 2025[6]. Intelligent automation in data security, compliance and governance is certainly the need of the hour in enterprise data security. In fact, our team at Chorology, Inc. believes the “compliance chasm” between compliance mandates and business practicality[7] is expanding (unfortunately), because fundamental data security problems and compliance challenges have not yet been solved. To date, no AI-powered “compliance engine” exists to intelligently automate the discovery, classification and risk assessment of structured and unstructured sensitive enterprise data.

With these facts in mind and led by its prolific founder Mr. Tarique Mustafa, Chorology set out several years ago to provide powerful data compliance, security and governance solutions that exceeded industry needs and expectations. Since our founding and emergence from stealth mode this August 2024, Chorology’s mission has been clear: to set a new standard for data integrity and regulatory compliance while future-proofing enterprises against new mandates. Chorology’s Deep-AI platform is a breakthrough capability in data security, global compliance and governance, providing intelligent automation and intuitive tools that help businesses stay ahead of security threats and evolving regulations.

To fully comprehend how different types of AI technologies provide value in data security, compliance and governance, (and which do not), it is useful to understand how AI solves data security problems, and which problems are best solved with which kinds of AI. By becoming conversant in the multiple forms of AI and their applicability in solving specific problems, data security, compliance and governance practitioners can understand the ”what’s” and “why’s” behind practical and valuable business uses of AI in protecting the sensitive data of their global enterprises and organizations.

3. Start With the Right Problem

A famous quote (often mistakenly attributed to the author of Alice in Wonderland Lewis Carroll), is “If you don’t know where you are going, any road will get you there”. This timeless wisdom has an analogy in the technology industry. *If you don’t know what problem you are solving, any technology will do.* In Chorology’s primary customer research, over 90% of CISOs, CIOs and CTOs interviewed agreed with the following statement:

---

[6] https://cybersecurityventures.com/cybercrime-damages-6-trillion-by-2021/
[7] Crossing the Compliance Chasm, Tarique Mustafa, 2024, All Rights Reserved

"We do not know what sensitive data we have or where it is located...
If we don't know our data, then how can we protect it or assess its risk?"

The depressingly simple answer to these strategically important questions is "You can't".

With first principles thinking, Mr. Mustafa and his Chorology R&D team realized that "knowing your data" was the first data security problem to solve with AI; the intelligent automation of discovering, classifying and mapping sensitive data - sprawled across the enterprise on-prem and in clouds - at industry-leading scale, speed and accuracy.

Given that fully automated data discovery is an extremely complex AI problem requiring deterministic knowledge representation, most of today's use of AI in data security and compliance platforms involves automating manual compliance functions and their respective workflow tasks such as accepting customer data subject access requests (DSARs) and routing tickets to IT professionals to manually investigate known types of data in structured repositories. This is automation at the top layer in black in Figure 1.

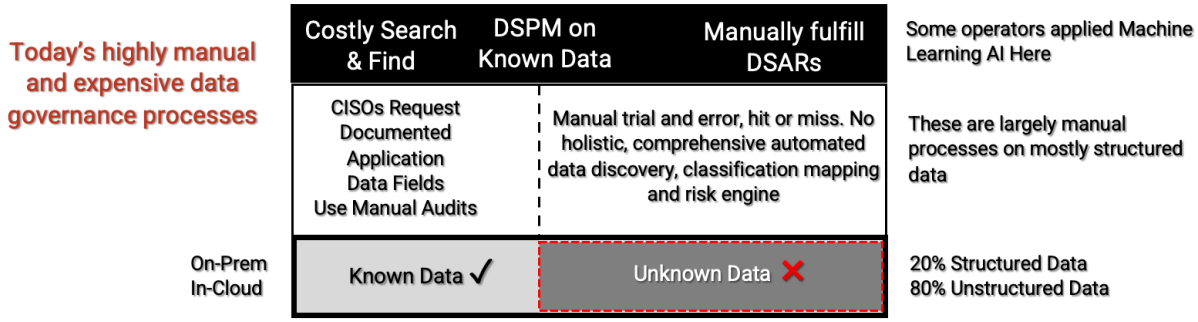

Figure 1: Data Security Product Stack

AI has been applied to the application layers above the data - to automate repetitive IT tasks such as searching, finding and tagging known data types and manually managing them according to enterprise security posture policies. Other application tools include assessing the risk of known data types to manage security posture scores (Data Security Posture Management) or processing consumers' inbound Data Access Subject Requests (DASRs).

But as we can see in Figure 1, if these semi-automated data governance platforms do not have the ability to comprehensively understand and classify structured and unstructured data that is "unknown", the upper-level applications will miss the risks associated with the data unknown to the enterprise.

In short, how do you assess the risk of sensitive data you don't even know you have?

If new platforms cannot discover and classify sensitive data in the 80% of enterprise data that is unstructured, (that is exponentially growing and sprawling), these platforms will miss value opportunities and (compliance and security) risks from unknown sensitive data in structured and unstructured data sprawled across global enterprise repositories.

Like the CISOs and CIOs stated, "*We can't secure or manage the security posture of data we don't know we have*. *We can't manage the risk of unknown sensitive enterprise data.*"

This unknown-unknown data challenge is even more important because 80% of today's enterprise data is now unstructured[8] (Think text, video, audio, web server logs, or social media activities). To effectively secure, manage and govern sensitive enterprise data, modern data governance platforms must include AI-powered engines that intelligently automate the discovery and classification of their sensitive data (both known AND unknown). Knowing how to auto-discover and classify sensitive enterprise data stored in structured and unstructured repositories, ensures data risk assessment of all sensitive data, on-prem and in-cloud. This was the foundational challenge posed by Chorology's primary research with CISOs represented in Figure 2 below.

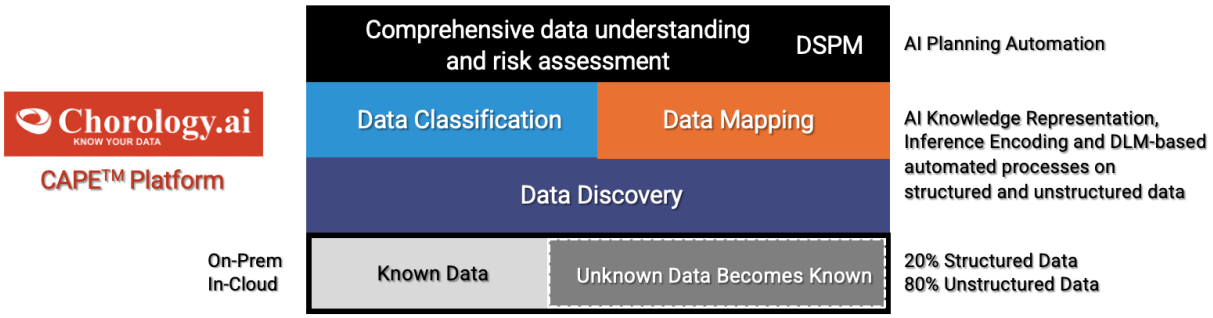

Figure 2: Manual Data Compliance Product Stack

To efficiently address data discovery and classification in both structure and unstructured data, sprawled across on-prem and cloud repositories at exponential scale, Tarique and his team took a fresh look at the data compliance and security technology problem using "first principles" thinking. Intelligently automating the data discovery, classification and mapping at global scale - while lowering the cost of ownership starts with solving the initial problem; using automation to confidently know what data an enterprise has and where it is located. But how does a machine scour thru corpuses with terabytes of structure and unstructured data in cloud and on-prem repositories - to discover different types of data, and then "know" (or classify) what data it finds, while finding it in real-time?

4. Not All AI Types are the Same

There have been three AI hype cycles[9] for emerging types of AI capabilities since the term "artificial intelligence" was coined in 1950 when Alan Turing published "Computer Machinery and Intelligence". In 1956, the first artificial intelligence program was presented at the Dartmouth Summer Research Project on Artificial Intelligence (DSRPAI) hosted by John McCarthy and Marvin Minsky[10] with the result that the birth of AI is generally attributed to this 1950-to-1956 timeframe.

In the time since the birth of AI and depending on which source reference is used, the consensus is that five to six core types of AI technologies exist:

1) Machine Learning
2) Deep Learning
3) Cognitive Learning
4) Neural Networks
5) Computer Vision
6) Natural Language Processing

[8] Statista, Taylor, 2023
[9] Gartner AI Hype Cycle, 2023
[10] History of AI, Wikipedia

These AI disciplines power the following “artificial” capabilities meant to replicate human abilities:

1) Learning
2) Reasoning
3) Problem-solving
4) Perception
5) Language understanding
6) Decision-Making or Planning

Some sources describe “decision-making, planning or explainable AI”, as closely related to “reasoning and problem-solving.”

Classical AI generally depends on rules-based systems for solving some clearly defined automation problems. However, due to inherent design limitations of rules-based systems, and limited rule complexity, volume, and computational performance, these systems cannot independently learn and therefore fail to scale. These early, overly simplistic systems struggle with learning complexity, instead relying on clearly defined, static informational rules that are highly customizable for smaller, individual domains.

AI Machine Learning involves algorithms that allow computers to learn from and make predictions or decisions based on training data and probability calculations. Unlike frame-based systems, ML doesn't rely on predefined structures but instead uses statistical methods to find patterns in data. Machine learning models are trained on large datasets with algorithmic statistical methods to recognize patterns and improve their performance over time. It can adapt and improve over time as more data becomes available.

Machine-learned models can be very powerful in the interpretation of images for people, places and things through image processing and recognition. For example, the use of AI/ML is helpful for recognizing people’s faces in large photo repositories or anatomical structures in medical images. Through the incorporation of neural networks, machine learning can evolve towards the accurate learning of people, places and things in different (static) contexts. It is important to note that neither Classical AI nor Machine Learning AI can learn continuously, nor can they generalize or make inferences. Both these disciplines of artificial intelligence depend on models created from large datasets and once deployed, can’t be changed without reprogramming or retraining these large models.

Deep Learning is closer to human-like intelligence as layered machine structures allow complex patterns from data to be modeled and understood. Deep learning has successfully addressed many commercial problems such as image classification when the image type is unknown ahead of time, and language translation of any verbal or textual corpus of words, phrases, sentences, and paragraphs.

Learning Continuously in a world that is constantly changing requires a model of the world that needs to update constantly as well. But most AI systems today do not learn continuously. There is typically a separate, often extensive “model training” process that must occur before machine-learned models are deployed. By way of comparison, our human brains learn new patterns by forming new synapses on dendrite branches. The new synapses don’t affect previously learned synapses, which is why humans are able to continuously learn new things, without forgetting prior things learned.

Learning via Movement in which the system learns by moving. Simple examples including learning to ride a bike to get “balance” or learning to brake by getting a “feel” of brake pressure vs. car stopping distance. We must explore these tasks because we cannot learn them by studying a set of rules.

Recent public examples include Google's Deep Mind robots on 60 Minutes[11] that learned how to play soccer over time via practice trial and error.

Learning Models That Generalize like how human brains respond quickly and precisely to previously unseen inputs by extrapolating from knowledge already learned and stored. Humans possess the ability to generalize and extrapolate with high confidence and in the correct context. Accurate responses in proper context or domain have been a known failing of AI systems for many years.

Learning Using Structured Representation (Reference Frames) Human brains store knowledge in "reference frames." Reference frames are like brain maps that humans use to understand how things relate to one another, and how specific kinds of knowledge relate to other kinds of knowledge. Our human brains use reference frames to make predictions, create plans, and perform movements. When human brains are thinking, our brains are moving through these reference frames, retrieving information associated at each location on the "brain map".

Frame-based computer learning relies on predefined structures and rules to represent knowledge and is more static, depending heavily on human expertise to design the frames upfront to provide proper context for distinguishing between concepts. Frame-based learning requires detailed knowledge engineering to set up the frames and rules and works well in domains where the knowledge is well-understood and can be explicitly defined.

5. Knowledge is Power – the Power of Certainty

With an accomplished background in all types of AI, Chorology's team led by CEO/CTO Tarique Mustafa, astutely recognized that to intelligently automate data compliance and posture enforcement, one needed to first intelligently automate data discovery, classification and mapping when data types within enterprise repositories are unknown in advance of data scanning. As depicted in Figure 2 above these automated data capabilities are foundational for automating higher-level data compliance and governance applications. As was already stated above, most of today's enterprise data is unstructured and sprawled across many on-prem and in-cloud data repositories that store unknown data types in texts, e-mails, audio, web server logs and social media content and other forms of unstructured content.

By knowing the multiple types of artificial intelligence, and how they are applied for different kinds of computer-based learning, Tarique and his team were able to structure the foundational data security problems of auto data discovery and classification by using a knowledge abstraction discipline known as "knowledge objects".

Knowledge objects are an abstraction layer above the data - used to represent any kind of data, from text to number expressions to documents or media files. The use of these knowledge objects gives rise to intelligent automation by individual or combined types of AI by abstracting data types and combinations of data types into AI-compatible knowledge objects that are in the proper form useful for one or more disciplines of AI. Using non-machine-learning disciplines of AI, the Chorology team holistically abstracted the very important and unsolved unknown data governance problem by applying these "knowledge objects" to represent any type of data, or combination of data types - in

[11] Robotic Soccer Players, 60 Minutes 2024

any format during data discovery, classification and mapping.

By envisioning how to algorithmically discover and classify data using structuring knowledge objects, a practitioner is then able to apply an extensive arsenal of AI expertise to invent, develop and commercialize ground-breaking solutions for other downstream data security, compliance and governance applications and services such as data risk assessment, security posture management and very hard and complex enterprise data tasks such as automatically executing inbound consumer Data Subject Access Requests (DSARs).

Figure 3: Core AI-Powered Data Compliance Capabilities

Framing the Problem Correctly to Create a Paradigm Break-through

Tarique had the brilliant insight that if Chorology's patented technologies could reframe enterprise data and documents as "knowledge data objects" – these knowledge objects would lend themselves to being automatically processed or "operated upon" by powerful AI capabilities across multiple AI disciplines. Based on these insights and data reframing, Tarique was able to intelligently automate critical data capabilities for solving multiple enterprise data security, compliance, and governance problems today, in a way that also future-proofs enterprises for tomorrow's coming mandates.

As a side note, the company received its name from the "structuring of enterprise data and knowledge". The noun "chorology" is defined as the study of the causal relations between geographical phenomena occurring within a particular region[12]. In short, "chorology" is the study of the structure of locality. Chorology, Inc, studies the (local) structures of data.

6. Enterprise Challenges for Data Security and Compliance

In today's fast-evolving digital world, where data flows freely and information is currency, enterprise data compliance and governance professionals face a multitude of operational challenges from a rapidly changing data landscape involving multiple contexts and regulatory mandates.

By their evolving nature, regulatory mandates are constantly changing and expanding. For example, 137 out of 194 countries now have legislation in place to secure data and protect consumer privacy[13]. This amounts to over 162 national laws as of March 2023 with at least twenty countries that have bills underway for new privacy laws because these countries lack data privacy laws.[14] Besides these

[12] https://en.wikipedia.org/wiki/Chorology
[13] https://unctad.org/page/data-protection-and-privacy-legislation-worldwide
[14] UNSW Sydney, May 2023

current laws, new forms of regulations are fast appearing on the horizon - and for good reason.

For example, AI machine learning innovations such as today's large language models (LLM's), show tremendous potential for transforming enterprise processes. Applications of Generative AI and LLM's with over four billion parameters modeled, rely heavily on the secure exchange of data at rest and enroute. Industry titans like Nvidia's CEO Jensen Hwang foresee "100's of AI's" being routinely used to automate processes between functional areas across organizations and enterprises and with their demand chain and supply chain partners. Securing data for these AI innovations is a massive challenge for today's AI industry. Protecting enterprise data as it is being ingested and used by these LLMs is an equally daunting, albeit valuation-critical challenge.

Most enterprises are struggling to understand how their data is interacting with LLM models, how enterprise data behaves with these new models, and the associated security risks. In addition, enterprises want to learn how to protect, secure and govern LLM's use of enterprise data. Probability-based LLMs are also incorrectly mixing data contexts in LLMs such as "Husband" and "Ex-Husband" or Customer ID vs. Social Security Number which can yield hallucinations. In imaging – AI is not able to distinguish a statue of a man on a horse from a real man on a horse. This is a problem of incorrectly confusing contexts (or "domains") of enterprise data - which in turn creates incorrect data meanings embedded in the data structures of structured and unstructured content ("Mustang" a horse vs. "Mustang" a Car with 350 horsepower or "Spouse" vs. "Ex-Spouse").

7. Challenging Volumes: A Tsunami of Enterprise Data

On top of the challenges of correctly identifying and classifying data across multiple contexts or domains, across expanding regulations, and the emergence of new digital AI technologies requiring even more regulation - enterprises are also facing significant challenges from the exponential rise in data volume and data sprawl.

Worldwide enterprise data volume is forecasted to rise to 181 Zettabytes by 2025.[15] (That is 181 followed by 21 zeros). With 50% to 90% of today's enterprise data unstructured[16], the problem of discovering and accurately classifying enterprise data at scale is becoming exponentially more complex. Today's data professionals are seeing data volumes grow by an average of 63% every month in their companies - with nearly six in ten data organizations saying they can't keep up.[17]

The oversimplification of describing "data security" as a "data challenge" given today's exponential rise in enterprise data volume and sprawl - is akin to saying that during a tsunami, one might experience moisture. To extend this ocean analogy even further, most enterprises are "drowning" in a sea of sensitive data while analysts, pundits and many data security vendors "describe the water."

By applying first-principles thinking, Chorology realized that to close the ever-widening compliance chasm and give back control to overwhelmed enterprise data practitioners - entirely new thinking, platform tools and practices, powered by smartly implemented AI technologies were required.

8. The Challenge of Data Discovery

The distinction between "data searching" and "data discovery" is critical to understanding the core

[15] ITC, IDC, Gartner
[16] ESOMAR's Global Market Research 2022
[17] Dataversity, 8/14/2023

value offered by Chorology's Deep-AI engine which lets an enterprise comprehensively "know their data" on-prem and across enterprise clouds to assess its security risk and posture.

In legacy data security and compliance tools and processes, the IT professional knows the data for which they are searching because they receive IT customer data tickets to fulfill in automated data request workflow platforms. IT personnel also know the data table structures of the enterprise and use their table headings to search for data by the data types. The compliance or IT personnel know these data tables usually because CISOs and CIOs request the data structures and tables specified in the PRDs of product managers – to comprehensively understand and audit - the sensitive data types that are designed into their system architectures and are being stored in their enterprise systems.

E.g. A data table created with the header of "Social Security Number" allows the IT personnel to know the data type and values stored in a table and to expect a specific data value structure ("xxx-xx-xxxx" in the case of SSN) before the IT personnel execute manual searches or run automated search scripts. The data types and their values are considered "known" ahead of the compliance data search process. In effect - IT and compliance personnel answer the question, "How can I ensure our sensitive data complies with legal mandates?" This results in data security postures based on the data an enterprise already knows it has.

As stated at the beginning of this paper, the data security problem for which Chorology designed its AI-powered solutions answers a different set of questions: "What sensitive data do I have and where? What is the total security posture risk based on all the sensitive data we store?" This is a data discovery problem, not just a data search problem as illustrated in Figure 4 below.

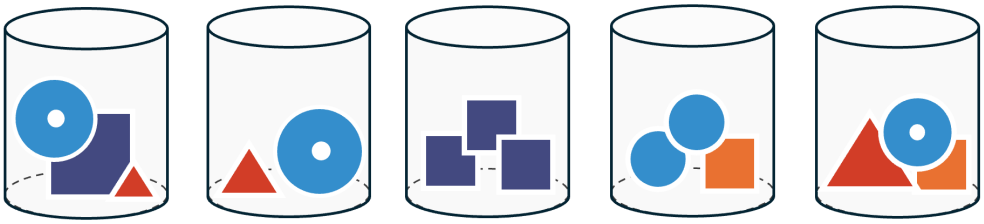

Figure 4 – The Data Discovery Problem

In this (more difficult) data discovery problem, scanning algorithms are scanning data repositories – (80% of which are unstructured and no data tables exist) that contain unknown data and data types. In the discovery problem the process is to conduct a complete audit of enterprise data first. The Chorology platform opens the data repositories, proprietary algorithms scan the data inside these repositories, while identifying and classifying the stored data in real-time - without knowing the data in advance.

Chorology is performing an audit of both known and unknown data sprawled across the enterprise or organization, stored in both structured and unstructured data repositories. "Data Discovery" means personnel are discovering both the data and the data type that exists in repositories without knowing what type of data is stored ahead of time, and without the benefit of knowing the structured data header fields of tables designed for the (structured) repositories. With Deep-AI-powered, intelligently automated data discovery and classification capabilities, the platform discovers and classifies both known and unknown data in real-time during the data scanning process across every repository in the scan, at ultra high-speed, accuracy and scale. In short, Chorology can discover and classify

"unknown-unknown" sensitive enterprise data.

After Tarique Mustafa recognized the real problem to solve were the foundational capabilities of data discovery, classification and mapping essential to "knowing your data" - he selected the appropriate AI disciplines and developed approaches using novel algorithms that required no machine learning, pre-processing overheads or software devices that persistently reside locally on enterprise repositories. These AI disciplines required the fields of knowledge representation, reasoning, inference calculus, planning and domain language modeling to intelligently automate data discovery, classification and mapping. Data governance applications like data security posture management (DSPM) or "record-linking" are then built on top of these core Chorology platform capabilities.

9. Why Machine Learning is Not the Right AI for Automating Data Security

As a reminder, the AI disciplines of machine learning and frame-based knowledge representation serve different purposes and are suited to different types of problems. Machine learning shines in data-rich, dynamic environments where discovering hidden patterns is essential, while frame-based systems excel in well-defined, structured knowledge areas. Furthermore, machine learning is a probabilistic-based AI methodology while frame-based knowledge representation is deterministic. Implementing the right AI discipline or combination of AI disciplines and technologies requires consideration of the true nature of the data security and governance problems being solved.

Through first principles reasoning and the help of CISOs and CIOs, the Chorology team quickly recognized that the AI discipline of machine learning - was not the right AI for solving the primary set of data problems Chorology's team identified; **Knowing what sensitive data an enterprise has and where.** If an enterprise does not know what sensitive data it has, then how can its personnel secure it, or assess its risk, or calculate its contribution to an enterprise data security posture score?

It is intuitively obvious that if one is solving the problem of data discovery as described in section 6 above (identifying and classifying both known and unknown data stored and sprawled across the enterprise), then using machine learning to "learn" the data being stored, is a very ineffective and inefficient way to scan repositories. One does not need to learn what an AMEX or VISA credit card number looks like – these are known data objects "determined" by formulas, and when combined with an expiration date and PIN – are well characterized knowledge objects containing known regular data expressions. Similarly, one does not need to learn through probabilistic methods - what a file containing salary data looks like, or a medical form, or learn what a sensitive health record looks like or personally identifiable information (PII) like a social security number. Again, these are known knowledge objects that contain known types of data expressions.

In addition, each of the domains of these knowledge objects are also essential for accurate classification. For example, a patient's medical form could contain medical data, HIPAA-regulated medical diagnoses, PII (the patient's identity information) and PCI data (their payment information). Each regulatory "domain" must be properly classified during the real-time data scanning.

In addition, if one is scanning terabytes of data, to identify and classify data discovered, it is not valuable to know a digital number discovered is an AMEX number with 87% confidence. One must definitively know that the discovered number is an AMEX number, so that it can be classified with proper domain-specific context ("credit card number" and "mandate = PCI compliance").

Deterministic models with frame-based knowledge representation use ontologies for accurate and

precise data discovery, while probabilistic machine learning models use statistical models and machine learning to estimate the likelihood of each data type in relation to other data types within the document. In a digital age, for the purposes of data security, compliance, and governance, an enterprise's market capitalization can't be based on "estimating" whether stored sensitive data is compliant with legal mandates. The CISOs and CIOs must accurately KNOW the sensitive data they have, and KNOW it meets (or doesn't meet) global regulatory mandates. Otherwise, a single regulatory mandate violation could result in 100's of millions of dollars reduction in the enterprise's market capitalization after the enterprise's lack of compliance is fined.

It then logically follows that the fields of data security, compliance and governance require precise, fast and scalable data discovery and classification to generate accurate data risk assessments. The conclusion then immediately follows that the proper AI solution should be based on the core AI discipline of frame-based structured knowledge representation, NOT machine learning.

10. Summary

In this paper we summarized the core AI disciplines and technologies to provide a proper context for the 4th AI hype cycle that exists today. With all of the AI disciplines and technologies as context, we identified the core AI data security, compliance and governance problems (accurate, ultra-fast, scalable data discovery, classification and mapping) – the solving of which opens a new paradigm and future possibilities for a fully automated data compliance and governance industry. We identified the key challenges happening in the marketplace including exponential data volume and sprawl – driving the need to rethink the entire data security and governance "stack". We distinguished between the foundational compliance, governance capabilities of searching for known data in structured repositories vs. discovering unknown data in unstructured repositories and how the latter is the core need now that 80% of today's enterprise data is unstructured and growing exponentially. Based on the needs of today's data security marketplace, and the need for solutions that future-proof the enterprise for tomorrow's mandates, we then showed why a machine-learning, probabilistic-based AI technology platform framework adds more risk to the enterprise vs. a platform based on the deterministic AI disciplines of knowledge representation, inference calculus and planning that eliminates data risk through its precision, speed and scale.