



Chorology is Solving DSPM Issues with DLM



Tom Hollingsworth ✓
The Networking Nerd & Event Lead at Tech Field Day



January 23, 2026

The industry has spent years trying to fix the data discovery problem with tools that don't have an engine. Companies spend millions on compliance platforms only to realize they bought a fancy workflow tool that can't find their data. You end up needing third-party discovery or legacy DLP engines just to get a baseline. It's a mess. Organizations still don't know what sensitive data they have or where it lives, especially when 80% of that data is unstructured dark data.

Chorology, which grew out of the team at GC Cybersecurity, is taking a different swing at this. They've launched an automated data intelligence platform that skips the usual machine learning hype in favor of something more reliable. They're using a Domain Language Model, or DLM, built on classical Deep AI principles like knowledge representation and inference. It's a move away from the probabilistic guessing that defines modern Large Language Models and their smaller, shrunken cousins. I sat down with CEO and Co-founder Tarique Mustafa to get an overview of the approach that Chorology is taking.

The Problem With Probabilistic Guessing

Most people today are obsessed with LLMs and SLMs. There is a fundamental flaw in using these models for data discovery. Machine learning is epistemic. It deals in beliefs and probabilities. When you're trying to identify a social security number or a specific bank loan document in a sea of unstructured text, "probably" isn't good enough. Probabilistic models lead to false positives and hallucinations. They guess and they get it wrong.

Data discovery is an ontological problem. It is about the nature of existence and identity. Either a piece of data is a sensitive record, or it isn't. Chorology's DLM uses knowledge representation to define data precisely. Instead of training a model on billions of parameters and

hoping it recognizes a pattern, they encode the actual knowledge of what a data object is.

There is a trend right now of taking an LLM and "shrinking" it down to create a Small Language Model for enterprise use. Tarique has a great way of describing this. He says shrinking an LLM is like chopping off an elephant's trunk, legs, and tail to try and make a small zebra. You end up with a broken animal. When you prune a neural network, you create holes in the knowledge representation. You break the connectivity of the graph.

The DLM approach is different because it's a canonical representation of the foundational model. It doesn't have those holes. It allows for the creation of specific domain languages for a company or a vertical without losing structural integrity. You get a model that is purpose-built, not a mutilated version of a general-purpose chat bot.

Composite Objects and the Planning Engine

Standard discovery tools look for simple metadata or file extensions. That's why they miss the unstructured data. Chorology uses composite objects. You can define a "Bank Loan Document" by telling the system to look for a social security number, an email, and an account number all within 50 characters of an IBAN. This logical proximity and context are what allow the system to illuminate dark data that traditional regex or metadata scans would sail right past.

Once the data is found, it's mapped into a Universal Data Map. This is a persistent record of every sensitive data point across on-prem and cloud repositories. Sitting on top of this map is a patented planning engine. Unlike standard workflows that require human intervention, this engine uses automated planning to calculate the exact sequence of steps needed to execute a Data Subject Request. Whether it's a Right to Access or a Right to Delete under CCPA or GDPR, the engine handles it autonomously. It cuts the error rate to basically zero because it's acting on a verified map rather than a probabilistic guess.

Speed and Total Cost of Ownership

We also need to talk about the hardware. LLMs and SLMs are resource hogs. They demand massive GPU clusters and enough electricity to power a small city. Chorology's platform runs blazingly fast in software on standard CPUs. It doesn't need specialized chips to function. This lowers the total cost of ownership significantly. Chorology is even talking about putting these algorithms directly onto a chip. When you aren't constantly retraining a neural network, your energy footprint and your budget both look a lot better.

Bringing IT All Together

We've reached a point where the industry is trying to solve every problem with the same hammer, and that hammer is probabilistic machine learning. It's the wrong tool for DSPM. For data discovery and compliance, you need precision and determinism. You need a system that understands the ontology of your data, not one that predicts the next likely token in a sentence.

Chorology's move toward Deep AI and DLMs is a necessary course correction. By replacing human-heavy workflows with a planning engine

and replacing guesswork AI with knowledge representation, they are actually solving the compliance gap instead of just throwing more compute at it. It is a smarter, leaner way to handle the data sprawl that is currently drowning most enterprise security teams.

If you want to learn more about Chorology and their approach to DSPM, make sure you check out their website at <https://Chorology.ai>