

# AI-Driven Analytics for Data Discovery

As organizations transform and push towards operational excellence, new strategies are needed that scale to meet modern data discovery workloads, contain costs, and mitigate risk while extending visibility into the value of corporate data.

## Voltage File Analysis Suite at a Glance

### Detect

Leverage AI-driven analytics to identify sensitive data, powered by IDOL

### Protect

Highly sensitive data, consumer trust and corporate reputation

### Evolve

To meet the growing data discovery workloads and drive operational excellence

## Understanding the Value of Data

With the continued growth of unstructured data comes the increased risk for understanding data and protecting sensitive and private information. The sheer volumes associated with these challenges make manual review nearly impossible. As a result, technologies like deep content analysis and artificial intelligence, can play a major role in helping organizations wrangle their data, understand its value and most importantly take protective actions on the high-value, highly sensitive and personal data your organization uses and has in its possession.

## Eliminating and Minimizing Data

As organizations transform IT, they strive towards operational excellence—lowering costs associated with data, mitigating risk and finding ways to gain a competitive advantage. These projects start with understanding the value of data. Data and application sprawl have constantly led to a bloat of redundant data (duplication, and convenience copies of data spread out across the enterprise and in cloud repositories). The same is true for out of date, stale data (data that has not been accessed for extended periods of time) or low-value data that simply takes up storage space and resources to manage (data like vacation photos or .DLL or .ZIP files). Voltage File Analysis Suite (FAS) by OpenText™ quickly identifies redundant, out of date data and trivial data through metadata delivering quick wins and early internal rate of return on these transformation projects.

## AI-Driven Detection of Sensitive Data

Determining the importance and value of what remains behind is critical to mitigating risk and operationalizing privacy and data insight. Voltage FAS leverages AI-driven rule sets and grammars out of the box to describe sensitive data entities that need to be identified and protected.

Our core grammar sets focus on:

- **PII**—Personally Identifiable Information, including 13 categories of entities across 38 different countries
- **PHI**—Protected Health Information, normally associated with the North American health industry
- **PCI**—Payment Card Industry data such as credit card and primary account numbers
- **PSI**—Personal security information, for account details access keys

Context is king and accuracy is essential. Powered by OpenText™ IDOL analytics, Voltage FAS provides curated, and optimized grammars built to assist today's global data privacy challenges—supporting entity extraction for 38+ country languages, data formats and economic regions.

- Our grammars also use context and “landmarks” used to generate a risk score we use to deliver more accurate results and filter out false positives
- Accuracy and risk score is based on proximity to the identified entity extracted or matched combined with natural language processing (NLP) techniques are used to determine the strength of the relationship and augment the score

- Extended language support in grammars also helps with context in determining the sensitivity or even presence of PII
- Multiple grammars can be combined to target entities more broadly but can result in higher compute costs and slower processing times
- Customizable risk scoring and weighting of categories provides greater flexibility in how you choose to implement and configure grammars to meet your needs

In many organizations' custom IDs or specific identifiers that are considered sensitive are used regularly. In these instances, Voltage FAS supports custom grammar configurations that allow you to identify these entities without the need for complex programming. Custom grammars can be defined using format descriptive regular expressions (RegX) or simple term lists, for example.

In addition, today's data discovery projects need to support modern workloads at petabyte scale. As a result, leveraging technology to assist and prioritize future phases and use cases is becoming more and more important in tackling these challenges. Voltage FAS is 100% cloud-ready—including public or private clouds. Built on a modern, cloud-first, microservices platform Voltage FAS can scale up and down as needed and can randomly sample data sets to indicate the presence of sensitive data, improving the accuracy and prioritization of future data discovery phases.

### AI-Driven Data Discovery: Benefits of Contextual Awareness

Once you have a better understanding of the value of your data Voltage File Analysis Suite by OpenText™ will help:

- Increase accuracy and detection of sensitive and high-value data
- Increase efficiency of privacy and data protection activities

- Reduce false positives
- Configure risk scoring and weighting provides added flexibility to tune

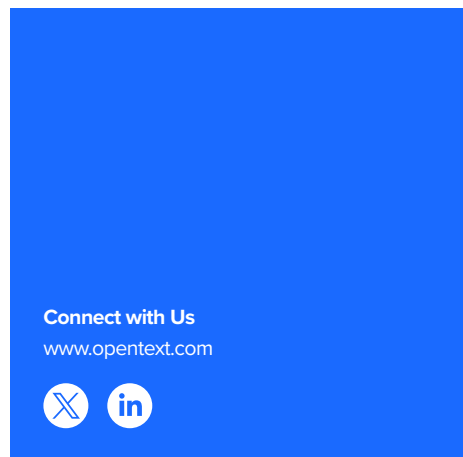
For example, with pattern matching a very simple format like a date can potentially produce a large volume of false positives. The entity and IDOL's deep understanding of the context around the entity can more easily decipher a date of birth from a date on an email, allowing for great confidence is what risk score is applied to that entity. Voltage FAS can also help easily identify entities split across columns or lines in a document using landmarks and parent entities. For example, first\_name/ last\_name combinations provide a strong confidence factor, compared to analysis that looks for separate entities.

Operational excellence is achievable—and Voltage FAS's AI-driven content analysis, powered by IDOL, is a critical tool in helping lower costs, mitigate risk and gain competitive advantage based on key AI-driven insight. The complexity of detecting and protecting sensitive data and PII require an evolving set of techniques that enable your organization to protect the crown jewels, securely share sensitive data with 3rd parties and comply with expanding global regulations.

### Our Solution for AI-Driven Analytics for Data Discovery Includes:

#### Voltage File Analysis Suite

Voltage FAS SaaS by OpenText™ is a file analysis solution enables organizations to quickly and efficiently reduce information risk, ensure data privacy, analyze, optimize and secure employee access to critical data that drive and protect the business. Voltage FAS ensures data protection and data preservation while mitigating the risk associated with managing sensitive data. Voltage FAS Suite drives operational excellence initiatives improving data visibility, data minimization efforts, data privacy readiness, and data protection while addressing long-term



preservation for high-value data (e.g., contracts, intellectual property, patents, etc.) and sensitive data (e.g., PI/ PII, PCI, PHI, etc.).

#### IDOL

OpenText™ IDOL is AI for unstructured data analytics, enterprise search, chatbots, and government open-source intelligence. It unifies text analytics, speech analytics, and video analytics with support for over 1000 data formats. It enables out-of-the-box access to 150 data repositories and indexes data without relocation, and disruption. Built on proven AI technologies such as NLP (Natural Language Processing), machine learning and deep neural networks, IDOL reveals trends, patterns, and relationships hidden in massive volume of unstructured data while its enterprise-grade security preserves and stays up to date on security entitlements without compromising performance by using strong data access protection.

Learn more at  
[www.microfocus.com/datadiscovery](http://www.microfocus.com/datadiscovery)  
[www.microfocus.com/IDOL](http://www.microfocus.com/IDOL)  
 (IDOL | Unified Text, Video & Speech Analytics | OpenText)