# HaystackID® Discovery Intelligence

Synergistically Harnessing Artificial Intelligence, Data Science, Machine Learning and the Human Touch to Power Next-Generation Discovery for Today's Challenges and Opportunities
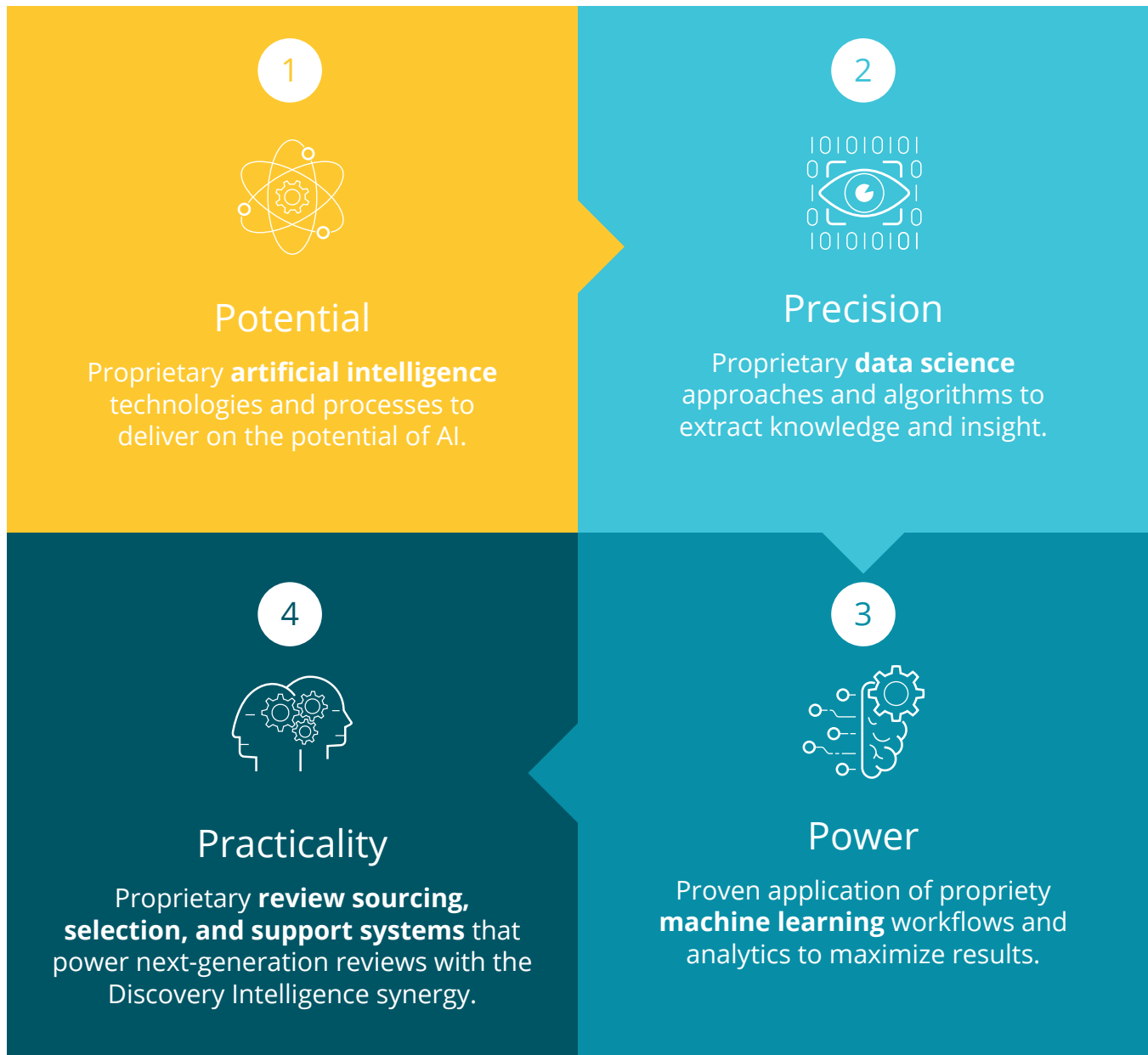
HAYSTACK®

# The Right Time is Now
# for Discovery Intelligence
# from HaystackID

HaystackID Discovery Intelligence (HDI) transforms how cybersecurity, information governance, and eDiscovery professionals consider the practice and process of discovering electronically stored information. By synergistically harnessing the potential of artificial intelligence, the precision of data science, the power of machine learning, and the practicality of expertly trained and managed reviewers, HDI delivers insight and intelligence that allows you to reach decision points faster and more economically than previously possible. The combined benefit of proprietary technology and processes used with optimized and tuned multimodal workflows for specific discovery objectives provides you with secure, defensible, and flexible capabilities to address the most complex and time-sensitive discovery challenges.

# The Four Synergistic Elements
# of Discovery Intelligence

### 1

## Potential

Proprietary **artificial intelligence** technologies and processes to deliver on the potential of AI.

### 2

## Precision

Proprietary **data science** approaches and algorithms to extract knowledge and insight.

### 4

## Practicality

Proprietary **review sourcing, selection, and support systems** that power next-generation reviews with the Discovery Intelligence synergy.

### 3

## Power

Proven application of propriety **machine learning** workflows and analytics to maximize results.

Whether supporting proactive exploratory analysis to gain insight into information early or concentrating on specific discovery objectives to drive intelligent decisions in audits, investigations, and litigation, the potential, precision, power, and practicality of HaystackID Discovery Intelligence provide you with next-generation discovery for today's challenges and opportunities.

# The Potential of Artificial Intelligence

According to the Electronic Discovery Reference Model (EDRM) [1], artificial intelligence (AI) refers to the capability of machines to mimic aspects of human intelligence such as problem-solving, reasoning, discovering meaning, generalizing, predicting, or learning from experience. AI includes both unsupervised and supervised machine learning and has many other processes, such as natural language processing (NLP). In the context of HaystackID Discovery Intelligence, AI describes the automated discovery processes used to classify, categorize, summarize, predict, and provide insight and intelligence regarding ESI. This application of AI by HaystackID is accomplished through the expert application of proprietary statistical, rule-based, and algorithmic means. The potential of AI is realized in HaystackID Discovery Intelligence in areas ranging from detecting sensitive data (e.g., from PII and PHI to SSNs and License Plate Numbers) and identifying entities to supporting cyber breach responses as part of ReviewRight® Protect™ services to evaluating and classifying legal document reviewers as part of HaystackID ReviewRight Match AI™ services. Three examples of AI-enabled services available as part of HaystackID Discovery Intelligence include:

**Cyber Discovery**

## ReviewRight® Protect™

ReviewRight Protect™ is a combination of advanced data detection technologies and processes, extensive legal and regulatory compliance expertise, and proven notification and reporting procedures that harness the power of the world's leading legal discovery and review services and orients them directly on the detection, identification, review, and notification of sensitive data-related breaches and disclosures.

Download ReviewRight Protect™ Fact Sheet [PDF]

**Enterprise Solutions**

# Protect Analytics™

Protect Analytics™ is an exclusive set of technologies and processes that allow client data set analysis for sensitive information ranging from PII and PHI to data breach code anomalies. Enabled by a collection of proprietary workflows and proven tools, Protect Analytics can proactively or reactively help determine sensitive data concentrations, locations, and relationships to inform notification lists, exposure assessments, and discovery targeting.

**Legal Discovery**

# ReviewRight Match AI™

ReviewRight Match AI™ is the industry's most advanced document reviewer ranking and selection technology powered by cutting-edge artificial intelligence. With over 12 years of review data, including millions of review hours and tens of millions of coded documents, our proprietary learning algorithm leverages the latest AI technology to evaluate over 25,000 tested, qualified, certified, and continuously graded candidates from around the world.

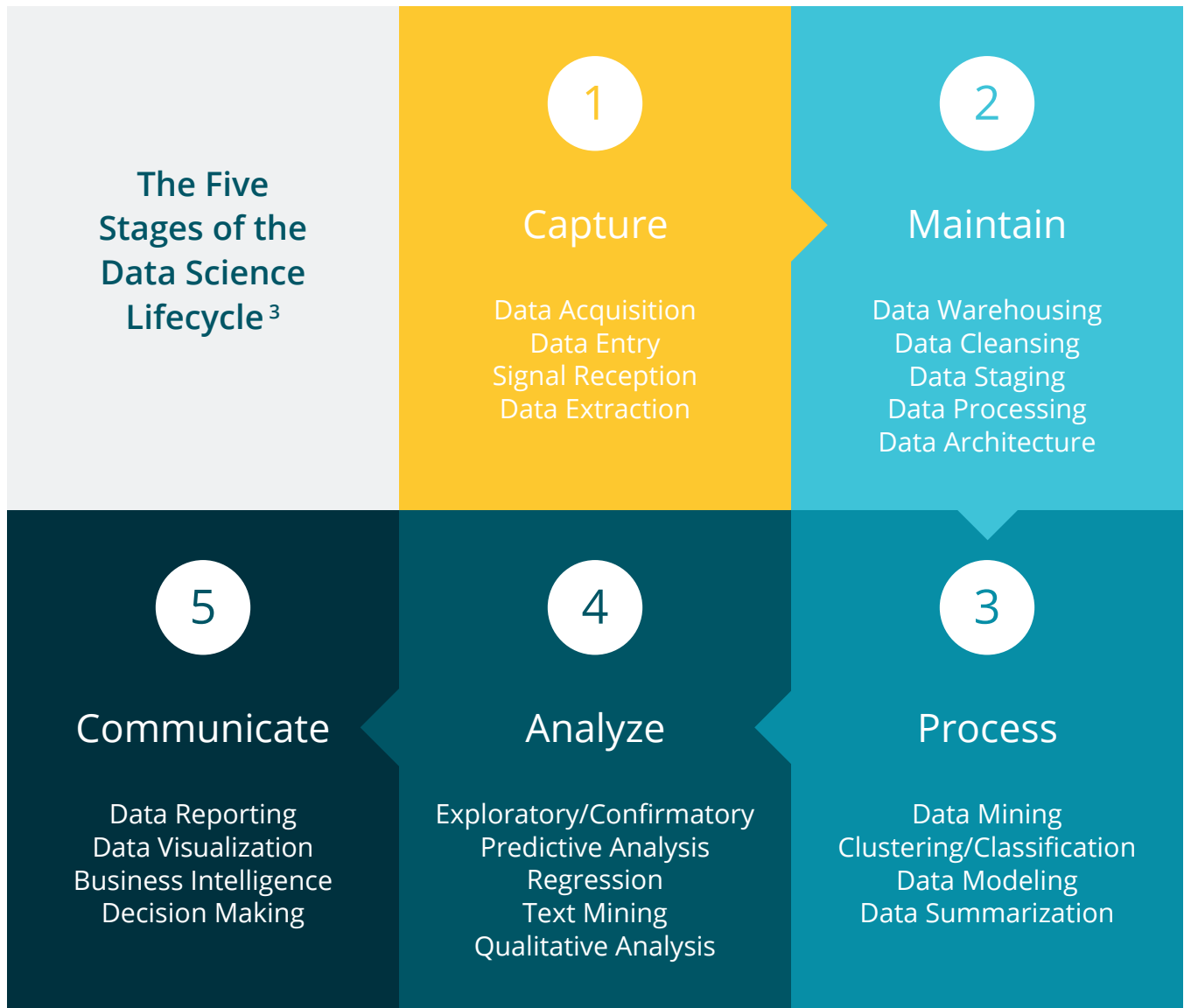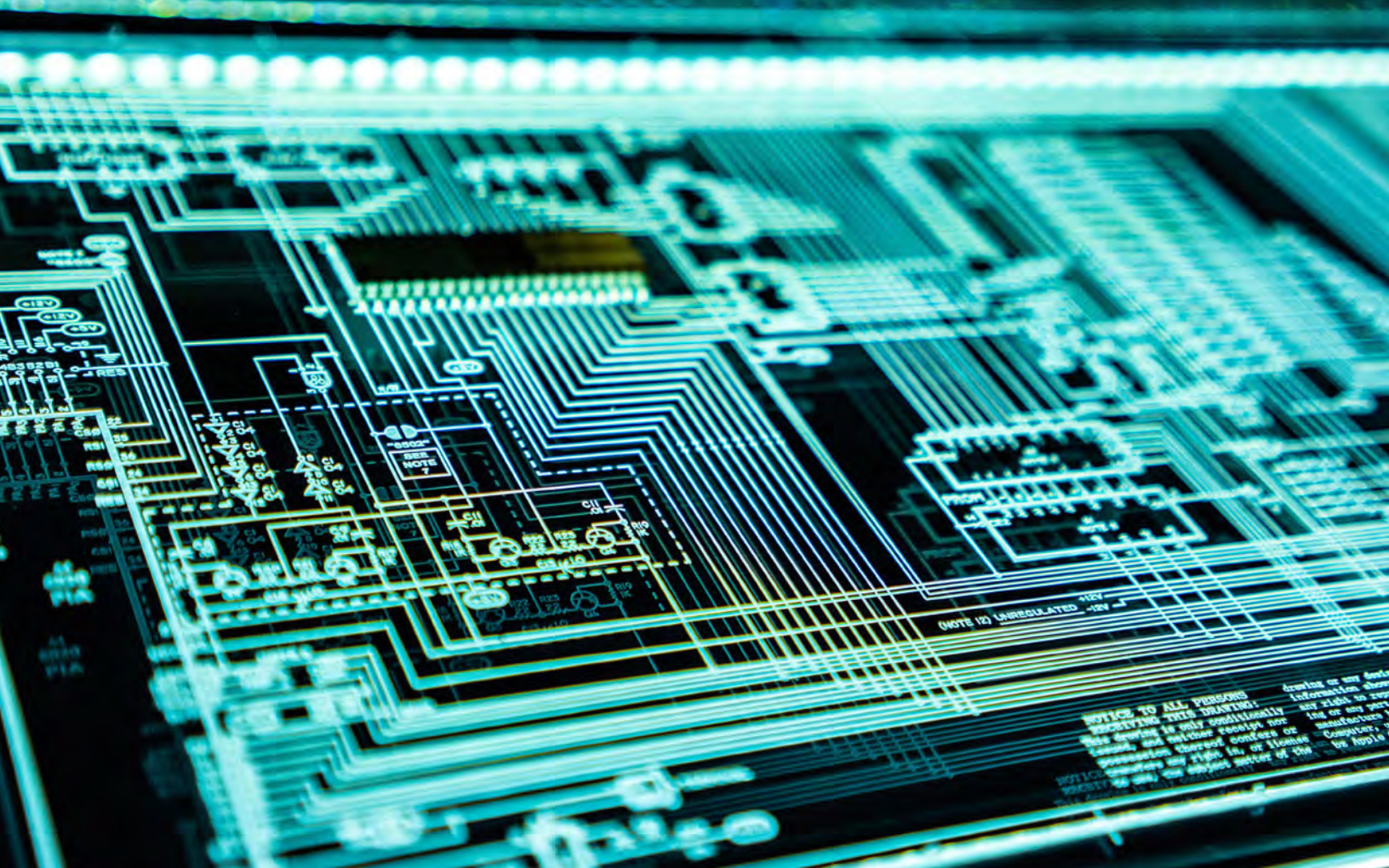Download ReviewRight Match AI™ Fact Sheet [PDF]

> From the point of data creation to the time of ESI review, HaystackID Discovery Intelligence helps you realize the potential of AI in all of your discovery efforts across the information lifecycle and litigation continuum.

# The Precision of Data Science

Data science can be defined as the process of using algorithms, methods, and systems to extract knowledge and insights from structured and unstructured data. [2] As part of HaystackID Discovery Intelligence, HaystackID's team of data science authorities have the expertise and experience to support the challenging and complex discovery tasks associated with the complete data science lifecycle. This lifecycle includes capturing, maintaining, processing, analyzing, and communicating intelligence to reach decision points and inform decision-makers.

**The Five Stages of the Data Science Lifecycle** [3]

### 1
## Capture

Data Acquisition
Data Entry
Signal Reception
Data Extraction

### 2
## Maintain

Data Warehousing
Data Cleansing
Data Staging
Data Processing
Data Architecture

### 5
## Communicate

Data Reporting
Data Visualization
Business Intelligence
Decision Making

### 4
## Analyze

Exploratory/Confirmatory
Predictive Analysis
Regression
Text Mining
Qualitative Analysis

### 3
## Process

Data Mining
Clustering/Classification
Data Modeling
Data Summarization

Whether detecting and identifying sensitive data or creating cutting-edge visualizations to help translate intelligence into action, HaystackID Discovery Intelligence provides you with industry-leading data science experts and expertise with conceptual and practical experience with:

- Algorithms and Models
- Analytics
- Coding

- Databases
- Frameworks and Libraries
- Knowledge of Key Business Domains

From data scientists to machine learning engineers and from data architects to application engineers, HaystackID Discovery Intelligence experts help you realize the precision of data science in all of your cybersecurity, information governance, and legal discovery efforts.

# The Power of Machine Learning

According to IBM [4], machine learning is a branch of AI and computer science that focuses on using data and algorithms to imitate the way humans learn, gradually improving its accuracy. In HaystackID Discovery Intelligence, HaystackID leverages predictive coding, a term used to describe a technology-assisted review (TAR) process which involves using a machine-learning algorithm to distinguish relevant and non-relevant documents based on a subject matter expert's coding a training set of documents. [5]

HaystackID Discovery Intelligence predictive coding is based on active learning. Active learning is a process, typically iterative, whereby an algorithm is used to select documents that should be reviewed for training based on a strategy to help the classification algorithm learn efficiently. [6] HaystackID Discovery Intelligence typically leverages a Continuous Active Learning® (CAL®) predictive coding protocol. In CAL, the TAR method developed, used, and advocated by Maura R. Grossman and Gordon V. Cormack, after the initial training set, the learner repeatedly selects the next-most-likely-to-be-relevant documents (that have not yet been considered) for review, coding, and training, and continues to do so until it can no longer find any more relevant documents. There is generally no second review because, by the time the learner stops learning, all documents deemed relevant by the learner have already been identified and manually reviewed. [7] Two examples of machine learning-enabled successes delivered by HaystackID as part of HaystackID Discovery Intelligence include:

**Enterprise Solutions Case Study**

## A Look at How Workflow, Analytics, and CAL® Can Make a Difference in Cost Savings

In this case study, a global manufacturer approached HaystackID to assist with processing, early case assessment (ECA), hosting, and managed review for a mid-sized, time-sensitive litigation matter. The client collected and securely transmitted approximately 650 GB of data, comprising several million documents, to HaystackID for processing and hosting. HaystackID, leveraging HaystackID Discovery Intelligence, worked collaboratively to create workflow efficiencies that yielded considerable cost savings. The combined team developed a workflow that used structured analytics and TAR (CAL) to deliver substantial and documented cost savings in the high six figures while completing the project. [8] Read More

Download Case Study [PDF]

**Legal Discovery Case Study**

# Speed, Slack, and Specialization – A Precision Approach to Second Requests

During the COVID-constrained summer of 2020, HaystackID supported a leading global finance platform company and its internationally recognized outside counsel in responding to a Department of Justice (DOJ) Second Request based on a proposed acquisition of a highly regulated company. As part of the Second Request, HaystackID collected and evaluated 18 TB of data, including critical Slack messages and files stores from onsite and remote locations from more than 17 types of data stores. Within 106 days, HaystackID completed approximately 300 collections and developed custom tools and processes. These custom tools and processes included innovative Slack-specific communications heat maps, predictive coding processes (CAL), and private message privilege identification to enable a compliant response to a complex investigation request, ultimately enabling a compliant response completion of the proposed acquisition. [9] Read More

Download Case Study [PDF]

From carefully considered identification of relevant documents to proactive support of information governance and data disposition tasks, HaystackID Discovery Intelligence helps you realize the power of machine learning in all of your discovery efforts.

# The Practicality of Expertly Trained and Managed Reviewers

While innovative technologies and techniques delivered by AI, data science, and machine learning are fundamental to discovery success, at the end of every process and practice, there is a requirement for human intervention or evaluation that can span the spectrum from document and source code review to deliverable authentication, validation, and verification. As the industry's most experienced and proficient provider of secure document review services with more than 2,000 virtual projects completed and more than 20,000 trained reviewers supporting onsite and remote reviews, HaystackID's ReviewRight offerings provide the expert and experienced human touch powering HaystackID Discovery Intelligence. HaystackID Discovery Intelligence managed review practicality in action includes:

**Case Study**

## Antitrust Agency Request Support: An Integrated Approach to Second Requests

In this high-velocity requirement to support six separate cannabis-related Second Requests within four months, HaystackID mobilized the complete review component of its HaystackID Discovery Intelligence capability. This capability allowed HaystackID to qualify and source domain experienced reviewers and managers, enabling rolling reviews with teams approaching 100 reviewers and cases as large as 600,000 documents. This capability also helped HaystackID complete review tasks, including specific privilege reviews and sensitive data (PII) reviews on time, within budget, and in compliance with antitrust agency requirements. [10] Read More

Download Case Study [PDF]

**Case Study**

# Shaping Outcomes with eDiscovery: A Complex Defense of Intellectual Property

In this matter between two technology-enabled manufacturing firms, HaystackID became a decisive contributor based on its aggressive yet controlled delivery of HaystackID Discovery Intelligence, including ReviewRight services to help its client find, understand, and learn from evidence. HaystackID Discovery Intelligence efforts collected 10 TB of ESI, reduced ESI from 10 TB to 430 GB through culling and processing, and further reduced ESI volume from 430 GB to 120 GB through review-centric analytics and TAR. The remaining 120 GB of ESI, about 550,000 documents, was completed within 90 days of initial collection through a combination of on-premises and remote reviews leveraging approximately 200 qualified expert reviewers. These efforts resulted in the identification of data-centric metadata changes that, when viewed within the context of all available ESI, substantiated the validity of the complaint of HaystackID's client, providing the proverbial smoking gun of evidence necessary for the successful defense of intellectual property. [11] Read More

Download Case Study [PDF]

## Core ReviewRight Offerings

- ReviewRight Staff™ (Global Document Review Marketplace)
- ReviewRight Virtual® (Secure Remote Review)
- ReviewRight Match AI™ (Reviewer Ranking and Selection)
- ReviewRight Protect™ (Data Breach Discovery, PII, PHI and Disclosure Review Services)
- ReviewRight Translate® (Foreign Language Review)
- ReviewRight Manage® (Managed Review)
- ReviewRight Host® (Review Hosting)
- ReviewRight Automate® (Technology-Enhanced Review)
- ReviewRight Secure Code Review (Virtual Source Code Review)
- ReviewRight Off-Shore Review (Off-Shore Review and Security)

From proactively certifying and sourcing domain expert reviewers and review managers to responsively supporting complex analytics and TAR tasks, HaystackID Discovery Intelligence helps you practically integrate the human touch into all of your discovery efforts.

# HaystackID Discovery Intelligence Synergy in Action

One example of the synergy provided by the combined application of HaystackID Discovery Intelligence artificial intelligence, data science, machine learning, and expert reviewers is manifested in HaystackID's Protect Analytics™ offering and the promise, precision, power, and practicality it provides.

Protect Analytics is an exclusive set of technologies and processes that allows client data set analysis for sensitive information ranging from PII and PHI to data breach code anomalies. Enabled by a collection of proprietary workflows and proven tools, Protect Analytics can proactively or reactively help determine sensitive data concentrations, locations, and relationships to inform notification lists, exposure assessments, and discovery targeting.

## A Proprietary Approach for Synergistic Sensitive Data Analysis

Key processes supporting data set analysis for sensitive information include:

- **Static Model Analysis:** The analysis and evaluation of a data set object structure through the use of class diagrams, domain models, context diagrams, and data flow diagrams that allow for the precise depiction of the logical aspects of data components.

- **Context-Sensitive (Markovian) Analysis:** The analysis and evaluation of current data variable behavior to predict the future behavior of those variables.

- **Regular Expression (Regex) Analysis:** The analysis and evaluation of data through the use of strings of text organized to create patterns that allow for matching, locating, and considering sensitive data.

The proprietary Protect Analytics approach to data set analysis, based on the combination and integration of these three data set processes, inform HaystackID data science, cybersecurity, and eDiscovery experts in their use of investigation, discovery, and visual analysis tools to translate analysis into actionable information for downstream decisions and deliverables.

## Cutting-Edge Command of Traditional and AI-Enabled Platforms

Key platforms supporting data-driven actions on sensitive information include:

- **Investigation Platforms:** This group of platforms enables data extraction, correlation, and contextualization to translate data information into actionable insight.

- **Discovery Platforms:** This group of platforms combines processing and review capabilities powered by artificial intelligence (e.g., machine learning) and expert intelligence (e.g., data scientists) to translate insight into actionable information.

- **Intelligence Platforms:** This group of platforms unifies, visualizes, and reports on data from analysis, investigation, and discovery activities, helping decision-makers translate data points into decision points.
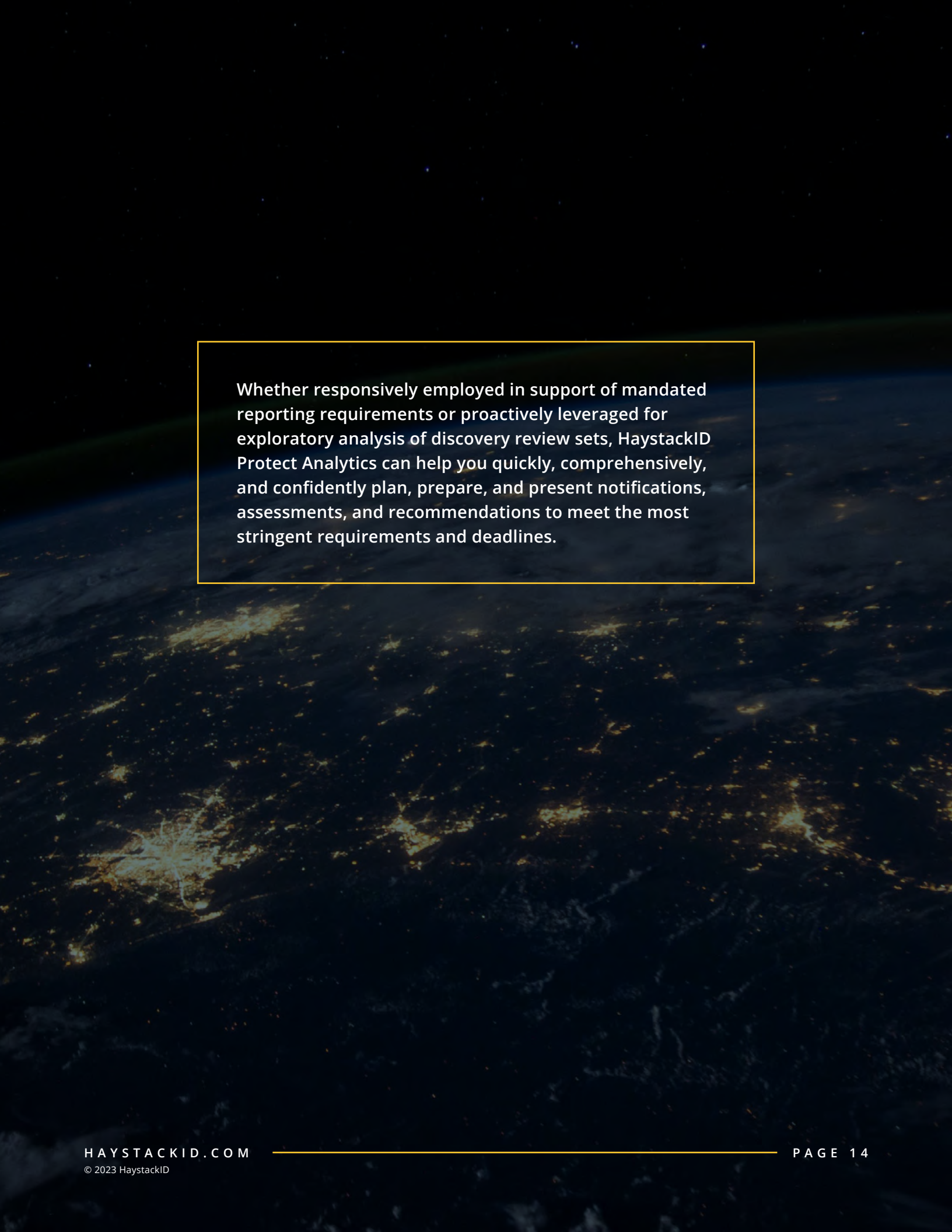
The proprietary workflows of Protect Analytics used with powerful and precise investigation, discovery, and intelligence platforms help HaystackID and client experts make data-driven decisions regarding sensitive information-centric deliverables.


## Comprehensive Answers for Decisive Deliverables

After analysis with proprietary processes and data-driven task accomplishment with powerful and precise investigation, discovery, and intelligence platforms, HaystackID is authoritatively prepared to meet client requirements in delivering sensitive data-centric deliverables ranging from notifications to assessments.

Key sensitive-data centric deliverables include:

- **Notification Lists:** From GDPR compliance to data breach notifications, HaystackID can help you determine regulatory and reporting requirements and prepare notification lists for use in notifying law enforcement agencies, business partners, and impacted individuals.

- **Impact Assessments:** From Data Protection Impact Assessments (DPIA) to Privacy Impact Assessments (PIA), HaystackID can help you identify and minimize risks related to personal data processing.

- **Discovery Targeting Recommendations:** From identification and location to concentration and density, HaystackID can help inform eDiscovery audits, investigations, and litigation of sensitive data by providing details, descriptions, and disposition sensitive data in discovery review sets.

Whether responsively employed in support of mandated reporting requirements or proactively leveraged for exploratory analysis of discovery review sets, HaystackID Protect Analytics can help you quickly, comprehensively, and confidently plan, prepare, and present notifications, assessments, and recommendations to meet the most stringent requirements and deadlines.

# Learn More. Today.

Contact us today to learn more about how HaystackID Discovery Intelligence can help you harness the potential of artificial intelligence, the precision of data science, the power of machine learning, and the practicality of expertly trained and managed reviewers to gain insight and intelligence that allows you to reach decision points faster and more economically than previously possible.

---

**About HaystackID®**

HaystackID is a specialized eDiscovery services firm that supports law firms and corporate legal departments and has increased its offerings and expanded with five acquisitions since 2018. Its core offerings now include Global Advisory, Discovery Intelligence, HaystackID Core™, and artificial intelligence-enhanced Global Managed Review services powered by ReviewRight®. The company has achieved ISO 27001 compliance and completed a SOC 2 Type 2 audit for all five trust principles for the second year in a row. Repeatedly recognized as a trusted service provider by prestigious publishers such as Chambers, Gartner, IDC, and The National Law Journal, HaystackID implements innovative cyber discovery services, enterprise solutions, and legal discovery offerings to leading companies across North America and Europe, all while providing best-in-class customer service and prioritizing security, privacy, and integrity. For more information about its suite of services, including programs and solutions for unique legal enterprise needs, please visit HaystackID.com.

**References:**

[1]   EDRM – The Use of Artificial Intelligence in eDiscovery (2021) View

[2]   IBM – Data Science (2021). View

[3]   Berkeley School of Information – What is Data Science? (2021) View

[4]   IBM – Machine Learning (2021). View

[5]   Predictive Coding Technologies and Protocols (2021). View

[6]   Ibid.

[7]   Grossman, M. and Cormack, G. (2016). Continuous Active Learning for TAR. [ebook] Practical Law. Available at: https://pdfs.semanticscholar.org/ed81/f3e1d35d459c95c7ef60b1ba0b3a202e4400.pdf [Accessed 27 October 2021].

[8]   Rubinger, A. and Tschannen, C. (2021) A Look at How Workflow, Analytics, and CAL® Can Make A Difference in Cost Savings. HaystackID. Available at https://haystackid.com/a-look-at-how-workflow-analytics-and-cal-can-make-a-difference-in-cost-savings/.

[9]   Sarlo, M. (2021). Speed, Slack, and Specialization – A Precision Approach to Second Requests. HaystackID. Available at https://haystackid.com/speed-slack-and-specialization-a-precision-approach-to-second-requests/.

[10]  Sarlo, M. (2020). Antitrust Agency Request Support: An Integrated Approach to Second Requests. HaystackID. Available at https://haystackid.com/an-integrated-approach-to-second-requests/.

[11]  Sarlo, M. (2018). Shaping Outcomes with eDiscovery: A Complex Defense of Intellectual Property. HaystackID. Available at https://haystackid.com/shaping-outcomes-with-ediscovery-a-complex-defense-of-ip/.