

CASE STUDY

# Leveraging Hybrid Generative AI + Technology-Assisted Review Workflows to Reduce Review Populations and Costs

How HaystackID® Leveraged Relativity® aiR for Review™  
and TAR to Deliver Defensible Results

*By Matthew Sinner, Esq., Bernie Gabin, and Young Yu*

HAYSTACK®



## Executive Summary

HaystackID® successfully tackled a common enterprise challenge: reviewing **891,527 documents** with limited resources while maintaining defensibility and controlling costs. Facing a low-richness corpus (~4%), single-attorney bandwidth constraints, and budget limitations, the team developed an innovative hybrid workflow combining generative AI (GenAI) from Relativity® (aiR for Review™), Technology-Assisted Review (TAR), and conceptual analytics under attorney oversight.

The approach prioritized high-value documents through TAR ranking, reducing the volume sent to GenAI processing by **more than 90%**—from **748,612** eligible documents to just **54,380** actually processed. This strategic prioritization maintained exceptional accuracy with **90.00% Precision** and **95.58% Recall** in the final production of **27,976 documents**, validated through rigorous statistical sampling at **95% Confidence Levels**.

**This case study demonstrates a defensible, cost-effective path for large-scale document review that balances automation with human expertise.** The methodology is directly applicable to cybersecurity incident response, information governance, and eDiscovery productions where teams face tight deadlines, limited resources, and stringent accuracy requirements.

# Methodology Overview

In this matter, conducted by HaystackID, the corpus comprised **891,527 documents**, with **748,612** eligible for GenAI analysis by aiR for Review. The remainder—**142,915 documents**—were excluded by file-type and text-size criteria (text **smaller than 0.2 KB** or **larger than 300 KB**) or because files were corrupt or unreadable. Anticipated attorney bandwidth was limited to a single attorney, making a Prioritized Review (Active Learning) approach impractical, and running the full eligible population through GenAI was not cost-effective. The strategy, therefore, centered on a hybrid workflow that joined aiR, TAR, and conceptual analytics, with counsel oversight guiding prompt development and quality control.

The workflow progressed in stages. Counsel's review of an initial random sample of **384 documents**, at a **95% Confidence Level** and a **+/- 5% Margin of Error**, yielded **17 Relevant documents**—approximately **4% richness**—and established a baseline for ground truth. Counsel's coding also seeded a TAR model to rank documents by similarity, enabling targeted sampling and focused submissions to aiR. Two subsequent tests—one rank-informed and one designed for greater conceptual diversity—refined the prompt language and stabilized TAR model performance. Submissions then scaled in tranches of **2,000 to 4,000 records**, blending high-scoring items with limited random mid- and low-scoring items to maintain index health and conceptual coverage. Targeted manual review addressed items not suitable for aiR analysis. Final validation used two statistical samples at a **95% Confidence Level**, supporting a **27,976-document** production with **90.00% Precision** and **95.58% Recall**, alongside a reduction of **more than 90%** in GenAI processing volume.

## Context and Objectives

The matter presented a large corpus, low overall richness, and a meaningful number of documents that initial GenAI runs could not confidently classify. The objective was to apply a hybrid workflow—aiR, TAR, and conceptual analytics—to reduce cost while maintaining accuracy and defensibility through attorney oversight and a documented validation plan.

## The Challenges

Four constraints shaped the design, including:



Anticipated attorney bandwidth limited the viability of a Prioritized Review training cadence.



Running the entire GenAI-eligible population through aiR would have been cost-prohibitive.



Random sampling alone, in a low-richness collection, risked inadequate conceptual coverage for effective prompt evaluation and improvement.



The prevalence of Borderline items required a measured plan that contained cost and preserved defensibility.

# Corpus and Eligibility

Eligibility for GenAI analysis followed defined file-type and text-size thresholds, excluding items with extracted text **below 0.2 KB** or **above 300 KB** and items that were corrupt or unreadable. These criteria, together with single-attorney bandwidth and cost controls, framed the workflow within Relativity's Review Center and related toolsets.

## Workflow Overview

The workflow used counsel-authored prompts and staged sampling to establish ground truth, seeded a TAR model with counsel's coding to rank the corpus, and then scaled submissions to aiR in controlled tranches while maintaining conceptual coverage and ongoing quality control. Targeted manual review addressed items not suitable for aiR, and outcomes were validated by HaystackID with two statistical samples before finalizing production and reporting Precision, Recall, and Elusion.

## Prompt Development and Initial Sampling

Counsel collaborated on initial prompt language that captured case context, entities, and target document descriptions. A statistically valid test set of **384** randomly selected documents produced **17** Relevant documents—approximately **4% richness**. Counsel's evaluation of False Positives, False Negatives, and Borderline items informed prompt refinements using aiR's rationales and considerations, keeping subsequent tests focused on classification clarity and conceptual coverage.

## Seeding TAR and Ranking the Collection

Counsel's coding from the first sample seeded a TAR model in Relativity's Review Center. The model ranked documents by similarity to counsel-coded Relevant and Not Relevant items, guiding selection for the next test set and defining a more efficient scope for submissions to aiR.

## Prompt Iteration Two

A second test combined the original **384 documents** with **500 additional documents** having high TAR relevance scores. The updated prompt surfaced more Relevant items, though many were conceptually similar, and a significant population of Borderline items remained. Counsel reviewed disagreements and Borderline items, informing both prompt updates and continued TAR training.

## Prompt Iteration Three

To expand conceptual coverage and challenge earlier results, counsel supplied exemplar documents from outside the review set. Using Relativity's conceptual analytics, the team identified similar items within the review set. The third sample consisted of **500 documents**, comprising **126** previously coded as Borderline, **174** previously coded as Not Relevant, **52** previously coded as Relevant, and **148 new documents** that were both conceptually similar to the exemplars and highly ranked by TAR. Outcomes improved measurably. Of the **126** prior Borderline items, **28** received definitive classifications, leaving **98**, and among non-Borderline items, aiR classifications matched counsel's coding **90% of the time**.

## Scaling Submissions with Tranches

With performance stabilized, the project scaled without submitting the entire eligible universe. TAR prioritized likely relevant documents, and submissions proceeded in tranches of **2,000 to 4,000 records**. Each tranche blended high-scoring items with a limited random selection of mid- and low-scoring items to maintain index health and conceptual coverage. New aiR classifications and counsel quality control continued to train the TAR model. When performance plateaued, a TAR cutoff was set, and the remaining documents above the cutoff were submitted to aiR.

## Manual Review Scope

Not every document was suitable for aiR analysis. A contract review team conducted linear relevance review for documents excluded at the outset that hit search terms, and for items that aiR labeled Borderline or Error above the TAR cutoff, where such review was appropriate. Borderline and Error documents below the TAR cutoff were excluded from manual review. In total, **12,923 documents** within the GenAI-eligible set required manual review.

## Validation Design

Validation relied on two samples at a **95% Confidence Level**. The Null Set comprised **696,693 documents** that had not undergone human review and were either classified Not Relevant by aiR or not submitted to aiR due to low TAR scores. A random sample of **2,393 documents** with a **+/- 2% Margin of Error** yielded four Relevant documents, which projected an Elusion rate of **0.17%** or approximately **1,165 documents**. The Presumptively Relevant Set comprised **20,412 documents** not previously reviewed by humans but classified Relevant by aiR. A random sample of **1,430 documents** with a **+/- 2.5% Margin of Error** confirmed that **86.29%** were truly Relevant.

## Production and Accuracy Metrics

The production set totaled **27,976 documents**. It consisted of **17,614 Relevant documents** projected from the Presumptively Relevant pool, together with **7,564 documents** coded Relevant through earlier human review. Of the produced documents, **25,178** were projected to be truly Relevant, yielding **90.00%** Precision. Recall was calculated as **25,178** divided by **26,343**, producing **95.58%**.

## Cost Outcome

Prioritizing submissions based on TAR scores while maintaining conceptual coverage through selective sampling meant that only **54,380** of the **748,612** GenAI-eligible documents were analyzed by aiR. This approach reduced GenAI processing volume by more than **90%** while meeting accuracy and validation standards.

## Conclusion

This matter presents a defensible path for large, low-richness reviews. **GenAI provided content interpretation and contextualization, TAR supplied ranking and focus, and counsel oversight anchored quality.** Structured sampling and staged submissions reduced cost and risk while supporting statistical validation. The same design principles—clear eligibility rules, disciplined prompt iteration, ranking-driven prioritization, and transparent validation—can guide similar matters where budgets, timelines, and accuracy standards must be met simultaneously.

# Postscript: By the Numbers

- **Corpus Size:** 891,527 documents
- **GenAI-Eligible:** 748,612 documents
- **Excluded:** 142,915 (file-type limits; text <0.2 KB or >300 KB; corrupt/unreadable)
- **Initial Random Sample:** 384 documents (95% CL; +/-5% MOE); 17 Relevant (~4% richness)
- **Second Test Set:** 384 + 500 high-TAR-score documents
- **Third Test Set (500 total):** 126 prior Borderline; 174 prior Not Relevant; 52 prior Relevant; 148 new (conceptually similar + high TAR)
- **Borderline Change:** 28 of 126 prior Borderline received classifications (22.22% reduction; 98 remaining in that cohort)
- **Non-Borderline Agreement:** 90% aiR classifications matched counsel coding
- **Tranche Size to aiR:** 2,000–4,000 records per run
- **Documents Analyzed by aiR:** 54,380 (of 748,612 eligible) → >90% reduction in GenAI volume
- **Manual Review (within eligible set):** 12,923 documents
- **Validation – Null Set:** population 696,693; sample 2,393 (95% CL; +/-2%); 4 Relevant → 0.17% elusion (≈1,165 documents)
- **Validation – Presumptively Relevant:** population 20,412; sample 1,430 (95% CL; +/-2.5%); 86.29% confirmed Relevant
- **Production:** 27,976 documents total
- **From Presumptively Relevant (projected Relevant):** 17,614
- **From earlier human review (coded Relevant):** 7,564
- **Projected Truly Relevant in Production:** 25,178 → Precision 90.00%
- **Recall:** 25,178 / 26,343 = 95.58%

---

## Learn More Today.

Contact us today to learn how HaystackID can help you solve your complex data challenges related to legal, compliance, regulatory, and cyber events.

### About HaystackID®

HaystackID solves complex data challenges related to legal, compliance, regulatory, and cyber requirements. Core offerings include Global Advisory, Cybersecurity, Core Intelligence AI™, and ReviewRight® Global Managed Review, supported by its unified CoreFlex™ service interface. Recognized globally by industry leaders, including Chambers, Gartner, IDC, and Legaltech News, HaystackID helps corporations and legal practices manage data gravity, where information demands action, and workflow gravity, where critical requirements demand coordinated expertise, delivering innovative solutions with a continual focus on security, privacy, and integrity. Learn more at [HaystackID.com](https://HaystackID.com).

## About the Authors

### **Matthew Sinner, Esq.**

*Senior Generative AI And Analytics Consultant, HaystackID*

Matthew Sinner joined HaystackID in 2020 and is currently a Senior Generative AI And Analytics Consultant. In this role, he develops and implements workflows utilizing structured analytics, conceptual analytics, machine learning, and GenAI. He works closely with project managers, clients, and review teams, providing guidance on appropriate strategies to address varying situations and ensuring the delivery of timely and effective solutions for client requests. Prior to joining HaystackID, Matthew was a Senior Analytics Consultant at NightOwl Global. He has been in various roles at NightOwl since 2015 and has been active in the eDiscovery space since 2012. Prior to joining NightOwl, Matthew worked with a political advocacy group and organized the intellectual property operations of a specialized construction materials company.

### **Bernie Gabin**

*Chief Data Scientist, HaystackID*

In 2022, Dr. Bernie Gabin joined HaystackID and is currently the Chief Data Scientist on the Data Science team. In this role, he works closely with the company's CDS, John Brewer, to apply data-driven metrics to improve our procedures and develop custom AI/ML-empowered solutions for our clients. Prior to joining HaystackID, Bernie received his Ph.D. in physics from Brandeis University. His doctoral work in brain-computer interface systems and machine learning/artificial intelligence led him to work on AI/ML-focused projects for the US Patent Office, Northrup Grumman, and the National Security Agency. At HaystackID, he brings his expertise in signal processing, AI design, and data modeling to create novel data-driven solutions to our most challenging problems.

### **Young Yu**

*Vice President of Advanced Analytics and Strategic Solutions, HaystackID*

Young Yu joined HaystackID in 2018 as a director and is currently the Vice President of Advanced Analytics and Strategic Solutions. In this role, Young is the primary strategic and operational adviser to HaystackID clients, focusing on the planning, execution, and management of eDiscovery activities. Young brings extensive experience to his position, having previously worked at IPRO Tech as a Professional Services Consultant/Product Manager for Analytics, Wilmer Cutler Pickering Hale & Dorr LLP as a Team Lead/Litigation Support Coordinator, Chadbourne & Parke LLP as a Project Manager, and Ikon Office Solutions in various roles, including Data Engineer and Global Database Administrator. He holds certifications as a Brainspace Certified Admin, Analyst, and Specialist and is affiliated with Agile, Scrum, and Six Sigma. Young is proficient in a wide range of software and platforms, including Microsoft Office 365, Salesforce, SQL, Relativity, Ringtail, and many others.

Assisted by GAI and LLM technologies.

Source: HaystackID