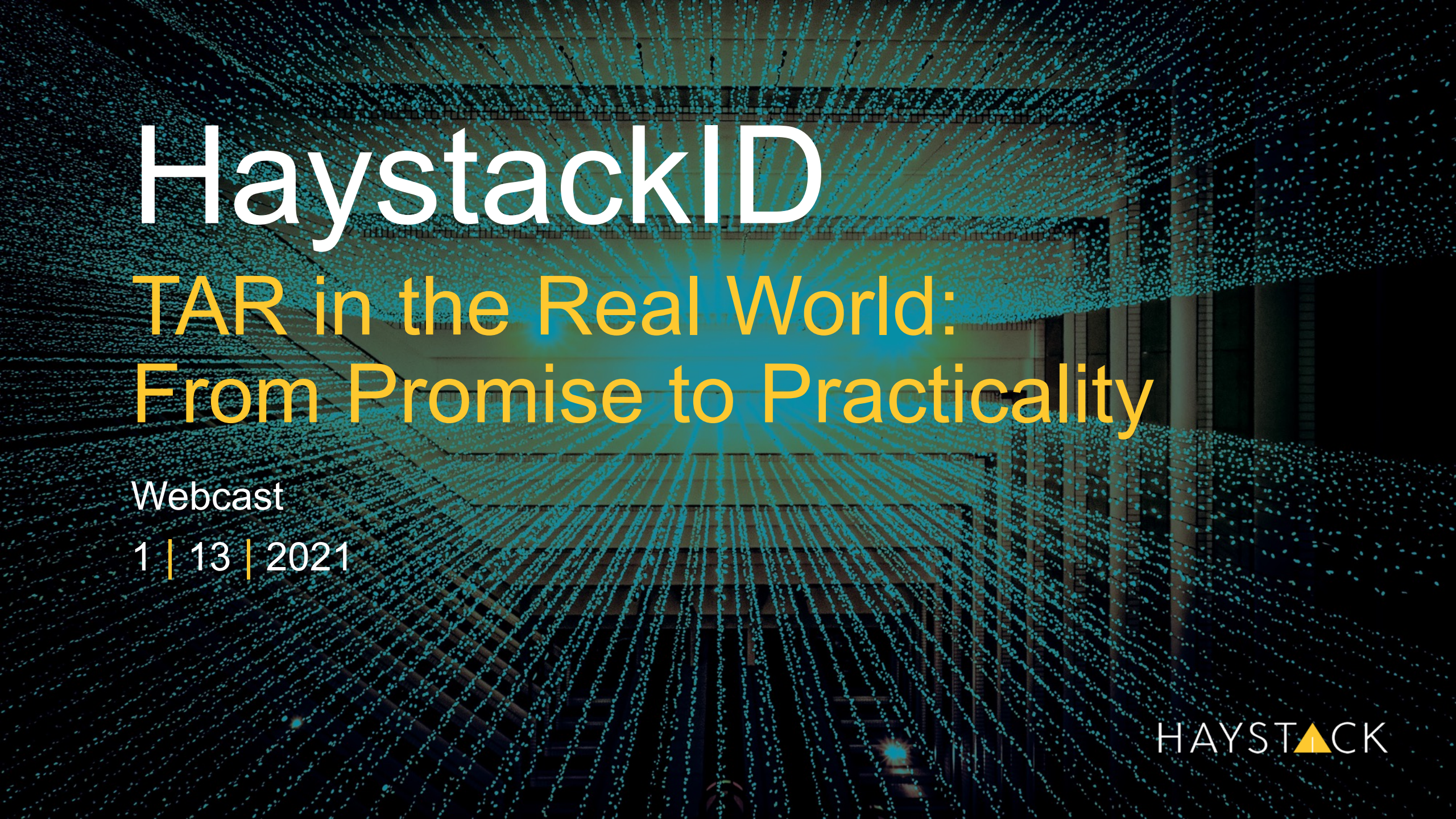

HaystackID
TAR in the Real World:
From Promise to Practicality
Webcast
1 | 13 | 2021
HAYSTACK

HaystackID Team

- **Michael Sarlo**, EnCE, CBE, CCLO, RCA, CCPA - Michael serves as the Chief Innovation Officer and President of Global Investigations for HaystackID.
- **Adam Rubinger, JD.** - As an EVP with HaystackID, Adam brings more than 20 years of experience and applied expertise in advising, consulting, and managing large-scale eDiscovery projects.
- **Anya Korolyov, Esq.,** Relativity Master - As Director of Project Management with HaystackID, Anya has 12 years of experience in eDiscovery with extensive expertise with Second Requests as an attorney and consultant.
- **Seth Curt Schechtman, Esq.** - As Senior Managing Director of Review Services for HaystackID, Seth has over 15 years of document review experience, including class actions, MDLs, and Second Requests.
- **Young Yu** - As Director of Client Service with HaystackID, Young is the primary strategic and operational advisor to clients in eDiscovery matters.

HAYSTACK

Agenda - the Everyday Tools of eDiscovery

- **Structured Analytics:** Threading, Near Duplicate Analysis, Name Normalization, Language ID
- **Conceptual Analytics:** TAR 1.0, CAL, Clusters
- **Brainspace** or **Relativity**
- **Stopping Point:** The Why and When of Workflow Decisions with Continuous Active Learning

HAYSTACK

How eDiscovery is Transformed by Analytics

Diminishes the document pool

Makes data More accessible

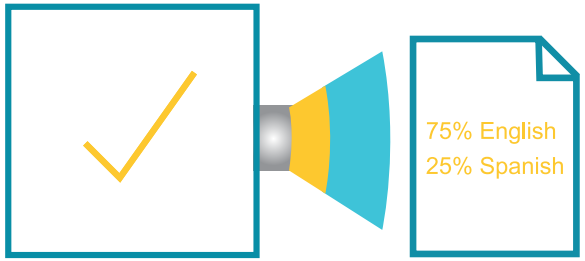
Allows for informed business decisions

Redirects billable hours and saves money

HAYSTACK

Structured Analytics

Near Duplicate Analysis



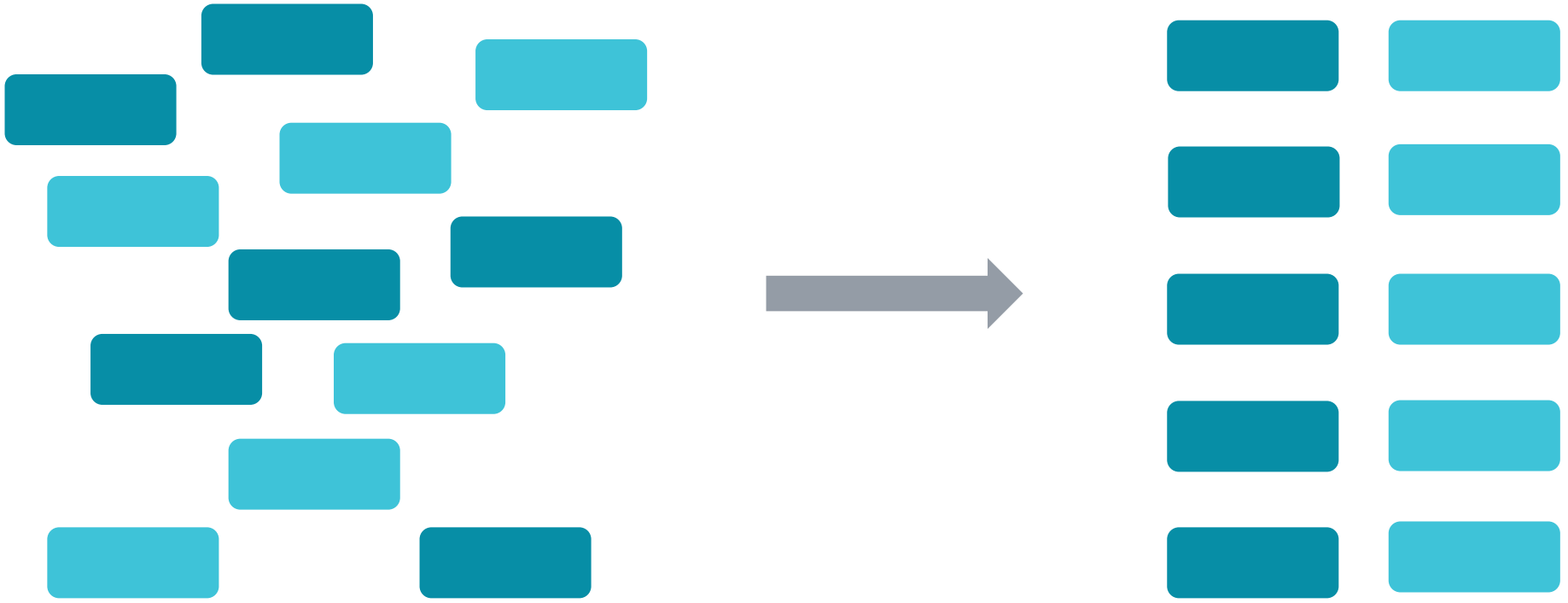
Language ID

Name Normalization

HAYSTACK

Email Threading

Groups a string of related emails together in a chain.



HAYSTACK

Poll Question

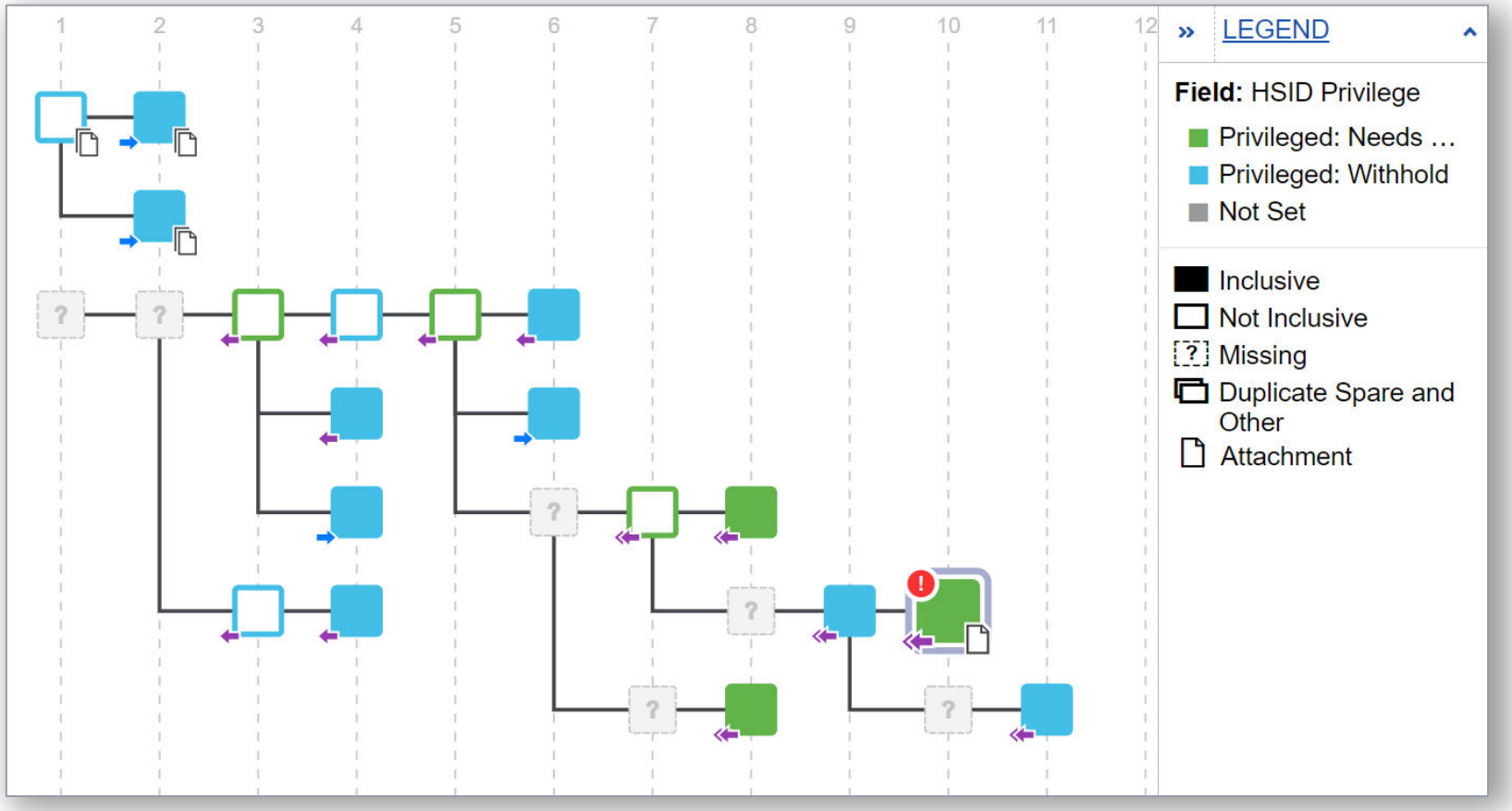
Over the past year, how often have you made use of threading to organize the review and/or assist with quality-control?

HAYSTACK

Threading Workflows

HAYSTACK

Thread Visualization



HAYSTACK

Machine Learning

Unsupervised Learning

Artificial Intelligence that automatically learns like a human brain. Patterns and themes are systematically detected and presented to the user in interactive data visualizations and transparent Concept Search. This automated learning is done without human guidance, examples Cluster Wheel and Concept Search

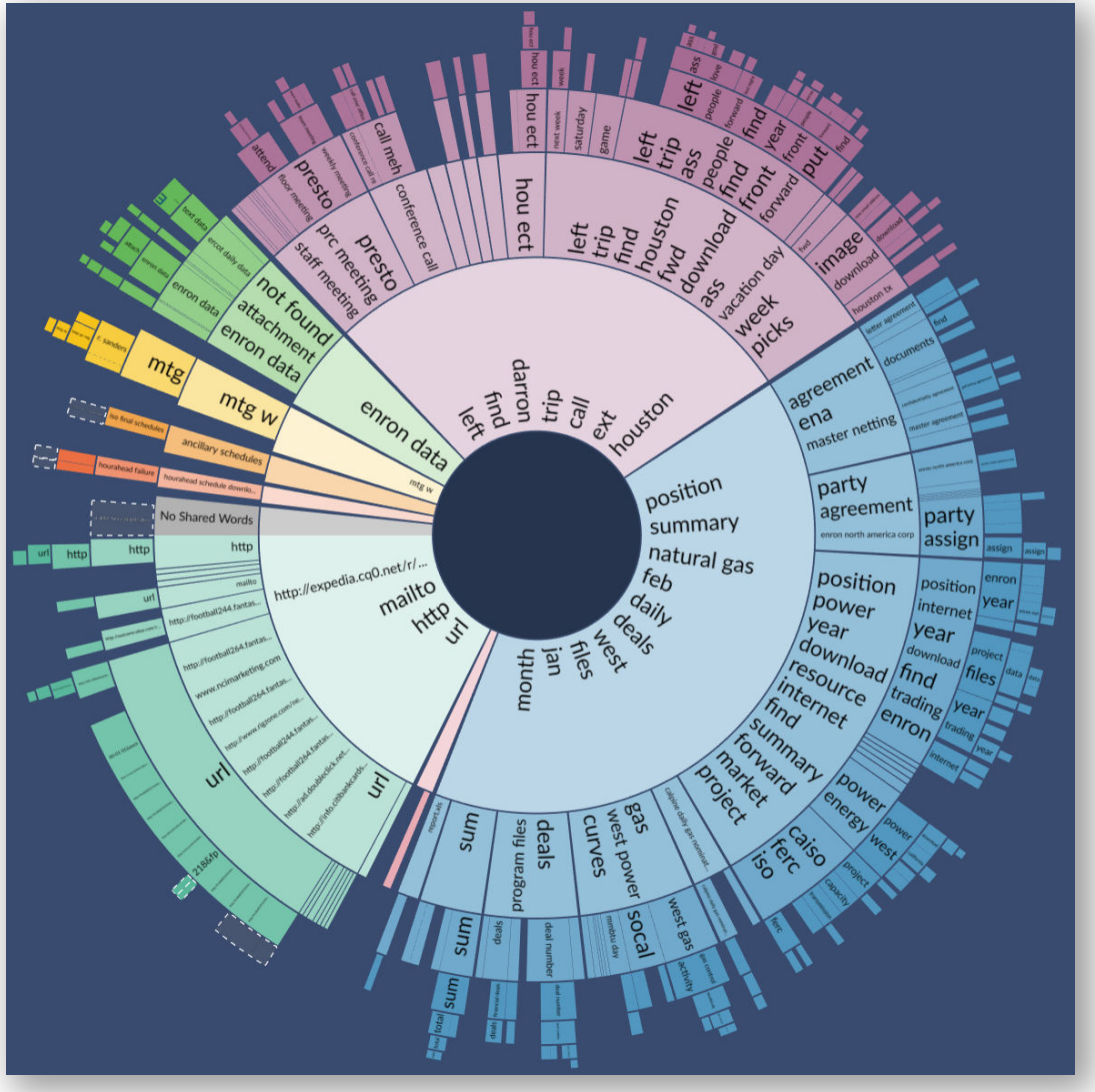
Supervised Learning

Human decisions are used to teach the machine what to look for and in turn the machine can surface insights previously unknown about your data. Human decisions can also be used to build predictive models which can sort and organize documents using positive and negative examples

HAYSTACK

Clusters

- Get to know your data
- Identify similar concepts and documents
- Visually key-in on privileged, responsive, or hot clusters

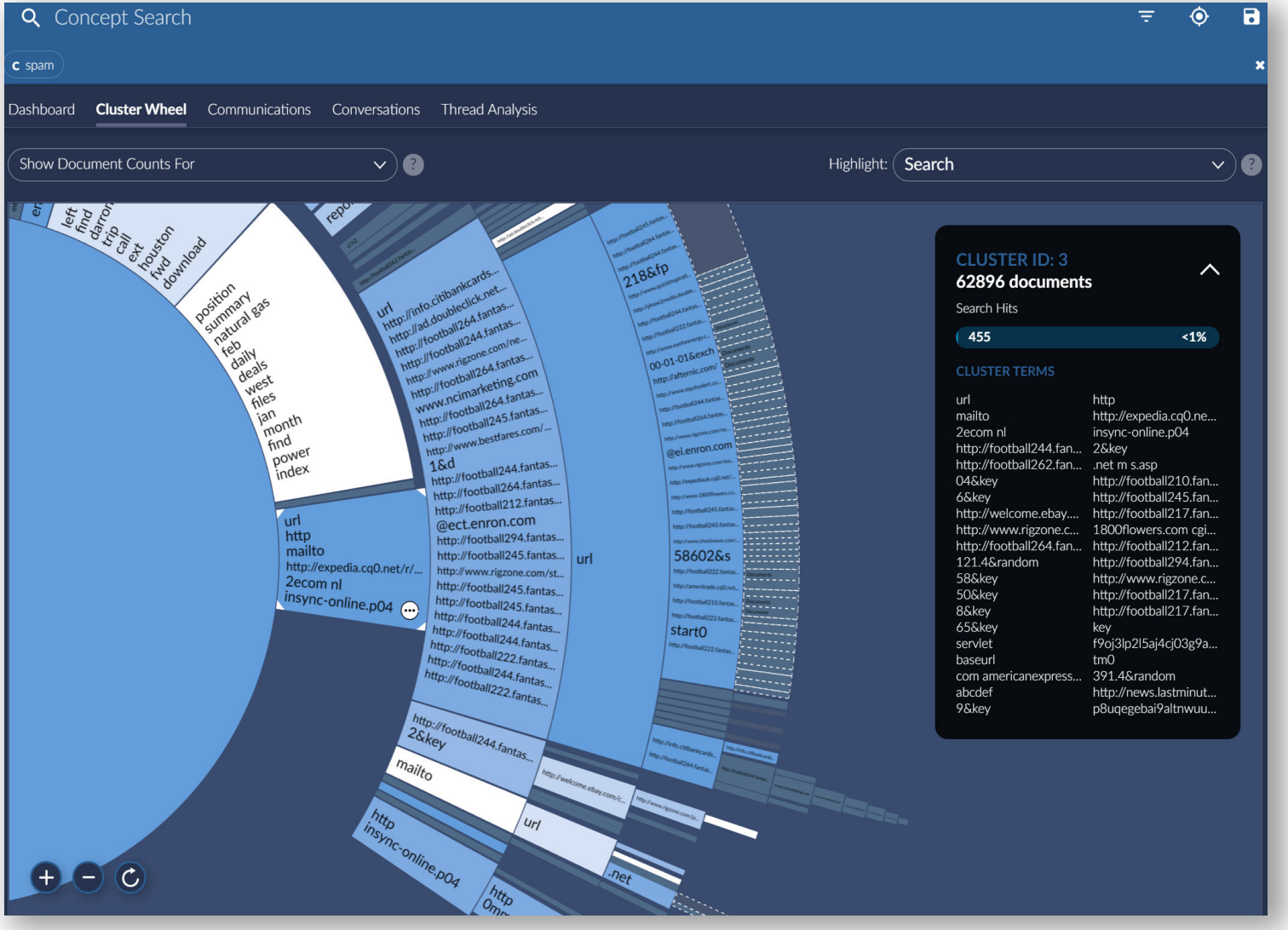


© 2021 HaystackID

HAYSTACK

Clusters

- Identify Relevant and Not Relevant concepts
- Analyze Cluster Terms

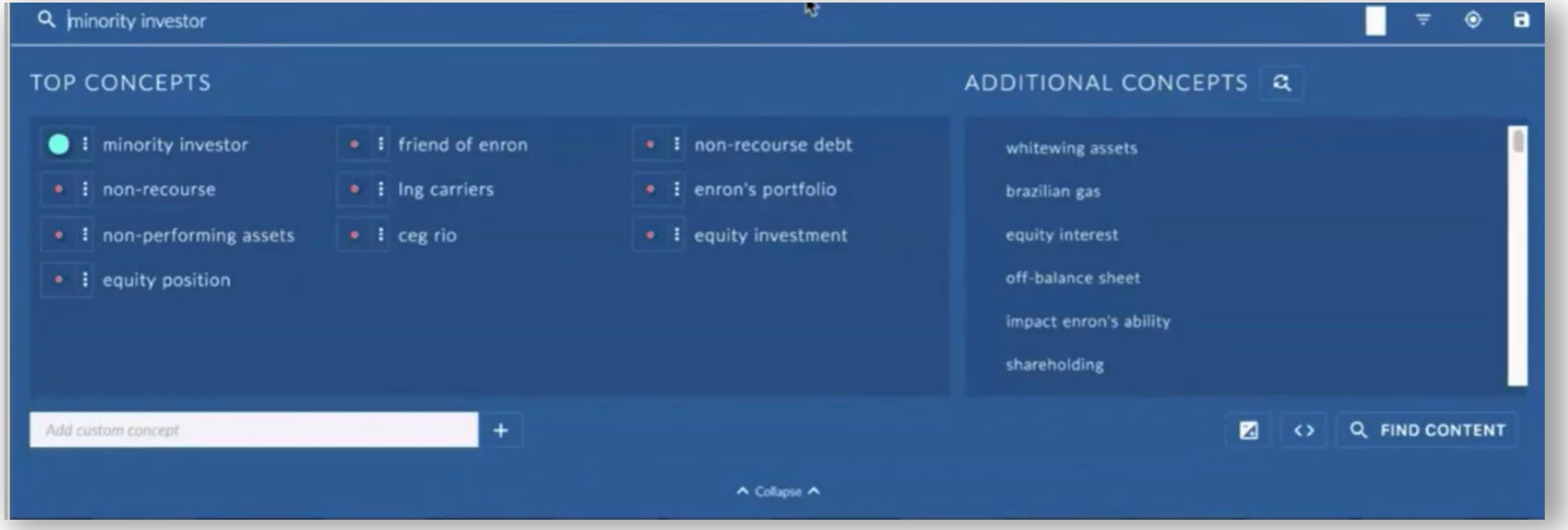


© 2021 HaystackID

HAYSTACK

Concept Searching

- Different from Keyword searching
- Benefits of Brainspace concept searching vs. Relativity
- Drill down into additional repeating concepts
- Investigative Benefits



© 2021 HaystackID

HAYSTACK

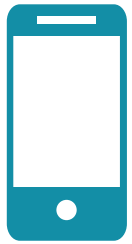
Issues and Solutions with TAR Standard Exclusions

- Documents with little text or substantive discussion/Large text files
- Outlook Calendar Invitations without content in the body
- Audio/Video/Image files
- Spreadsheets

HAYSTACK

Solutions for Short Format Messages

Teams



Mobile

Bloomberg

Slack

HAYSTACK

Poll Question

Over the past year, what percentage of matters that have required review have you used TAR 1.0 or 2.0?

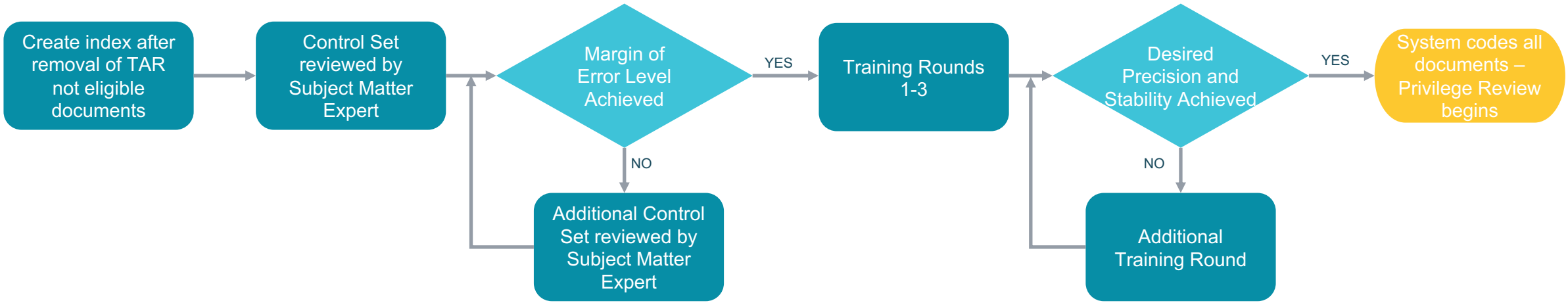
HAYSTACK

Defining Relevance

- The scope and definition of Relevance should be considered carefully when utilizing any TAR workflows.
- Defined too narrowly, the model may not identify peripherally relevant documents.
- Defined too widely, the model may be overinclusive and identify marginally relevant documents.

HAYSTACK

TAR 1.0 Workflow



HAYSTACK

Training Round Consideration

Relativity

Stratified: Groups the round saved search documents into subgroups based on the documents' concepts and returns the documents that cover most of the conceptual space or until the Maximum sample size or Minimum seed influence has been met. This type of sampling allows RAR to effectively train with as few documents as possible. Selecting this type makes the Maximum sample size and Minimum seed influence fields available and disables the Calculate sample button. The Stratified sampling option is only available when you select the Training round type.

Statistical: Creates a sample set based on statistical sample calculations, which determines how many documents your reviewers need to code in order to get results that reflect the project universe as precisely as needed. Selecting this option makes the Margin of error field required.

Percentage: Creates a sample set based on a specific percentage of documents from the project universe. Selecting this option makes the Sampling percentage field required.

Fixed Sample: Creates a sample set based on a specific number of documents from the project universe. Selecting this option makes the second Fixed sample size field required.

Brainspace

Random: Simple random sample of documents not already used for training. Best Use: When necessary to guarantee that documents are selected independent of user input.

Fast Active: This selection method favors documents that 1) appear in clusters distant from each other and from those of previous training documents, 2) are similar to many other data set documents, and 3) have a score near 0.5 under the current predictive model. Use When: Fast training is necessary in large batches.

Influential: This selection method favors documents that 1) are different from each other (and from previous training documents if this is not the first batch), and 2) are similar to many other data set documents. Use When: This is your first training round and there is no manual seed set available.

Diverse Active: This selection method favors documents that 1) are different from each other and from previous training documents, 2) are similar to many other data set documents, and 3) have a score near 0.5 under the current predictive model. Use When: Accelerated training is necessary and to avoid any manual influence.

Reporting Brainspace

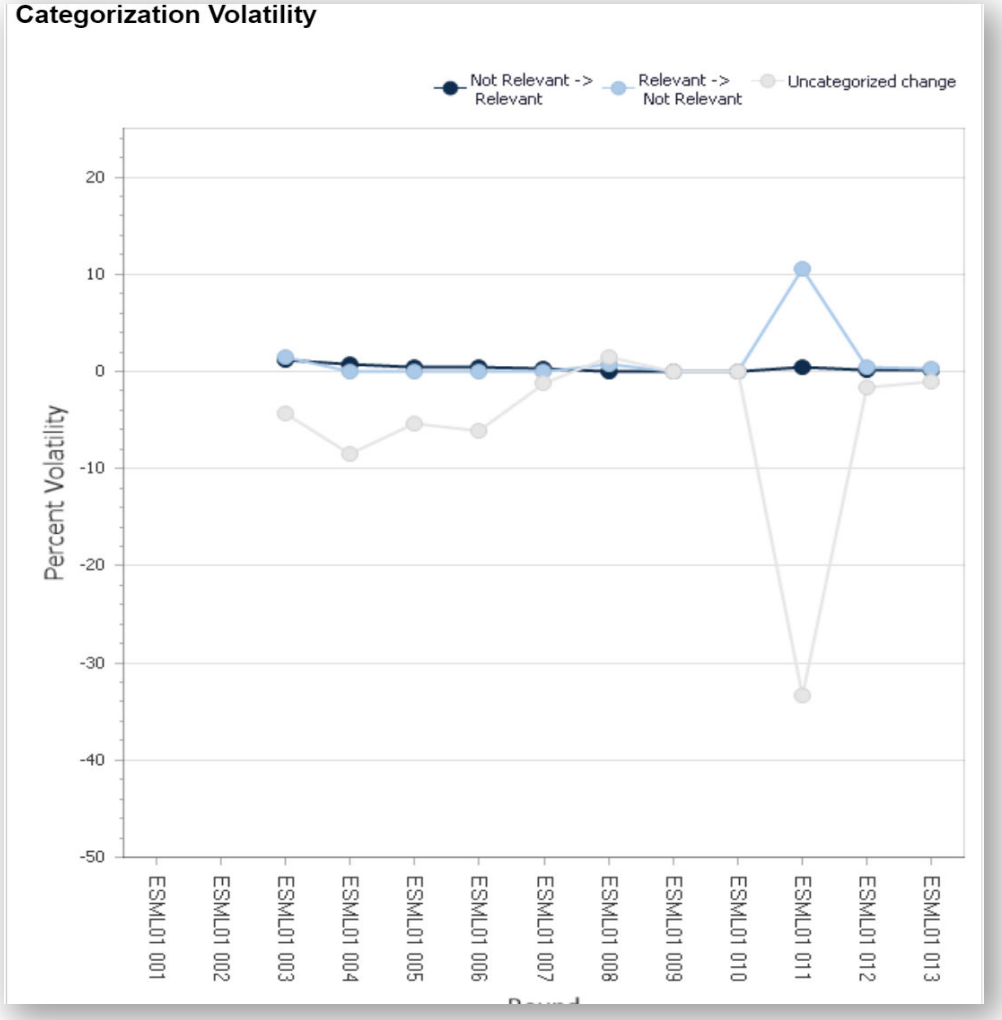
Control Round	# of Docs	Recall Goal	Confidence	Max Margin of Error	Estimated Richness	Additional Documents Recommended		
Control Round 1	712	75%	95%	10.53%	9.13%	3,185		
Control Round 2	1046	75%	95%	6.11%	10.98%	1090		
Control Round 3	1090	75%	95%	4.69%	11.52%	N/A		
Control Round 4	35	75%	95%	4.65%	11.55%	N/A		
Control Round 5	380	75%	95%	4.36%	11.62%	N/A		

Training Rounds	# of Docs	Classification Model	Consistency	Depth for Recall (%)	Depth for Recall (Docs)	Recall	Precision	F-Score
Training (Round 1)	200	INFLUENTIAL	98.00%	80.70%	81,661	75.00%	10.92%	19.07%
Training (Round 2)	600	FAST_ACTIVE	98.38%	37.30%	37931	74.70%	23.27%	35.48%
Training (Round 3)	800	FAST_ACTIVE	97.88%	29.90%	30450	75.61%	29.52%	42.47%
Training (Round 4)	799	FAST_ACTIVE	97.33%	26.40%	26841	75.00%	34.41%	47.17%
Training (Round 5)	800	FAST_ACTIVE	96.09%	28.50%	28739	75.00%	32.11%	44.97%
Training (Round 6)	800	FAST_ACTIVE	95.77%	27.50%	27584	75.00%	33.79%	46.59%
Training (Round 7)	250	DIVERSE_ACTIVE	95.41%	26.80%	26779	75.00%	34.75%	47.49%
Training (Round 8)	250	DIVERSE_ACTIVE	95.15%	24.10%	24120	74.70%	38.52%	50.83%

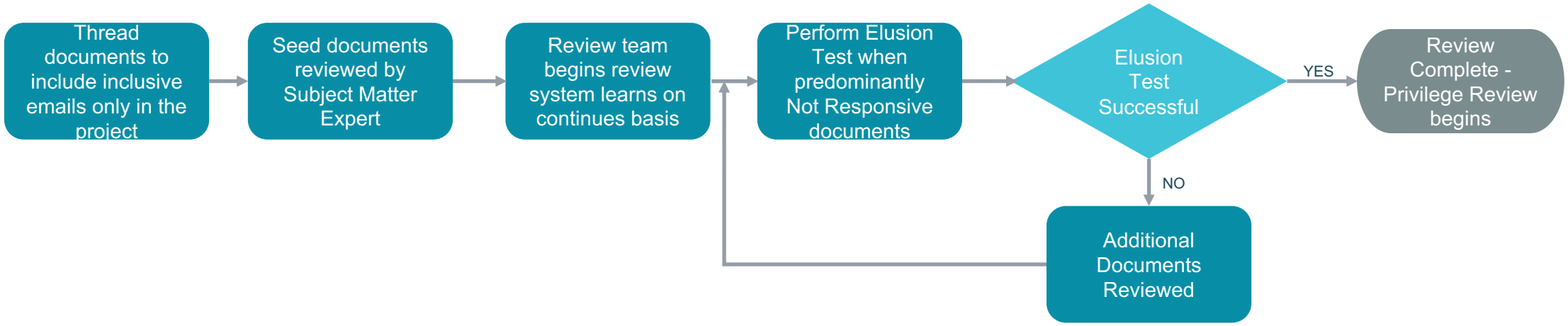
HAYSTACK

Reporting Relativity

Round name	Categorized Relevant		Categorized Not Relevant		Uncategorized	
	#	%	#	%	#	%
ESML01 001	0	0.00%	0	0.00%	2,934,899	100.00%
ESML01 002	380,213	12.95%	409,752	13.96%	2,144,934	73.08%
ESML01 003	397,318	13.54%	518,770	17.68%	2,018,811	68.79%
ESML01 004	669,385	22.81%	495,258	16.87%	1,770,256	60.32%
ESML01 005	837,995	28.55%	482,192	16.43%	1,614,712	55.02%
ESML01 006	1,030,837	35.12%	470,362	16.03%	1,433,700	48.85%
ESML01 007	1,069,193	36.43%	465,771	15.87%	1,399,935	47.70%
ESML01 008	1,002,812	34.17%	487,297	16.60%	1,444,790	49.23%
ESML01 009	1,003,142	34.18%	488,015	16.63%	1,443,742	49.19%
ESML01 010	1,003,142	34.18%	488,015	16.63%	1,443,742	49.19%
ESML01 011	799,566	27.24%	1,671,235	56.94%	464,098	15.81%
ESML01 012	799,159	27.23%	1,718,902	58.57%	416,838	14.20%
ESML01 013	795,473	27.10%	1,751,784	59.69%	387,642	13.21%



TAR 2.0 Workflow



HAYSTACK

Poll Question

Over the past year, what percentage of matters that have used TAR 2.0 employ a workflow where the learning algorithm is trained and the review is cut off prior to placing eyes on all responsive documents that are produced?

HAYSTACK

TAR 2.0 Considerations

- Coverage Review
- Prioritization Review
- Families
- Privilege
- Responsive Changes
- Cut off

HAYSTACK

Portable Models

Classifier: Compilation of terms and phrases that represent that is in the process of being refined using positive and/or negative example documents using either algorithmic or manual selection methods.

Predicative Rank: The output of training the classifier which results in a score between 0 and 100 for each document in the dataset where higher ranking documents are likely to be positive and lower ranking documents are likely to be negative.

Portable Models: Machine learning weighted key words that allow to create a predictive model from one matter and apply it to multiple other matters.

Examples and Uses: Investigation matters – employment, antitrust, FCPA; models to remove junk and auto replies; identification of key custodians across matters.

Benefits: accelerated review, minimizing resources, consistency, defensibility, security.

HAYSTACK

The Difference Between TAR 1.0 and TAR 2.0

TAR 1.0: "Predictive Coding"

1. One-time training before assigning documents for review. Does not allow training or learning past the initial training.
2. Trains against small reference set, limiting ability to handle rolling uploads; assumes all documents received before ranking. Stability based on training against reference set.
3. Subject Matter Expert handles all training. Review team judgments not used to further train the system.
4. Uses random seeds to train the system rather than key documents found by the trial team.
5. Doesn't work well with low richness/prevalence collections; impractical for smaller cases because of stilted workflow.

TAR 2.0: "Continuous Active Learning"

1. Continuous Active Learning allows the algorithm to keep improving over the course of review, improving savings and speed.
2. Ranks every document every time, which allows rolling uploads. Does not use a reference set but rather measures fluctuations across all documents to determine stability.
3. Review teams train as they review, working alongside expert for maximum effectiveness. SME focuses on finding relevant documents and QC'ing review team judgments.
4. Uses judgment seeds so that training begins with the most relevant documents, supplementing training with active learning to avoid bias.
5. Works great in low richness situations; ideal for any size case from small to mega because of flexible workflow.

Case Studies – TAR 1.0

Case	Documents Processed	Documents Reviewed to Train the Model	Documents NOT Reviewed due to TAR	Savings from TAR %
Case 1	2,973,164	11,929	2,961,235	**99.60%**
Case 2	257,122	8,487	248,635	**96.70%**
Case 3	636,523	3,372	633,151	**99.47%**
Case 4	79,346	15,464	63,882	**80.51%**
Case 5	6,539,175	7,423	6,531,752	**99.89%**
Case 6	1,168,910	2,466	1,166,444	**99.79%**
Case 7	174,192	2,450	162,450	**93.26%**
Total	**11,828,432**	**51,591**	**11,767,549**	**99.49%**

HAYSTACK

Case Studies - CAL

Case	Documents Processed	Documents Reviewed	Documents NOT Reviewed due to AL	Review Savings
Case 1	144,853	18,207	126,646	**87.43%**
Case 2	47,363	11,205	36,158	**76.34%**
Case 3	480,200	83,740	396,460	**82.56%**
Case 4	8,673	5,477	3,196	**36.85%**
Case 5	11,831	9,891	1,940	**16.40%**
Total	**692,920**	**128,520**	**564,400**	**81.45%**

HAYSTACK

What's Next in Analytics

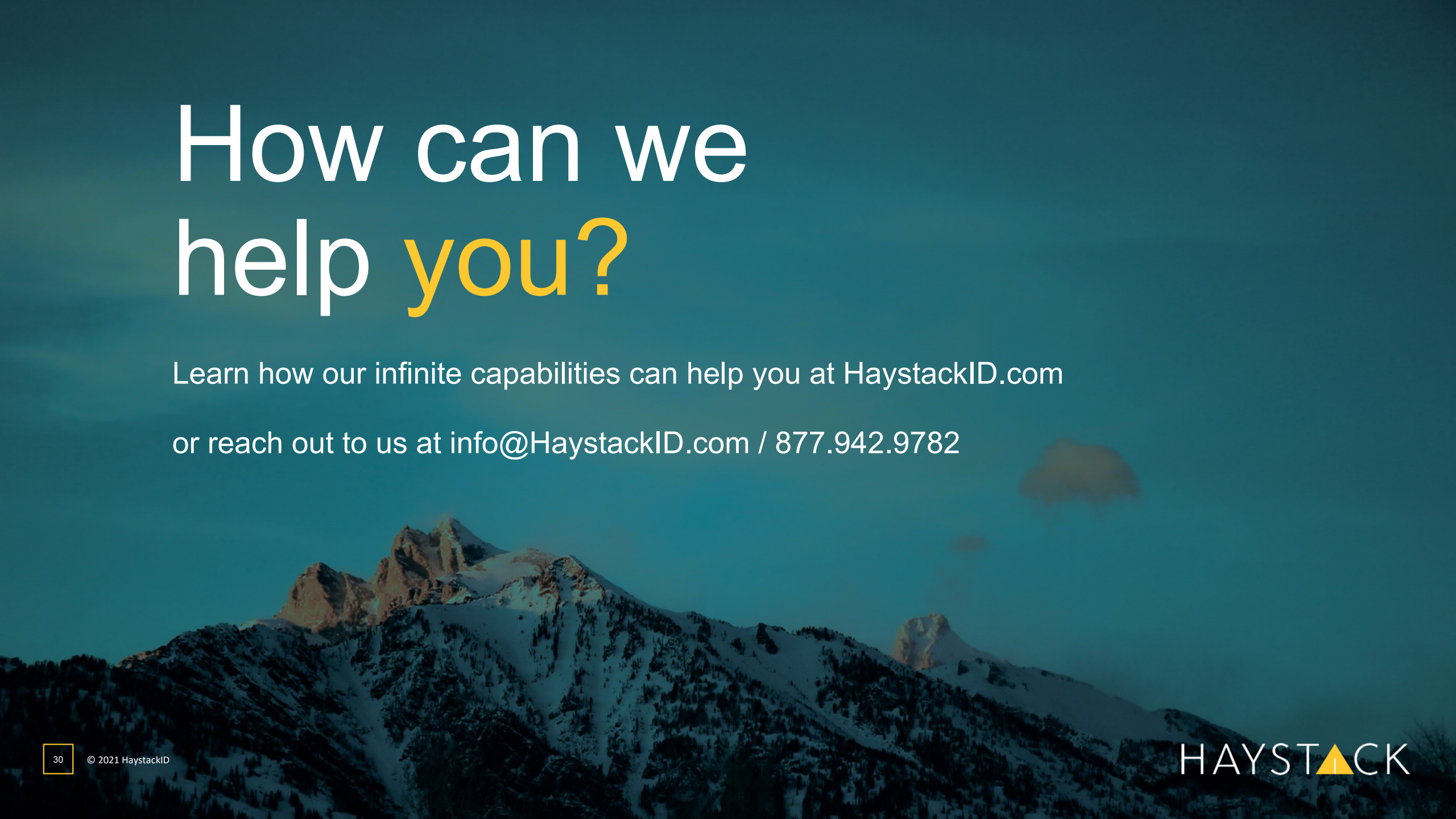
- Hybrid Model
- Information Governance & Application of Analytics
- Analytics for Forensic Assessment & Collection
- Sentiment Analysis & Emojis
- Analysis of Financial Data
- GDPR, PII & PHI

HAYSTACK

How can we help you?

Learn how our infinite capabilities can help you at HaystackID.com

or reach out to us at info@HaystackID.com / 877.942.9782



© 2021 HaystackID

HAYSTACK