Optimize Surveillance using Machine Learning

Contents

Introduction

Veritas Surveillance™ and Veritas Alta™ Surveillance are market-leading on-premises and SaaS communication surveillance solutions. They provide an end-to-end workflow for organizations that need to comply to regulatory requirements such as those from the Securities Exchange Commission (SEC), the Financial Industry Regulatory Authority (FINRA) and the revised Markets in Financial Instruments Directive (MiFID II).
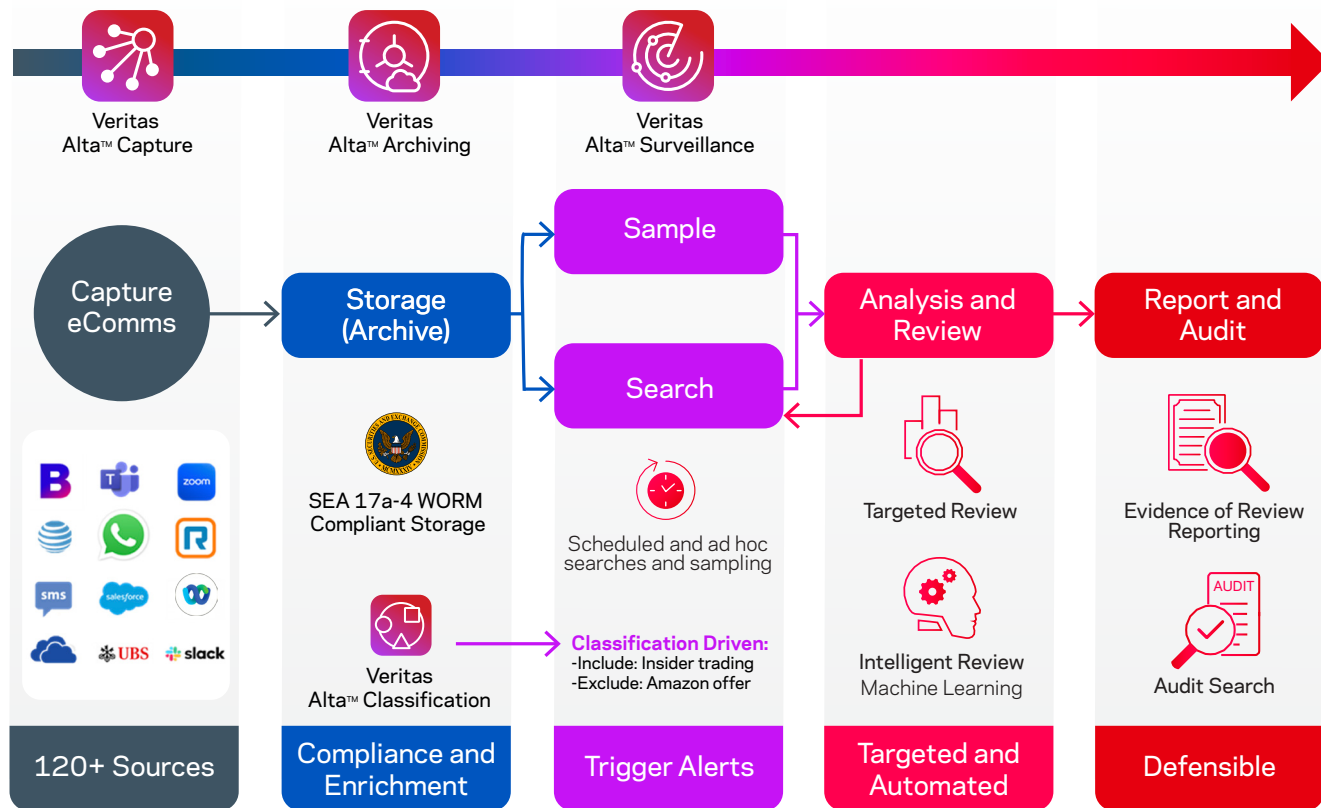
Veritas Surveillance is a data compliance solution that:

- Captures and searches data from more than 120 content sources

- Archives on-premises or in cloud, to ensure compliant storage of regulated content

- Performs sampling and search using machine learning-based classification, pre-trained policy-based classification, and keyword- and metadata-based criteria

- Delivers workflows to manage review teams and target specific employees for surveillance review

- Provides an intuitive web interface to conduct reviews and take required actions

- Offers extensive and detailed reporting and auditing of the compliance process

Traditional approaches to Communication Surveillance are no longer tenable. The ever-growing forms of e-communications such as IM chats, audio and video calls, along with emails, have drastically increased the volume of content needed to be reviewed. This puts organizations under a lot of strain, making it difficult for compliance teams to cope, and increases the likelihood of missing out on violations that should be flagged. The solution is to automate the classification of content using AI to mute non-relevant content—often referred to as noise, or false positives—while surfacing truly relevant content to the reviewers.

Veritas provides two kinds of automated classification engines, both supervised and unsupervised. Together, they coordinate to optimize, and modernize, surveillance workflows.

Classification

Unsupervised	Veritas Alta	Supervised ML	Intelligent
Rules-Based + NLP	Classification	Continuous Active Learning	Review

The unsupervised engine, Veritas Alta™ Classification, is a key component embedded with many Veritas solutions to deliver expert-trained, customizable, out-of-box policies to address a wide range of use cases in surveillance, corporate compliance, human resources, legal, privacy, public sector, and security.

Intelligent Review, the main topic of this whitepaper, provides supervised machine learning that automatically classifies new content for relevance, while providing full transparency into its predictions. The AI engine learns continuously from previous reviews conducted by your compliance teams to create and maintain its prediction models. Intelligent Review has seen the following recent enhancements:

- Improved prediction accuracy
- Consideration of message content (and not just metadata) for classification
- Relevancy score and Intelligent Review details about how messages are classified

In this white paper, we will highlight how Intelligent Review works, and how to extract maximum value to modernize and optimize your surveillance process.

Intelligent Review

Intelligent Review is a feature of Veritas Alta Surveillance that uses machine learning to classify content based on its relevance to review. Content is classified as relevant or irrelevant and as a result, sampling and search processes are optimized. The illustrations below will help you understand how sampling and review functions with and without Intelligent Review.
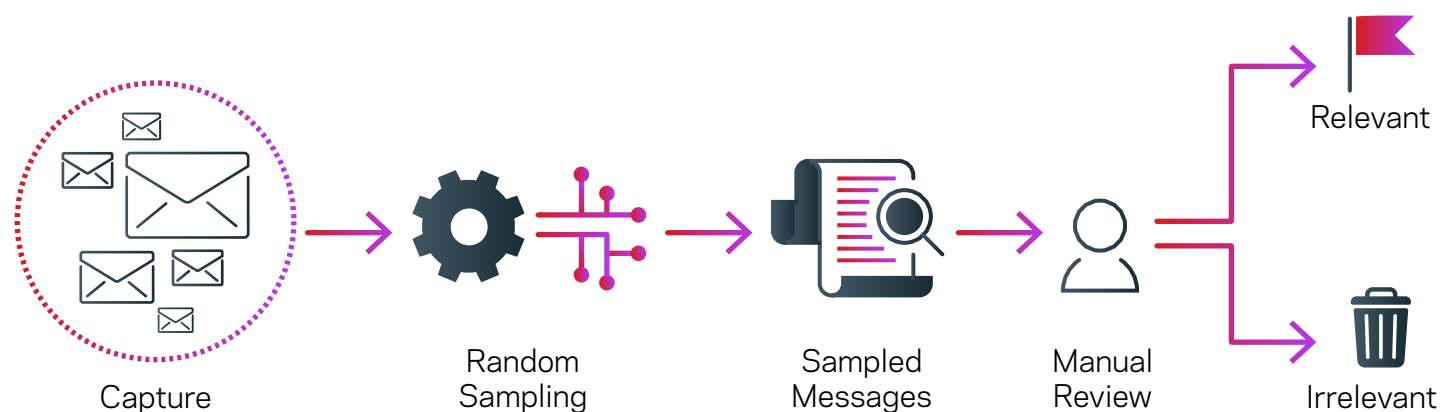

Capture → Random Sampling → Sampled Messages → Manual Review → Relevant / Irrelevant

Figure 3. Sampling without Intelligent Review
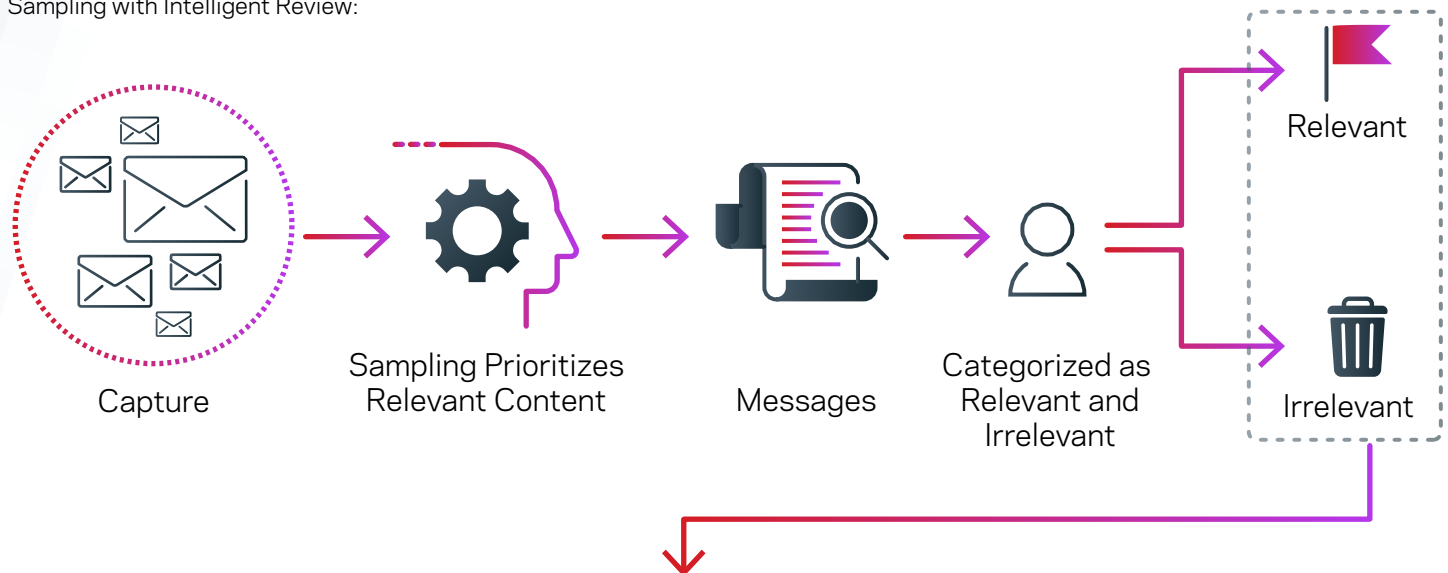
Sampling with Intelligent Review:

Figure 4. Sampling with Intelligent Review

Optimize Review with Intelligent Review

Intelligent Review learns passively by observing how reviewers are marking items relevant or irrelevant as part of their regular review. It does not require a discrete training regimen, and once enabled, it is always learning. It works optimally to classify content with high accuracy when certain guidelines are followed. Refer to the following sections for more details.

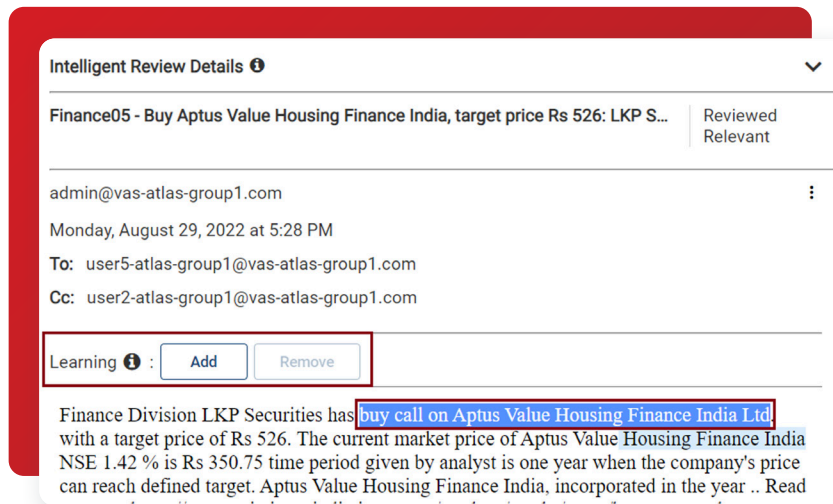**Training Intelligent Review**

Intelligent Review employs continuous learning, constantly tracking how reviewers are marking items as relevant and irrelevant, and it refreshes its models every day. It considers messages reviewed in the last 90 days (except if content snippets are added during review, in which case such messages are considered for 180 days from their review date). Intelligent Review uses the following metadata fields of the message along with content snippets (if any) selected by reviewers from the message body or attachment:

- Author

- Recipients

- Subject

- Message direction

  - Internal: Author and all recipients are part of the organization

  - External inbound: Author is external to the organization

  - External Outbound: Author is part of the organization, but one or more recipients are external

- Classification tags

Tags generated by Veritas Alta Classification based on Veritas Alta Classification policy rule matches.

- Content Snippets

Intelligent Review also provides an advanced feature that allows permissioned reviewers to select the portions of text from the message body or attachment that prompted them to mark the item as relevant or irrelevant. Once the snippets have been selected and the item marked as relevant or irrelevant, the selected text snippets will then be processed to remove stop words, numbers, and punctuation marks, and then used to train the engine. Text snippets should be as specific as possible, rather than selecting a large chunk of text, as this trains the system more effectively. Selecting content snippets is optional, and if not done, Intelligent Review considers only metadata for classification.

Understanding Intelligent Review's Output

Once sufficient items have been marked by reviewers to train the system, Intelligent Review starts to classify new messages as relevant or irrelevant. While classifying, it also produces a Relevancy Score and Intelligent Review Details which are explained below.

The time it takes to train the system can vary. Training starts the day a new department is created and items are reviewed. Departments that existed before intelligent review settings were configured (more on this later) are likely to see items classified by Intelligent Review very quickly, as it trains upon the review that has already occurred within that department over the past 90 days. New departments, or departments that have not conducted any review to date, will require a period of time (from a couple of days to a few short weeks) before Intelligent Review begins to classify messages. This typically requires review of a few hundred messages within the department.

Review Filters

Messages classified by Intelligent Review are added to the review set as Unreviewed Relevant and Unreviewed Irrelevant (if you have set up searches or samples to include only relevant content, you may not have items in Unreviewed Irrelevant, see more on this later). Reviewers can filter messages predicted as relevant or irrelevant from this interface. They can prioritize the review of relevant messages first, followed by the irrelevant ones. They can also take bulk actions on messages through the filters seen in the image below.
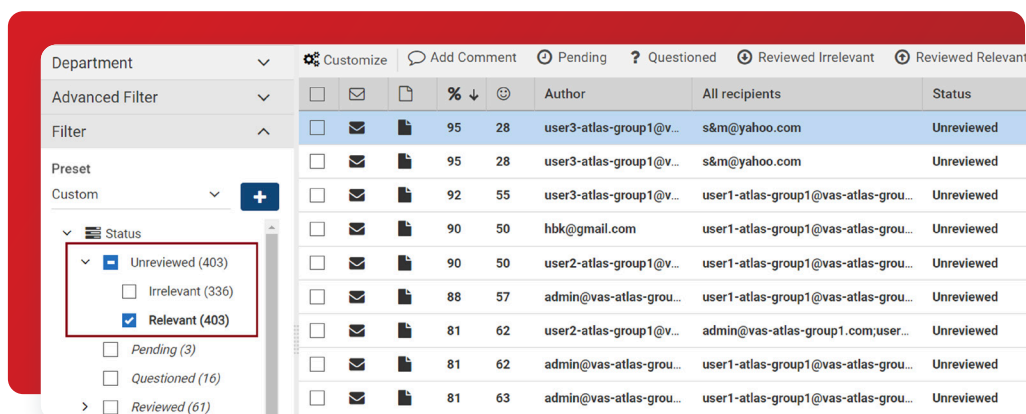
Relevance Score

The Relevance Score is a number between 0 and 100 assigned to all classified messages. It indicates the probability of a message being relevant. The higher the number, the more Intelligent Review predicts this message to be relevant. It is used in conjunction with the Relevance Threshold of the department to decide whether the message will be classified as relevant or irrelevant (more on this later). It is also available in the review grid, where reviewers can sort it to prioritize review based on relevance.

Intelligent Review Details

Intelligent Review Details are shown in the Message Preview pane within Veritas Surveillance. The details display how different fields (Author, Subject, etc.) of the message contribute to the overall Relevance Score by scoring a message's Relevant and Irrelevant contributions for each field. Any message may have some contributions that point toward relevant, and at the same time, there might be other contributions that point toward irrelevant. Depending on how these field contributions add up, Intelligent Review predicts relevance with a certain probability.

The range of scores for Relevant and Irrelevant are normalized between 0–100. Any given field may contribute to a Relevant score, an Irrelevant score, or both. Fields such as author and direction are single valued and may contribute to only one, either the Relevant or Irrelevant score. However, other fields such as subject, recipients, classification tags, and content, may contribute to Relevant and Irrelevant scores, as there may be more than one. For example, a message may contain two recipients, one strongly associated toward relevant messages, and another, perhaps weakly associated toward irrelevance. One may be displayed in the details for Relevant and other may be displayed in the details for Irrelevant.

Intelligent Review Details will show for subject and content, values for both single words, and bigrams (two words together). These represent words in the subject line, and words in the content, which have the most scoring impact.

For every field, Intelligent Review Details displays up to the top five contributions that impact the score of each contributing field. If a field does not contribute toward a particular direction, it will not be displayed.

Apart from the message's fields, Department Influence may also have an impact on prediction. It shows the extent to which prediction is inclined in favor of, or against, the marking based on the department's learning. If reviewers in a department tend to mark more messages as irrelevant, then Department Influence will contribute toward irrelevance.

If the total score of relevant contributions is more than the total score of irrelevant contributions, the message will a have higher Relevance Score (greater than 50). Otherwise, the Relevance Score will be less than 50.

Access to the Intelligent Review Details can be limited to specific reviewer roles using the Show Intelligent Review Details in Review permission (which is on by default for all of the reviewer specific roles).
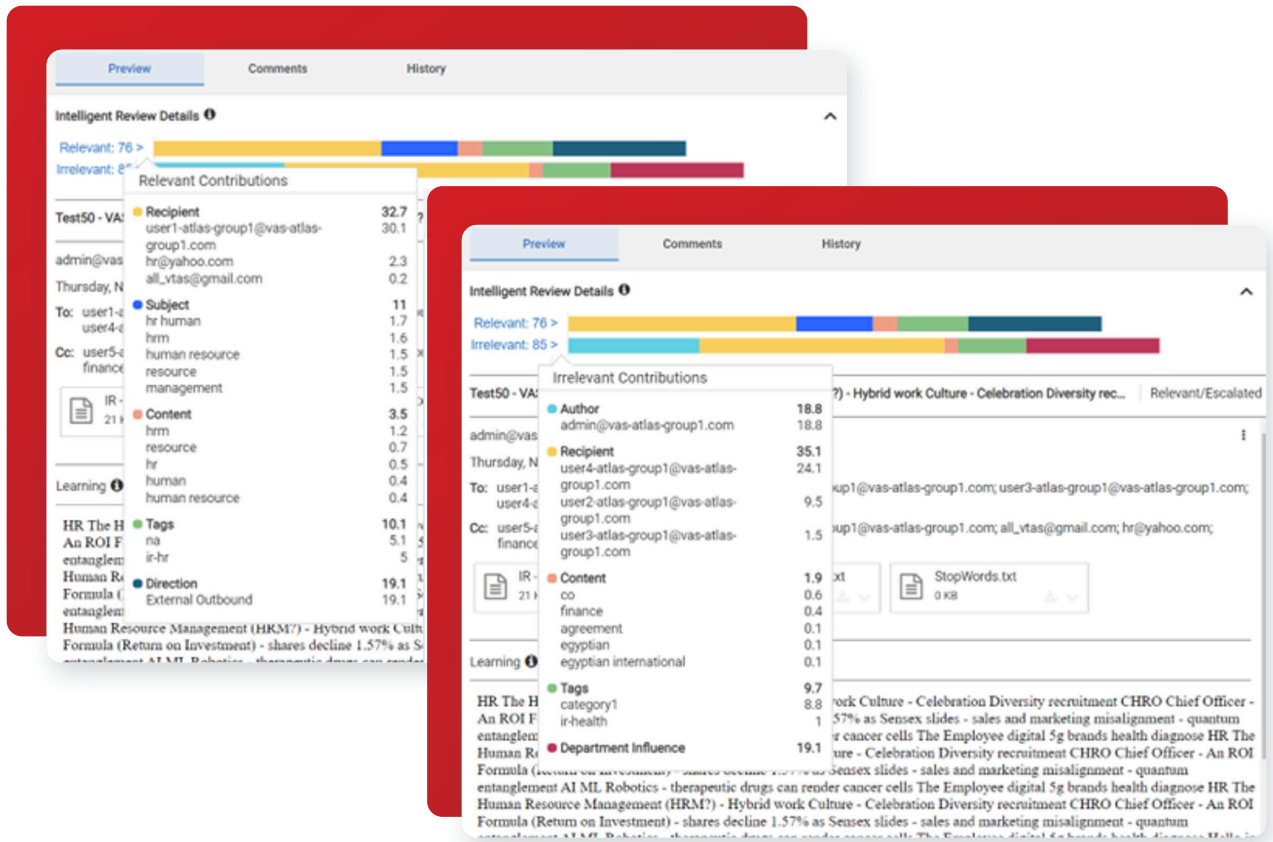
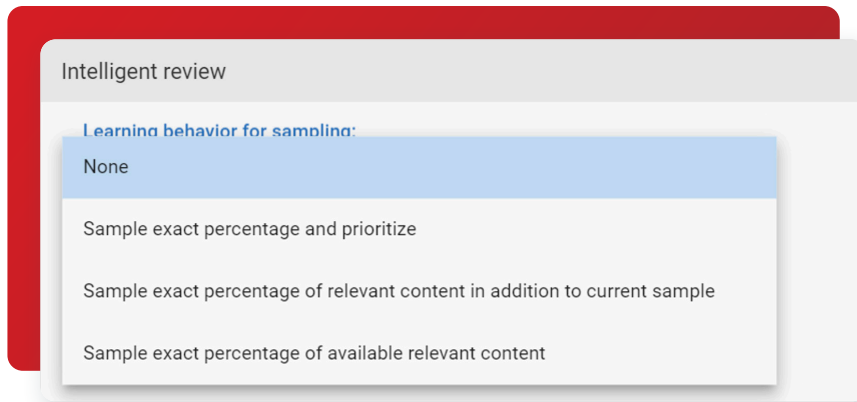## Configuring Intelligent Review

Department Level Configuration

For a department, Intelligent Review can be enabled when creating a new department or by using the Edit Department option.
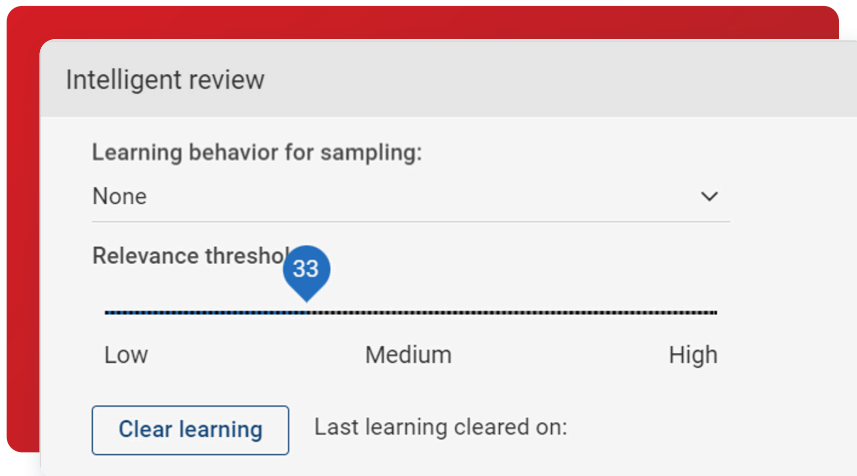
Learning Behavior for Sampling

This option determines how messages are randomly sampled considering their relevance during random sampling.

- **None**

  With this default option, Intelligent Review doesn't classify messages and the reviewer will not see items classified as Unreviewed Relevant and Unreviewed Irrelevant within Veritas Surveillance.

- **Sample exact percentage and prioritize**

  Intelligent Review classifies messages as relevant and irrelevant, but messages will be randomly sampled without considering their relevance. Reviewers will be able to see messages classified as Unreviewed Relevant and Unreviewed Irrelevant within Review. This is the recommended option for the organization starting with Intelligent Review, as it has no impact on which messages are brought into review, it simply adds messages to the Unreviewed Relevant and Unreviewed Irrelevant statuses, which can then be used to prioritize Review.

- **Sample exact percentage of relevant content in addition to current sample**

  Intelligent Review classifies messages as relevant and irrelevant, and guaranteed sampling adds Unreviewed Relevant and Unreviewed Irrelevant messages to the review set until it has captured the specified percentage of relevant messages that you have configured in the Department Properties. For example, if your Monitoring Policies within the Department Properties is specified to randomly sample 10 percent for All Policies, it will randomly sample relevant and irrelevant messages until 10 percent of all relevant messages have been randomly selected. With this option, the review set may contain more messages than your monitoring policy demands, since you are collecting irrelevant messages as well. Choose this option if you want a specified percentage of relevant messages but also want irrelevant ones for review.

- **Sample exact percentage of available relevant content**

  Intelligent Review classifies messages as relevant and irrelevant, and only relevant messages will be sampled. If there aren't enough relevant messages to fulfill the sampling percentage specified in the Monitoring Policies within Department Properties, then only irrelevant ones are sampled until the specified percentage is achieved. Unlike the previous option, the percentage sampled represents the total percentage of all messages and not just the percentage of relevant messages. Choose this option when you are confident about the accuracy of Intelligent Review's predictions.

Relevance Threshold

Relevance Threshold is a number between 0 to 100 that is used to classify messages as relevant based on the probability calculated by Intelligent Review. The correlation between this threshold value and probability is explained below:
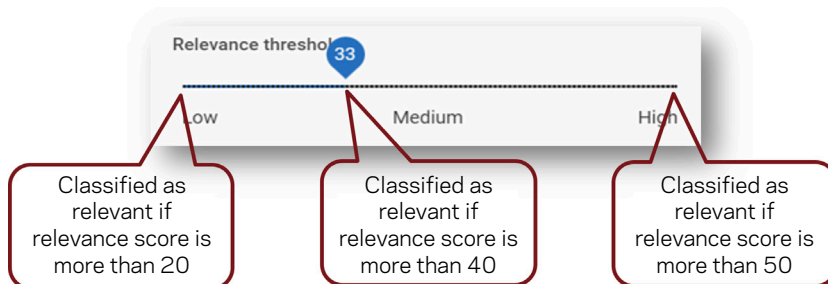
- At the default value of 33, a message will be classified as relevant when there is at least a 40 percent probability of it being relevant. If it is less than 40 percent, the message is deemed irrelevant.

- At the highest value of 100, the required probability to be classified as relevant is 20 percent. If it is less than 20 percent, the message is deemed irrelevant.

- At the lowest value of 0, the required probability to be classified as relevant is 50 percent. If it is less than 50 percent, the message is deemed irrelevant.

- During review, if you find too many messages are classified as Unreviewed Irrelevant, move the Relevance Threshold slider to the right. This will result in less messages classified as Unreviewed Irrelevant because Intelligent Review will have a higher threshold to meet before declaring a message as irrelevant.

- During review, if you find too many messages are classified as Unreviewed Relevant, move the Relevance Threshold slider to the left. This will result in less messages classified as relevant because Intelligent Review will have a lower threshold to meet before declaring a message as Irrelevant.

Clear Learning

This option can be used to reset the learning data. Use this when you want Intelligent Review to only consider new data that is reviewed after you have cleared the learning.

Search Level Configuration

When searches are used, Intelligent Review can be configured when creating a new search or editing an existing search.
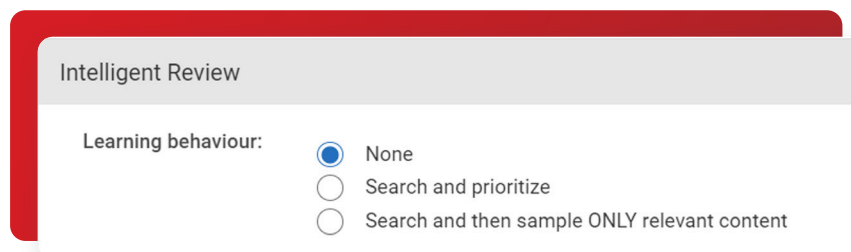
This option determines how messages searched are made available for review by also factoring in Intelligent Review classification.

- None
  Search results are not influenced by Intelligent Review classification. Reviewers won't see messages classified as Unreviewed Relevant and Unreviewed Irrelevant. Choose this option if you do not want Intelligent Review.

- Search and prioritize
  Search results are not influenced by Intelligent Review classification. However, reviewers will see messages classified as Unreviewed Relevant and Unreviewed Irrelevant. Choose this option if you do not want to add or remove messages because of Intelligent Review, but you want to use the Unreviewed Relevant and Unreviewed Irrelevant facets during review for prioritization.

- Search and then sample ONLY relevant content
  Search results are influenced by Intelligent Review classification. Only messages that are classified as relevant and meet all other search criteria are made available for review once the search is accepted. Choose this option if you only want messages classified as relevant added to review.

Permissions

You can manage the access to some advanced features provided by Intelligent Review. Select roles by assigning the following permissions:

Add or Remove Content Snippets

Users with a role with this permission can add or remove content snippets from the message body or attachment. These snippets are then used to train Intelligent Review.

Show Intelligent Review Details in Review

Users with a role with this permission can view the Intelligent Review details while reviewing messages. These details show why a message is flagged as relevant or irrelevant.

Conclusion

As this white paper demonstrates, Intelligent Review is a cutting-edge Machine Learning tool built into Veritas Surveillance that automatically reduces noise and surfaces relevant content for your review teams. It does this using a continuous learning methodology that requires no discrete training regimen or periodic re-training of the AI engine. Turn it on and it just works. It also leverages the best in AI explainability with its Relevancy Scoring and Intelligent Review Details. Use Intelligent Review to:

- Enhance review sets by prioritizing relevant content over irrelevant content
- Expedite the review process by laser focusing on the relevant content first
- Find more violations early from a more concentrated and smaller review set without noise
- Avoid regulatory penalties that could cost millions

Documentation

Veritas Surveillance User Guide

References

- Veritas Alta Surveillance
- FINRA 3110
- FINRA 3170
- SEC Charges 16 Wall Street Firms
- Powerhouse Banking Firm Fined for Supervision Deficiencies