

# Classification Review

Train Proofpoint's Intelligent Classification and Protection AI engine to deliver the highest quality results tailored to your needs

# Overview

Many organizations use manual processes to identify, classify and protect their data. As data volumes continue to swell, however, this approach is becoming less and less viable. Manual processes often cannot keep up with today's ever growing workloads. And they are prone to human error. To secure data in this day and age, you must be able to ensure speed and accuracy even at scale.

To this end, Proofpoint has developed revolutionary artificial intelligence (AI)-based technologies that can modernize and enhance your data security posture. They help you solve some of the most daunting problems associated with the management of data security. They also provide state-of-the-art security but at a fraction of the cost of other solutions or manual approaches. For each hour worked, for example, our AI can analyze as many documents as a human can in 10,000 hours. And the AI is three times more accurate on average. However, although AI can function independent of human activity, it first needs to be trained. And to ensure results with the highest quality, performance and explainability, humans must be in the AI's learning loop. That is where Classification Review comes in.

Classification Review is an integral part of Proofpoint Intelligent Classification and Protection, our AI-powered data discovery and classification solution. It comes as a streamlined module that can easily be integrated into your own processes. Classification Review lets you validate the results generated by the solution's AI engine. But it does more than just ensure quality. It also lets you custom-tailor your solution by aligning on asset classification policies.

# Table of Contents

<b>1</b>	<b>Setting Up the Classification Review</b> . . . . .	4
	Alignment on the asset list . . . . .	4
	Alignment on the process . . . . .	4
	Alignment on user involvement . . . . .	5
<b>2</b>	<b>Sampling</b> . . . . .	5
<b>3</b>	<b>Interface</b> . . . . .	6
<b>4</b>	<b>Retraining the Classification Models</b> . . . . .	7
<b>5</b>	<b>Performing the Classification Reviews</b> . . . . .	7
<b>6</b>	<b>Conclusion</b> . . . . .	8

---

# Setting Up the Classification Review

## Alignment on the asset list

Before you even start the Classification Review, you and Proofpoint must be aligned on the asset list to leverage. This list defines the classification dimensions of your data. These dimensions may include things like business category (Finance, Client Data and Technology) and confidentiality level (Internal and Confidential).

You have three options to define this asset list:

- You can leverage your own asset list to classify documents.
- We can help you to define a customized list that leverages your own asset list as well as Proofpoint's key categories.
- If you do not have your own asset list, we can provide you with one that includes a default group of categories specific to your industry.

## Alignment on the process

Once you and Proofpoint are aligned on data classification, we must also align on the goals of the review. When you define the process clearly, you'll get a smooth review that is tailored to your needs.

We must define the following before you can kick off the Classification Review:

- **The expected improvement in accuracy after classification review.** Reviewing more documents may be slower, but you will see more improvement in performance. Reviewing fewer documents may be faster, but the improvement in performance won't be as significant.
- **The departments that should take part in the review.** You can involve one (or a few) departments at a time or you can organize a company-wide Classification Review campaign with all departments.
- **The folders and data sources that the user reviews should target to focus on business-critical data.**

## Alignment on user involvement

We help you find the right balance between the quality premium offered by the Classification Review and the time that reviewers can afford on the exercise.

- Our Active Learning module can determine the optimal number of documents to review to achieve your expected improvement in accuracy. This is typically around 20 documents per category.
- We can also recommend the most appropriate set of reviewers. This group is often the smallest set of high-risk users who have access to recent business-critical data. You'll typically need 10 to 15 reviewers per campaign.

On average, a document takes three minutes to be reviewed.

“A good human plus a machine is the best combination.”

—Garry Kasparov, Russian chess grandmaster and former world chess champion

---

## Sampling

We leverage the following key principles to choose the most significant documents to sample and review:

- **Diversity.** Machine learning (ML) models are very good at recognizing patterns. So, an ML model will perform better if it can see the most diverse set of documents possible.
- **Uncertainty.** ML models do not only predict the labels of documents. They can also determine how certain they are in their predictions. This lets you select for review the documents that the model is most uncertain about. These documents will often come from underrepresented classes.

Proofpoint has also developed an access-based assignment of documents. This means that reviewers only receive documents they have been granted access to, instead of documents from a particular category.

# Interface

Classification Review’s user interfaces are seamlessly integrated with your processes to facilitate efficient reviews. The interfaces include previews of the documents to review.

The interface for data owners (or project managers) lets them set up review campaigns and monitor their progress. Data owners oversee and protect a data domain. And to ensure that all reviewed documents are classified correctly, data owners are in charge of validating the proposed classification. They can also challenge it if they think another label would be more appropriate.

Classification Review also has an interface for key users who will review documents. Key users are high-risk profile individuals who have access to a large volume of the most sensitive data.

Pending 66/75

Category Confidence Score  
(Proofpoint prediction)

IF File	IF Language	IF Category	IF	IF Confidentiality
Table	English	Corporate Entity > Audit - Internal > Reports Previous: Corporate Entity > Clients > Contracts	-	Secret Previous: Top Secret
Slides	English	Corporate Entity > Clients > Contracts	98%	Top Secret
Text Document	French	Corporate Entity > Clients > Contracts	98%	Secret Previous: Top Secret
Table	English	Corporate Entity > Clients > Contracts	98%	Top Secret
Table	English	Corporate Entity > Clients > Contracts	98%	Top Secret
Text Document	English	Corporate Entity > Audit - Internal > Reports Previous: Corporate Entity > Clients > Contracts	-	Top Secret
Text Document	French	Corporate Entity > Clients > Contracts	98%	Secret Previous: Top Secret

Figure 1: The Classification Review interface.

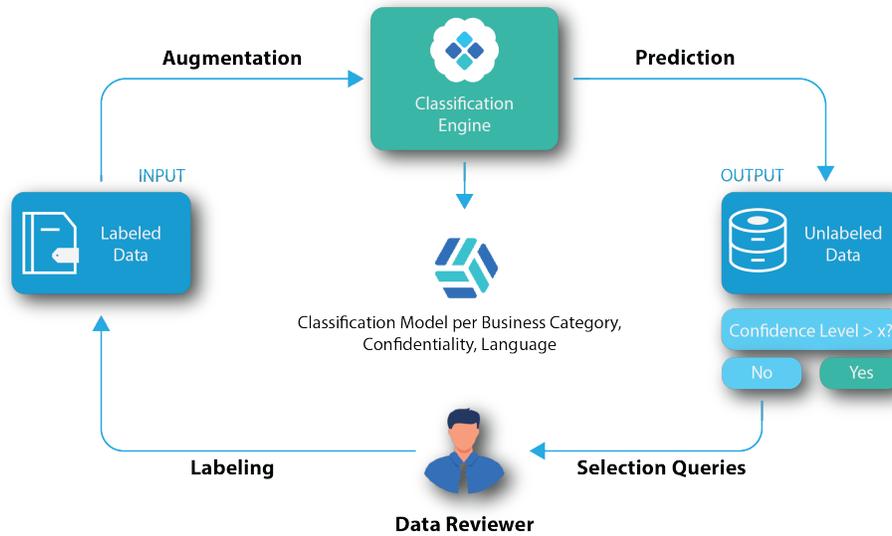


Figure 2: The Classification Review process.

## Retraining the Classification Models

Retraining classification models after Classification Review leads to a significant increase in performance. Doing so can drive up the accuracy by as much as 15%.

The table below shows four real examples that showcase the performance increase after model retraining.

## Performing the Classification Review

After its setup, the Classification Review exercise follows a simple four-step iterative process that usually lasts about two weeks. The process is as follows:

1. Once your data is classified, Proofpoint uses Active Learning to sample a set of the most significant documents. These documents have the highest impact on the classification model's quality.
2. The key users receive the documents to review. They can validate Proofpoint's predictions or they can correct them.
3. Reviewed labels are then propagated to the most similar documents. This improves the effectiveness of the review.
4. Proofpoint's classification model retrains itself and yields optimal results by leveraging the new information.

To ensure consistent accuracy of the models, we recommend that you organize a Classification Review campaign every nine months to a year.

### Performance Increase per Classification Review Campaign

	CLIENT 1	CLIENT 2	CLIENT 3	CLIENT 4
No. of Reviewers	13	6	10	14
No. of Reviewed Documents	315	190	360	1200
Business Category Performance Increase	+18%	x	+16%	+15%
Confidentiality Performance Increase	+33%	+19%	+11%	x

---

## Conclusion

Through its Classification Review module, Proofpoint leverages AI to allow data owners and key users to validate the results of document classification. More specifically, Classification Review aims to:

- Validate the accuracy of Proofpoint's classification module
- Improve the performance of the classifier
- Align Proofpoint's predictions with your own internal policies

This offers unprecedented performance. It also allows you to improve the quality of the classification mechanism over time as it evolves with your own data. And it lets you tailor the AI mechanism to your own asset list by aligning any prediction with your own internal policies.

### LEARN MORE

For more information, visit [proofpoint.com](https://www.proofpoint.com).

---

#### ABOUT PROOFPOINT

Proofpoint, Inc. is a leading cybersecurity and compliance company that protects organizations' greatest assets and biggest risks: their people. With an integrated suite of cloud-based solutions, Proofpoint helps companies around the world stop targeted threats, safeguard their data, and make their users more resilient against cyber attacks. Leading organizations of all sizes, including 75 percent of the Fortune 100, rely on Proofpoint for people-centric security and compliance solutions that mitigate their most critical risks across email, the cloud, social media, and the web. More information is available at [www.proofpoint.com](https://www.proofpoint.com).

©Proofpoint, Inc. Proofpoint is a trademark of Proofpoint, Inc. in the United States and other countries. All other trademarks contained herein are property of their respective owners. [Proofpoint.com](https://www.proofpoint.com)