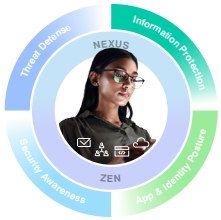
Proofpoint Automate

Streamline review processes and reduce false positives through advanced machine learning

The digital communications revolution is spawning more content—and more types of content—than ever before. This information deluge is driving a revolution of its own. Artificial intelligence (AI) and machine learning (ML) are set to transform compliance and e-discovery review. But while compliance, surveillance and legal teams are inspired by the promise of AI, they are struggling with the reality. Many of these technologies are inefficient. Some just aren't effective. And most don't scale

Proofpoint Automate can help. As part of the Proofpoint Digital Communications Governance family of products, it augments the rich detection and analytics features of Proofpoint Supervision to further streamline and improve your review process.

This solution set is part of Proofpoint's integrated human-centric security platform, mitigating the four key areas of people-based risks.



Monitoring communications, as required in financial services and other regulated sectors, has always been a balancing act. Companies must cast a wide net to find all relevant compliance issues. But casting the net too widely results in more false positives—and heavier workloads.

Proofpoint Automate is an add-on for Proofpoint Supervision that uses advanced ML to greatly reduce your reviewers' workload. You can easily deploy, validate and implement Automate models (the code used to evaluate communications of all types) to improve supervision decision-making and automate key parts of your workflow.

Unlike standalone ML tools, Proofpoint Automate builds upon the already-rich workflows and reporting you get from Proofpoint Supervision. By adding Proofpoint Automate, your supervision and surveillance efforts can become even more efficient and effective, further lowering costs.

Real-world outcomes

In lab tests, false positives have been reduced by up to 84% when using Automate with the Flag Deduplication feature. Of course, results will be different for every customer; AI models, data sets and training can vary widely by company. But Proofpoint Automate customers have reported the following real-world outcomes:

- A 25–50% reduction in false positives
- Ongoing ROI of 125% per year

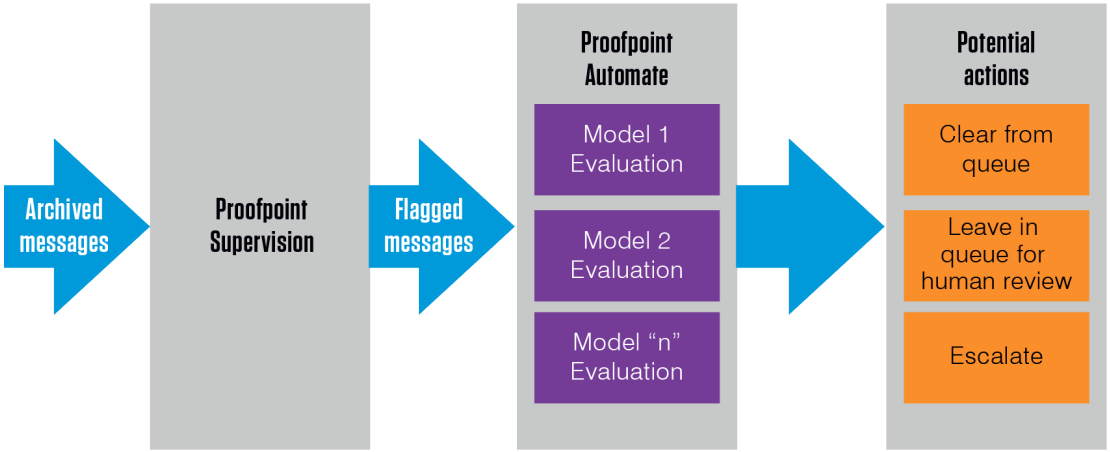


Figure 1: Workflow of Proofpoint Supervision with Proofpoint Automate.

Proofpoint Automate models have already helped identify the following improvements for early customers:

- **Disclaimer detection.** Suggests disclaimer text to add to Supervision's configuration for pre-approved content
- **Exclusion detection.** Suggests senders and subjects, such as newsletter senders, to add to Exclusion Rule configuration
- **Auto review.** Secondary evaluation of flagged messages with the ability to intelligently clear or escalate the messages as required

And these are just the start. Proofpoint Automate's open, scalable AI engine can tackle more complex and varied classes of communications-related decisions. These include detecting fraud, harassment, discrimination, coercion and other policy breaches. With Proofpoint Automate, your Proofpoint Intelligence Compliance deployment can make an even bigger business impact.

Models to automate and refine supervisory review

With traditional compliance tools and static rules, supervision review queues can quickly fill up with low-risk content. Boilerplate text, standard disclaimers and automated notifications are just a few examples. Reviewing it all takes time and resources. For already-stretched compliance teams, that means higher costs and longer time to value.

Proofpoint Automate refines your review queues, intelligently winnowing out low-risk content to prevent false positives in future reviews. With other compliance review tools, identifying new disclaimers and newsletter senders can take weeks or even months. Automate shortens this cycle, thanks to ML models that enable much faster review decision-making.

Proofpoint Automate removes the need for human action in the process of clearing or escalating secondary content reviews. Reviewers can then focus on the most business-critical content without worrying about missing out on the entire scope of data.

Technical view: Models and review automation

Proofpoint Automate uses ML models to evaluate messages flagged for review by Proofpoint Supervision. (See Figure 1.) These can be any message type supported by Proofpoint Archive, such as:

- Email
- Social media (such as Twitter or LinkedIn)
- Collaboration platforms (such as Slack or Microsoft Teams)
- Mobile (such as WhatsApp or voice)

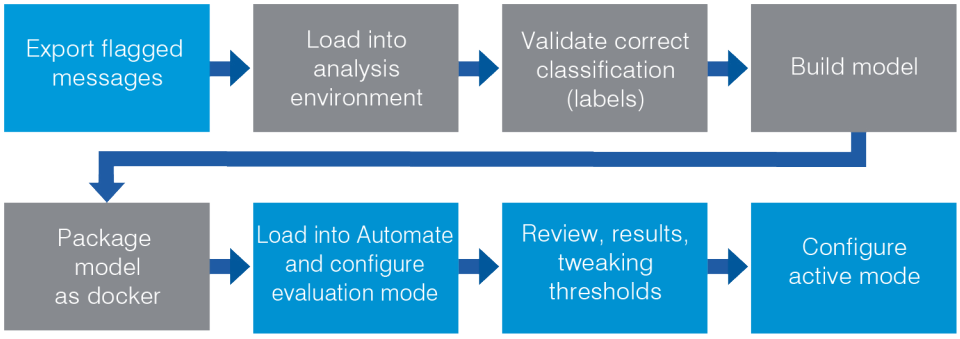


Figure 2: Proofpoint Automate training and deploying ML models.

Models are configured for items flagged for one or many rules. Each model evaluates the relevant items and assigns a score. That score is then added to the messages within Proofpoint Supervision.

In evaluation mode, reviewers still review messages; reporting enables you to compare between model-driven and human-driven review decisions. Once you have high enough confidence, you can activate the model to take the same actions as a human reviewer would. In practice, it's automated, autonomous review.

For flagged messages with corresponding Automate models, the scoring ranges and actions may be set up as follows within Proofpoint Supervision:

- **0.0 to 0.15.** High confidence that the flagged item is not an issue. Automate clears the issue.
- **>0.15 to <0.8.** Low confidence in the message evaluation. Automate leaves the message in the queue for human review.
- **0.8 to 1.0.** High confidence that the flagged item is an issue. Automate escalates the message for further investigation.

When the models are activated, evaluations, actions and related data are all recorded to the same standard as if a human reviewer is making the decision.

Key attributes of automate models

Like most ML models, Proofpoint Automate's are developed using data to train ML algorithms.[1] Before any model is produced, the data first needs to be prepared. (This step is sometimes called data wrangling.) The data is gathered, prepared and divided into two sets: a training set and a testing set. At this stage, the data must be classified and labeled to help describe the meaning a human would derive from it.

Proofpoint Supervision provides a natural mechanism for labeling data—simply train your staff to properly categorize issues as they review them. The message data (including metadata about which flags matched and why) and review decisions can easily be exported from Proofpoint Supervision. (The format is well suited for building Automate models.)

With the data ready, you have a baseline model. You can measure the effectiveness of the model by testing it on the test data set. Effectiveness is typically measured using two scores: precision and recall.[2]

In simple terms, precision is the percentage of the results correctly classified as being of interest. Recall is how many messages were correctly identified (that is, not missed).

1 Javatpoint. "Machine learning life cycle." Accessed December 2024.
2 Wikipedia. "Precision and recall." Updated December 2024.

Training and deploying models

With the baseline model complete, you can train and compare new models to see which ones offer better precision and recall scores. Sometimes, the two scores may be a trade-off. Recall may be more important in some situations, while precision may matter more in others.

To be used in the Proofpoint Automate framework, a model is packaged as a web service that implements our REST API specification and is delivered in a Docker container.[3]

The model can then be set up to operate one of these ways in production:

- **Evaluate.** The model evaluates every message in a particular class but takes no action. This mode allows the model to be tested in production on real-world data flows without performing actual reviews.
- **Enabled.** The model evaluates communications in a class and acts on them. You can also specify a sampling rate to evaluate a sample of messages but take no action on them. This approach enables discrepancy reports that compare model performance against that of human reviewers.

This support for running new and updated models offers two main benefits. First, it lets you evaluate them effectively. Second, it provides the evidence of successful testing you may need under validation mandates. You can compare updated models to both human reviewer decisions and those made by the model.

Flexible Ecosystem for Model Execution

Proofpoint Automate is a highly flexible platform. It runs models written in any language, using virtually any analytics framework that fits your needs.

You can use the out-of-the-box models directly. Your data science team can also build new models that run within the framework. This flexibility makes use of your team's unique knowledge of the business—without the work of building and supporting a scalable platform from the ground up. If you need more support, Proofpoint professional services can also help you create custom Automate models.

Conclusion and next steps

Proofpoint Automate achieves the promise of AI and ML at scale. Our unique framework helps streamline and improve the way you supervise digital communications to comply with regulations and corporate mandates. It deploys quickly, makes training and testing easy and offers an unmatched return on investment. With Automate, you can safely modernize your supervision practices for new levels of efficacy and efficiency.

3 Docker. "What is a Container?" Accessed December 2024.

Learn more at proofpoint.com