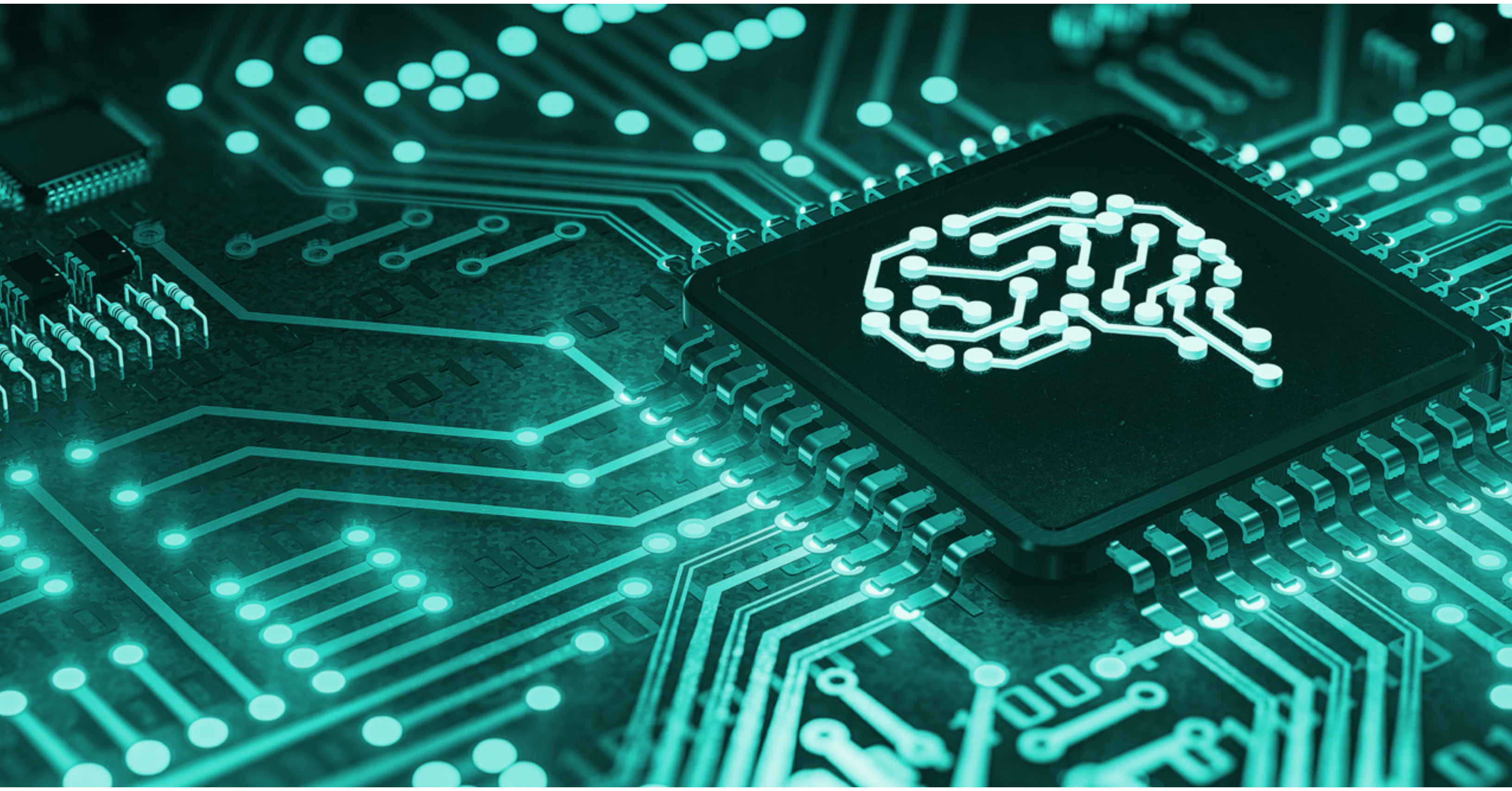
netskope

WHITE PAPER



Protecting Data Using Machine Learning

INTRODUCTION

Every day the Netskope Security Cloud processes billions of events and files, capturing a wide variety of user activities in:

- SaaS applications, such as Microsoft Office 365, Box, Salesforce, or G Suite
- Public cloud infrastructure services, such as Amazon Web Services, Microsoft Azure, and Google Cloud Platform
- Websites that users have visited

Netskope also has a comprehensive understanding of all enterprise data stored and transacted in the cloud. For example, Netskope has intimate knowledge of how a file was uploaded, downloaded, or shared within a managed cloud storage app and transferred to unmanaged cloud apps and personal devices. This contextual understanding of user activities and corporate data, coupled with a complete view of cloud and web traffic, enables Netskope to protect the sensitive data an organization owns or its employees use.

Netskope's ability to reliably distinguish specific, typically sensitive data from all other data within an environment means that customers can achieve the highest levels of accuracy in detection, and avoid the pain of false-positives. Netskope's continuous effort to improve accuracy and efficiency in data classification has driven our investment and development in artificial intelligence and machine learning.

Artificial Intelligence and Machine Learning 101

Artificial intelligence (AI) is an umbrella term for a computer program that behaves similarly to the human brain. AI is not new, having been studied since the 1950s. However in recent years AI's ability to "learn" and "problem solve" has been applied to almost every sector of the technology industry, including cybersecurity.

Machine learning (ML) is a subset of AI, where software programs are built to learn from examples. Within cybersecurity, an example of ML would be a program that has learned to categorize spam email through being exposed to many emails that are known to be spam. ML uses sample data or "training data" to build a mathematical model to make predictions or decisions about future input data.

A subset of ML called deep learning which tries to operate as the human brain does; the adjective "deep" reflects the multiple layers used in its artificial neural network. Within the past decade, deep learning and artificial neural networks have set new records in accuracy for many important problems, such as image recognition and natural language processing.

Artificial Intelligence and Machine Learning at Netskope

A leader in cloud security, Netskope is integrating the latest AI/ML technology into its data and threat protection features, as well as its business operations. Our team of dedicated data scientists, security researchers, and engineers with track records of solving security and fraud problems in different domains; together, this team has developed more than 100 patents. Leveraging this expertise in AI/ML and security, Netskope is developing large-scale AI/ML solutions for cloud security. Customer use cases for which Netskope uses AI and ML include:

- User and entity behavior analytics (UEBA) to detect malicious insiders, compromised accounts, brute force attacks, and data exfiltration attacks
- Detection of malware using ML models as a complementary approach to anti-virus signatures, threat intelligence, heuristics, and sandboxes
- Categorization and detection of malicious web domains, URLs, and web content

And finally, the subject of this paper:

- Helping organizations adhere to compliance regulations such as GDPR, CCPA, PCI, HIPAA; or their own internal data protection policies; through the detection of sensitive information in documents, images, and other data exchanged with the cloud and web.

Data Loss Prevention

Netskope's award-winning cloud-based data loss prevention (DLP) has always excelled at protecting customers' data. The Netskope Security Cloud provides organizations with numerous capabilities designed to reduce cybersecurity risk. In particular, Netskope is deployed to help an organization protect the sensitive data it owns or its employees process. Netskope understands the context of cloud and web access, including the user, device, app, instance, activity, and content involved, to accurately identify violations and data risks. From there, it can then allow, challenge, block, quarantine, encrypt, or apply a legal hold, as well as integrate with on-premises solutions to prevent data loss and exposure.

Our customers typically use its Next Generation Secure Web Gateway (NG SWG) to inspect data moving in real-time to and from cloud applications and websites, or they use API-based cloud access security broker (CASB) capabilities to examine data at rest in SaaS applications (such as Microsoft Office 365) or IaaS services (such as AWS S3). Regardless of whether data is inspected inline in real time or when it is stored at rest, Netskope performs accurate inspection through 3,000+ out-of-the-box data identifiers, 25 predefined legal and regulatory compliance templates, and various matching techniques (proximity expression, custom regex and dictionaries, file fingerprinting, exact data matching, and so on).

MACHINE LEARNING FOR DLP

File classification using machine learning provides a fast and effective way to identify sensitive information, enabling users to work inline with granular real-time DLP policy controls. ML classifiers are able to accurately classify documents and images into different categories, such as tax forms, patent documents, source code, passports, driver's licenses, payment cards, screenshots, etc., without the need to identify specific pieces of sensitive information contained in those files. Security administrators can then create DLP policies based on these categories. In this way, the ML file classification works as a complementary approach to traditional DLP rules and makes sensitive PII, PCI, and HIPAA data more secure.

Image classification

Image classification is a key part of DLP and keeps track of specific sensitive data within images that organizations may be storing or sharing. The traditional approach to identifying sensitive data in an image, for example, a patient's information in a medical X-ray image, has been to use optical character recognition (OCR) to extract text from the image. The extracted text is then used for regex or exact matching. OCR technology, although effective, is resource intensive and delays the detection of security violations. OCR also has difficulties identifying violations in low-quality images. In many cases, it is only necessary to determine the classification of the input image. For example, to identify whether an image is a passport or not, without needing to extract the person's name and other details in the image. As a result, ML-enabled image classification presents a more cost-effective, accurate, and secure option using deep learning and convolutional neural networks (CNN). We can also combine image classification with OCR to generate more detailed violation alerts.

Deep learning and CNN were huge breakthroughs in image classification in the early 2010s. Since then, CNN-based image classification has been applied to many different domains, including medicine, autonomous vehicles, and cybersecurity, with accuracy close to that of humans. Inspired by how the human visual cortex works, a CNN is able to effectively capture the shapes, objects, and other qualities to better understand the image's contents. A typical CNN, depicted in Figure 1, has two parts: the convolutional base and the classifier.

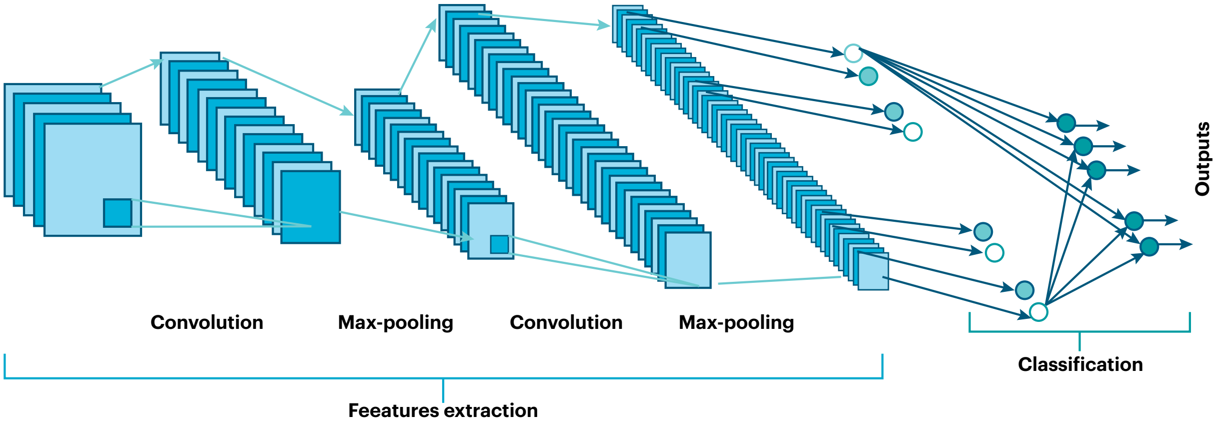
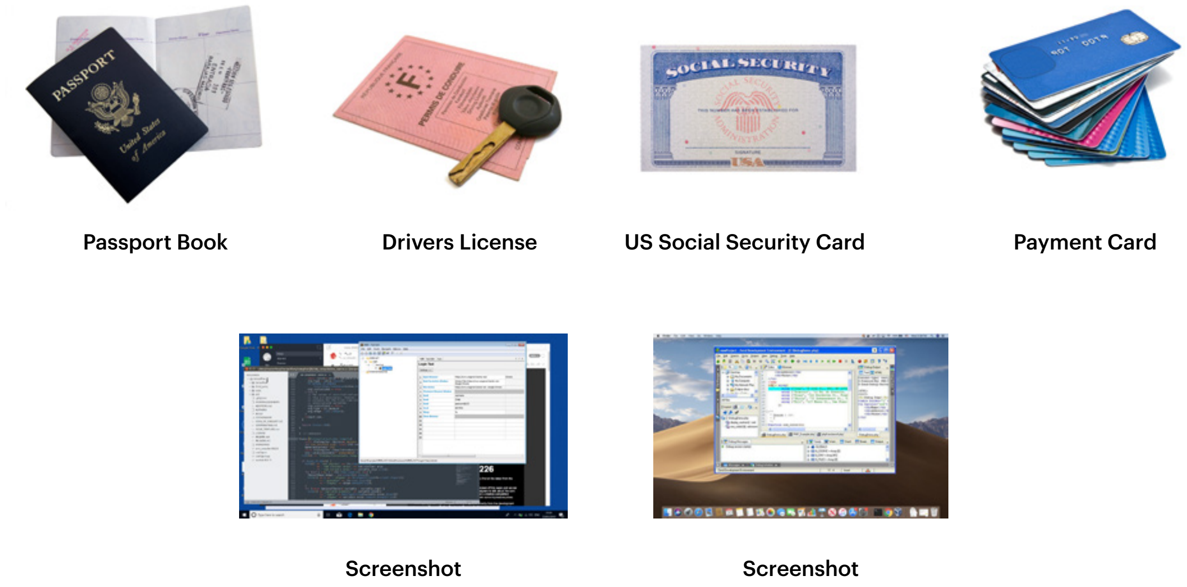


Figure 1: A diagram of a typical CNN

Think of how you would identify a passport, for example. Given an image, you look for unique things that describe a passport (i.e., features): a headshot picture, the word passport, name, a few dates, bar codes, and so on. That's what the convolutional base does with a stack of convolutional and pooling layers. The main goal of the convolutional base is to generate features from the image. It builds progressively higher-level features out of an input image. The early layers refer to general features, such as edges, lines, and dots in the image. Meanwhile, the latter layers refer to task-specific features, which are more human interpretable, such as the logo on a credit card, or application windows in a screenshot. The classifier is usually composed of fully connected layers. Think of the classifier as a machine that sorts the features identified in the convolutional base. The classifier will tell you if the features identified are a cat, dog, driver's license, or X-ray.

Netskope has developed CNN-based image classifiers for use with its NG SWG and Cloud Inline solutions. The classifiers can accurately identify images containing sensitive information, including passports, driver's licenses, US Social Security cards, credit and debit cards, fullscreen and application screenshots. These inline classifiers allow for granular policy controls to be applied in real time.


Passport Book

Drivers License

US Social Security Card

Payment Card

Screenshot

Screenshot

Figure 2: Examples of sensitive data that can be identified using image classification

Document classification

Similar to image classification, document classification is a key part of DLP that uses ML to improve accuracy and therefore reduce false positives. While the process to identify sensitive data is similar, there are some key differences to note.

Many of the documents that an organization's users transfer through or store in the cloud contain sensitive information, including confidential legal and financial documents, intellectual property, and employee or user personally identifiable information (PII). ML-based document classifiers automatically classify documents into different categories, including source code, tax forms, patents, and bank statements. ML classifiers complement the more traditional text matching or regex-based DLP rules. In many cases, manually configured regex rules can generate excessive false positives or false negatives when looking for specific patterns in documents. In comparison, ML classifiers automatically learn the patterns that identify sensitive data in real time, without the need for traditional DLP rules.

Text classification is one of the standard natural language processing (NLP) tasks. As illustrated in Figure 3, text content is extracted from documents and a pre-trained language model is used as an encoder to convert documents into numeric values that capture the contextual and semantic information of the documents' words. Based on these document encodings, document classifiers are trained using fully connected neural network layers.

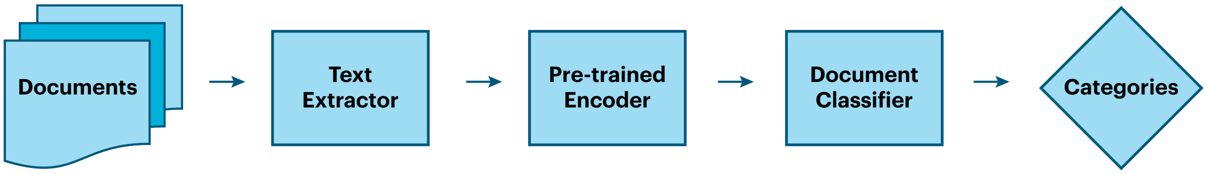
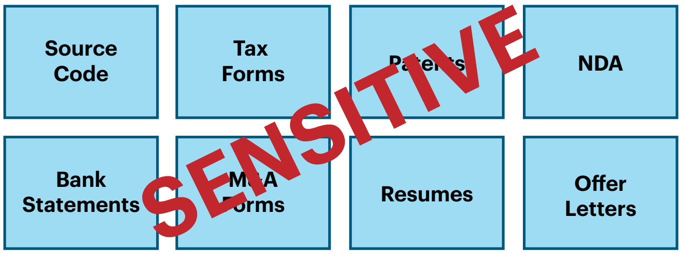


Figure 3: The NLP process of document classification

Currently, Netskope document classifiers can accurately identify more than 12 types of documents containing sensitive information, including:

- Source code
- IRS tax forms
- M&A forms
- Resumes
- U.S. patent files
- Offer letters
- Bank statements
- Nondisclosure agreements
- Consulting agreements
- Partner agreements
- Stock agreements
- Medical power of attorney forms



These lightweight, high-speed document classifiers are integrated into the Netskope DLP offering and can be used in conjunction with inline or API-based policies to provide data protection for Netskope customers.

TYOC: Train Your Own Classifiers

Because different industries and organizations have different types of sensitive data. As a result, their sensitive information classifiers need to reflect those differences. This could be a particular type of identity card, special HR documents, or critical infrastructure images, As an exploratory technology, Netskope has devised a new way for customers to train their own image or document classifiers. This method allows organizations to keep track of the information that matters most to them. Netskope achieves this functionality by providing customers with an on-premises Docker container that allows the training of ML classifiers while preserving customer privacy.

Within the Docker container, documents and images are first converted to numeric features, which act as an abstract representation of the input data. This process ensures that sensitive customer data is protected and kept in-house. For documents, the features are encodings of the text, while for images, the features represent the shapes, objects, and other qualities needed to understand the contents of the image. The random and non-linear transformations that take place in the feature extraction process make it impossible to retrieve the original input files or any specific sensitive data from the features. This irreversible training process has been designed to address any data security or privacy concerns that organizations may have with sharing their sensitive data with Netskope. The Netskope Security Cloud, where DLP policy enforcement takes place, only sees the learned features and never the original sensitive data.

After receiving the extracted features from a customer training process, additional training samples from Netskope's own corpus are added, and deep learning classifiers are trained with Netskope's ML engine. After this additional training, the customer can test the newly trained classifier with more samples in their on-premises container. After the classifier has achieved satisfactory accuracy based on customer testing, it will be deployed into the customer's tenant and used to detect sensitive information in documents or images within the customer's web and cloud traffic. While TYOC is currently available for customers to build out these custom classifiers and fine-tune their efficacy as a part of the exploratory program, the ability to deploy these classifiers will be on our roadmap.

See figure 4 for an illustration of how this TYOC process will work.

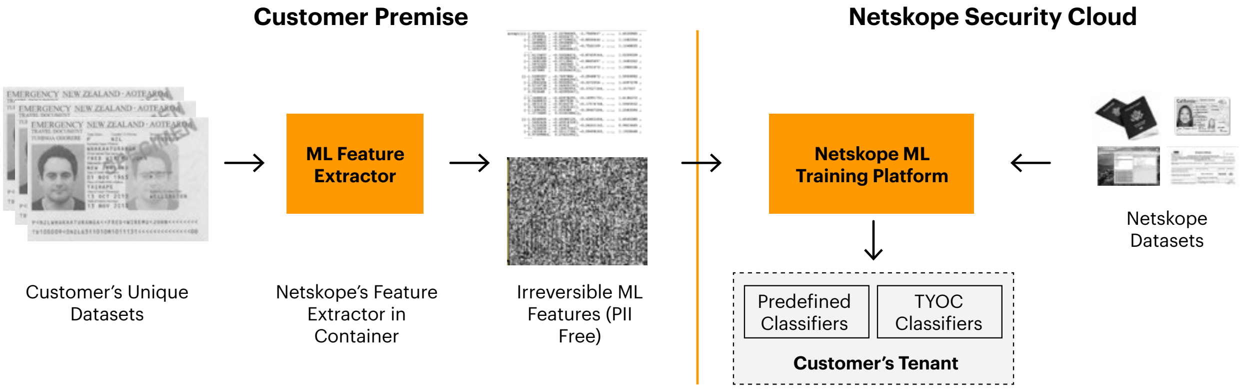

Customer Premise
Netskope Security Cloud
ML Feature Extractor
Netskope ML Training Platform
Netskope Datasets
Customer's Unique Datasets
Netskope's Feature Extractor in Container
Irreversible ML Features (PII Free)
Predefined Classifiers
TYOC Classifiers
Customer's Tenant

Figure 4: Overview of TYOC process

Not only does this system of "training your own classifiers" allow organizations to get the most from their Netskope DLP capability, but it also avoids any risk of exposing sensitive customer information to Netskope at any stage of the process.

Document Classification - Case Study

A real estate company uses Microsoft Office 365 OneDrive for Business as corporate storage for all of its documents, including sensitive tax and HR forms. It needs to ensure that these documents are not leaked and shared publicly.

A DLP policy was configured within Netskope to detect the download or share activities of Schedule K-1 (Form 1065) tax forms, which report on a partner's share of the income, deductions, credits, and more of its business's information. Initially, the DLP policy used document fingerprints to identify K-1 forms for the past 10 years. However, only 70 percent of the K-1 forms were successfully detected by this policy due to different variations of the form. The DLP policy was updated to use the tax form document classifier combined with keywords such as "Schedule K-1" and "Form 1065". As a result of the ML classifier being used, 100 percent of the K-1 forms were identified with zero false positives, effectively preventing sensitive tax forms from being leaked.

SUMMARY

At Netskope, we are embracing AI/ML technology, wherever it is applicable, to build better products, improve efficiency, and use the power of technology. Our advances in integrating ML-based image and document classification with our DLP is a testament to our approach. ML classification is more cost efficient, accurate, and secure than the traditional capabilities you might find in other DLP solutions. We also offer organizations the ability to securely train their own classifiers.

Netskope's full library of ML-based image and document classifiers are available as part of the Netskope Advanced DLP license, which is also included with Enterprise versions of Netskope products such as Next Gen Secure Web Gateway (NG SWG) Cloud Inline and CASB API. Netskope's Standard DLP license includes ML-based document classification for resumes and source code. We are continuously expanding our portfolio of inline file classifiers to meet our customers' needs. We welcome the opportunity to demonstrate our capabilities to you whether you are an existing Netskope customer or interested in learning more about our products. Please contact us for more information.

If your organization would like to be a part of our "train your own classifier" exploratory program, please contact contact-datascience@netskope.com.

Finally, you can learn more about Netskope Data Protection at:

https://www.netskope.com/products/capabilities/data-protection.