

GateScanner DLP

AI-Powered Data Loss Prevention for File Security

Protect sensitive information wherever files move.



GateScanner DLP helps organizations prevent sensitive information from leaving their boundaries through files. It combines **AI-powered content analysis**, traditional detection techniques and **document fingerprinting** to identify sensitive information across file formats, representations and contexts.

Unlike approaches that rely mainly on predefined patterns, GateScanner DLP lets you **define your own sensitive-information policies in natural language**. Security teams can detect not only recognizable patterns such as credit card numbers, but also information whose sensitivity depends on its meaning and business context – for example, a document stating the minimum discount the sales team may offer.

Available as a **standalone decision engine** or integrated with Sasa Software's GateScanner products, it adds an outbound data-protection layer to Sasa Software's file security portfolio.

Key benefits

Context-aware detection

Identify sensitive information by meaning and context, not just keywords or fixed patterns.

Flexible policy definition

Describe what counts as sensitive in plain natural language.

Multiple detection techniques

Combine AI analysis, regular expressions and document fingerprinting.

Deployment flexibility

The software runs on-premises; the AI is separate, so reuse your existing AI infrastructure, on-prem or in the cloud.

Integration flexibility

Use as a standalone API, with Sasa Software products, or with other file-handling systems.

Testing & feedback loop

Test policies on sample files and refine them until production-ready, then monitor activity.

How it works

1

Submit a file

A connected application or workflow sends a file, and names the policy to apply.

2

Extract content

Supported content is extracted; OCR exposes text embedded in images.

3

Evaluate

AI analysis, Regex and fingerprinting evaluate the policy.

4

Validate

Optional deterministic rules refine the findings.

5

Return a decision

Pass, Detection or Error is returned to the integrating system.

Three possible outcomes

PASS

No policy violation was detected.

DETECTION

A policy violation was detected.

ERROR

The evaluation could not be completed.

Errors block by default: an error is never treated as permission to release a file. The integrating system should block or hold the file until it can be evaluated successfully. The exact response to each outcome depends on the integration and its enforcement configuration.

Three complementary detection methods

AI-powered semantic detection

Interprets content by meaning, context and your organization's policies. Particularly useful for complex or business-specific information that cannot easily be described with fixed patterns. It can also recognize unusual encodings, such as card numbers written as words (depending on model and policy).

Regular expression (Regex)

Deterministic, efficient detection of structured information with recognizable formats, such as IP addresses and credit card numbers. Validation rules help separate valid findings from false positives.

Document fingerprinting

Register sensitive reference documents without retaining the original in the fingerprint gallery. Extracted content is compared against fingerprints using a configurable similarity threshold, catching modified documents and format conversions (e.g. Word to PDF).

Content extraction & OCR

Sensitive information also appears in images and screenshots. GateScanner DLP extracts text from a wide range of file types and applies OCR to text embedded in images, so detection and fingerprinting work on the underlying content.

- Detects structured data, such as credit card numbers, in images when successfully recognized
- The AI can also "see" images by itself, without OCR, when the selected model supports it

Natural-language policy management

A web-based administrative interface lets administrators define what to detect and add guidance explaining its meaning – for example, an *internal pricing strategy* policy that includes the minimum authorized discount for a customer.

- Create policies from natural-language descriptions
- Refine instructions after testing reveals missed or unwanted detections
- Use simple predefined patterns where no contextual interpretation is needed

Validation, testing & auditability

Deterministic validation

- Act only when a finding passes a precise check (e.g. a valid card number)
- Check values against predefined lists (e.g. specified IP addresses)
- Reduce unwanted blocking from overly broad patterns or AI instructions

Policy testing

- Upload a file and review how each detector responds
- Refine the policy and test again
- Validate changes without relying on production traffic

Administration & audit

- Logs of who changed policies, and policy change history
- Scan history
- Analytics on which policies block the most files, to spot overbroad rules

Deployment & integration

Deployment

- Decision engine hosted on a Windows server, on-premises or in a cloud environment
- Web-based administrative interface, accessed via browser
- API access for external systems
- The AI model is separate from the engine, so it can run on-premises or in the cloud, and existing AI infrastructure can be reused
- Detection engine is separate from the systems that enforce its decision

Integration

- **Standalone API:** an application submits a file and a policy name, and receives Pass, Detection or Error
- **With Sasa Software products:** GateScanner Mail Protection, Kiosk, Multi-source and CDR Integration Server
- Customers can build their own connectors; Sasa Software can develop connectors for commonly used systems

AI architecture & data privacy

On-premises AI

Connect to an AI model running in your own infrastructure, keeping content analysis in your controlled environment. You provide the GPU infrastructure and model; detection quality depends on the model and available compute.

External AI services

Where permitted, connect to supported services such as OpenAI and Google Gemini through APIs. Data handling and retention depend on the provider, service and your contract – verify these before sending sensitive content.

Complementing GateScanner CDR: GateScanner's established capabilities address threats in *incoming* files. GateScanner DLP adds the complementary question – does a file *leaving* your organization contain information that should not be disclosed?