OPSWAT

DATASHEET

AI Content Inspector

AI Content Fraud Detection at File Ingest

OPSWAT AI Content Inspector is an AI-driven content authenticity and document fraud detection engine for MetaDefender. It analyzes submitted images, PDFs, and text-bearing files to identify indicators of AI-generated content, manipulated media, and suspicious documents before they reach claims, payment, onboarding, or review workflows.

Files can pass malware checks and still contain fabricated or misleading content. AI Content Inspector adds a content-authenticity inspection layer to existing MetaDefender workflows, helping organizations evaluate not only whether a file is safe, but whether the content inside the file shows signs of AI generation or fraud.

How It Works

AI Content Inspector analyzes supported files using multiple detection methods across visual, textual, metadata, and document-structure signals. Results are returned as AI-generation and fraud-indicator verdicts that can support allow, flag, block, or route-for-review decisions.

File Ingestion & Normalization

Submitted files are validated, normalized, deduplicated, and prepared for inspection, including metadata and EXIF handling for supported image workflows.

Multi-Signal Content Analysis

The engine evaluates image artifacts, text patterns, metadata, and document structure to identify signals commonly associated with AI-generated, manipulated, or suspicious content.

Detector-Level Results

Detection results include category-level findings such as AI-generated image detection, AI-generated text detection, and fraud detection, with supporting detector outputs and confidence details.

Policy-Ready Verdicts

Verdicts can be used to support automated routing decisions, manual review queues, or workflow-specific actions before content reaches downstream approval points.

Key Features

AI-Generated Image Detection
Identifies indicators of synthetic or manipulated images used in property claims, accident reports, expense documentation, and other visual evidence workflows.

AI-Generated Text Detection
Analyzes text-bearing files for signals associated with AI-generated writing, including linguistic and structural patterns.

Document Fraud Screening
Inspects invoices, receipts, PDFs, and other supported documents for AI-generated content and fraud indicators before they reach accounts payable, expense, claims, or approval workflows.

PDF Support
Supports PDF inspection workflows that include both AI text detection and AI image detection.

Invoice Fraud Detection
Helps identify suspicious invoices, altered financial documents, fabricated receipts, and AI-generated supporting documents.

Insurance Claim Image Detection
Helps detect AI-generated or manipulated property damage images, accident photos, and other claim evidence submitted through digital intake workflows.

Detector Breakdown and Confidence Details
Provides deeper inspection results from individual detectors, including category-specific analysis and confidence scoring.

Uncertain Hits Threshold
Provides configurable control over uncertain detection results to help teams tune review sensitivity for their workflows.

Google SynthID Watermark Detection
Detects SynthID-marked image content as part of supported image authenticity workflows.

MetaDefender Platform Integration
Runs as a separate engine within MetaDefender Core and MetaDefender Cloud, alongside existing technologies such as Metascan Multiscanning, Deep CDR, and Proactive DLP.

Supported Environments and File Types

Deployment	Platforms	Supported File Types
• Cloud • On-premises • MetaDefender Core • MetaDefender Cloud	• Windows • Linux	AI Content Inspector supports common image and text-bearing formats used in fraud workflows, including: .pcx, .fpx, .jpg, .bmp, .emf, .avif, .apng, .png, .tga, .webp, .ico, .jpx, .j2k, .eps, .cur, .psd, .wmf, .txt, .md, .markdown, .pdf

OPSWAT.
Protecting the World's Critical Infrastructure

opswat.com/get-started