

3 THINGS YOUR FIREWALL ALREADY BLOCKS IN GENAI TRAFFIC

AND ONE IT CAN'T (YET)

The PAN-OS controls behind it, what each one needs, and where they run: on-premises, virtual, and cloud.

PREPARED BY ON2IT AUDIENCE EXECUTIVE AND SECURITY LEADERSHIP DATE JUNE 2026

STATUS CURRENT AS OF PAN-OS 11.X / 12.X

If you already run a Palo Alto Networks firewall, you already hold an AI control point. The same PAN-OS engine that classifies applications and inspects threats now sees GenAI traffic as a first-class citizen. Some controls are free with your content subscription. Others are licensed add-ons. The decision in front of you is not whether to deploy something new. It is how far up the stack you want to turn capability on.

TWO JOBS, OFTEN CONFUSED

"Securing AI" splits into two very different problems. They use different controls and different SKUs. Naming which one you mean is the first move in any AI security conversation.

REGULATE AI USE

GOVERNING WHAT YOUR PEOPLE DO WITH GENAI

Discovering shadow AI, classifying apps as sanctioned, tolerated, or unsanctioned, and stopping sensitive data from leaving in a prompt or an upload. This is the consumption side: ChatGPT, Claude, Copilot, and the hundreds of tools your staff already reach for.

PROTECT AI YOU RUN

SECURING THE AI APPLICATIONS YOU OPERATE

Defending the LLM apps, models, and agents you build and deploy against prompt injection, sensitive data leakage, insecure output, and model denial of service. This is the production side: your own AI workloads in the data center and the cloud.

THE CAPABILITY STACK

From coarsest to most specialized. Each layer sits on the firewall you already own and earns its place by what it does with AI traffic.

CAPABILITY	WHAT IT NEEDS	WHAT IT DOES WITH AI TRAFFIC	JOB	FORM FACTORS
App-ID and GenAI tagging Application visibility and policy	CONTENT SUB	Identifies named GenAI apps and their child functions through dynamic content updates. Lets you allow, block, and tag apps as sanctioned, tolerated, or unsanctioned, and see who is using what.	REGULATE	All NGFW form factors and Prisma Access
App-ID Cloud Engine (ACE) Cloud-delivered App-IDs	SAAS SEC INLINE	Adds on-demand App-IDs for new and shadow GenAI tools that would otherwise be seen only as ssl or web-browsing, closing the gap between the monthly content cadence and the pace of new app launches.	REGULATE	NGFW and Prisma Access
Enterprise DLP Content-aware inspection	SUBSCRIPTION	Inspects prompts, files, and non-file traffic for PII, source code, and intellectual property using ML classifiers. The detection engine that makes "do not leak this into a prompt" enforceable.	REGULATE	NGFW and Prisma Access
AI Access Security GenAI governance	SUBSCRIPTION	A 500+ app dictionary with risk scores, fine-grained control over functions, uploads, plugins, custom GPTs and browser extensions, DLP-powered prompt inspection, response threat scanning, user coaching, and posture recommendations against your existing rulebase.	REGULATE	NGFW and Prisma Access, managed in Strata Cloud Manager
Advanced Threat Prevention, URL, DNS Inline threat inspection	SUBSCRIPTIONS	Applies the same inline, ML-powered prevention to GenAI flows as to the rest of the network: command and control, exploits, malware, malicious URLs in GenAI responses, and DNS-layer abuse.	PROTECT	All NGFW form factors and Prisma Access
Prisma AIRS AI runtime security platform	ADJACENT PLATFORM	Protects the AI you operate. Network Intercept inspects traffic between your apps and LLMs inline. API Intercept embeds scanning in code. Catches prompt injection, data leakage, insecure output, and model DoS, plus model scanning, red teaming, and agent controls.	PROTECT	Cloud network architectures, containers, and in-application

REGULATING AI USE

Four layers, each broader than the last. You can start free and licensed-up only as the risk and the adoption warrant it.

LAYER 1

CONTENT SUB

APP-ID AND TAGGING

Your content subscription already delivers App-IDs for the major GenAI tools and their sub-functions. Write a positive-enforcement policy: allow the tools you sanction, block the ones you do not, and tag the rest.

BENEFIT

Immediate shadow-AI visibility and coarse allow or block on named GenAI apps, with no new license to buy.

LAYER 2

SAAS SEC INLINE

APP-ID CLOUD ENGINE

New GenAI tools appear faster than monthly content can track. ACE serves cloud App-IDs on demand for apps that would otherwise hide inside encrypted or generic web traffic.

BENEFIT

Far broader and faster app coverage, so a tool launched last week does not sit invisible in your logs.

LAYER 3

SUBSCRIPTION

ENTERPRISE DLP

Allowing a GenAI app is not the same as trusting everything sent to it. Enterprise DLP reads the actual content of prompts, files, and uploads for sensitive data before it leaves.

BENEFIT

The difference between "we permit ChatGPT" and "we permit ChatGPT but our source code never reaches it."

LAYER 4

SUBSCRIPTION

AI ACCESS SECURITY

The full governance layer. Risk-scored discovery of hundreds of GenAI apps, control down to functions and plugins, response scanning, user coaching, and recommendations that tighten your own rulebase.

BENEFIT

Safe, auditable GenAI adoption: enable the business while keeping data and posture under control.

PROTECTING THE AI YOU RUN

Threat-prevention subscriptions already harden the GenAI channel. For the AI workloads you operate yourself, Prisma AIRS extends the same firewall thinking into the model layer. Few organizations have turned this on for their own AI applications yet, which is exactly why the conversation is worth starting now.

ON THE DATA PLANE

SUBSCRIPTIONS

THREAT PREVENTION ON AI FLOWS

Advanced Threat Prevention, Advanced URL Filtering, and Advanced DNS Security do not govern AI, but they inspect the traffic to and from it. A malicious link returned in a GenAI response is still a malicious link, and your firewall already blocks it.

BENEFIT

The GenAI channel inherits the same inline prevention as every other flow, with no separate tooling.

IN THE AI STACK

PLATFORM

PRISMA AIRS

Network Intercept inspects traffic between your applications and their LLMs inline, including east-west workloads. API Intercept puts the same scanning inside your code. Both target AI-native threats: prompt injection, data leakage, insecure output, model DoS.

BENEFIT

Your own AI apps, models, and agents get a deterministic security control, not just good intentions in the prompt.

ONE POLICY, EVERY FORM FACTOR

The controls above are not tied to a box. PAN-OS is the same engine wherever it runs, so the AI policy you write follows the workload, managed centrally through Panorama or Strata Cloud Manager.

HARDWARE NGFW

PA-Series on-premises, data center and campus edge

VM-SERIES

Virtual NGFW in private and public cloud

CN-SERIES

Containerized NGFW for Kubernetes

CLOUD NGFW

Managed service in AWS, Azure, and GCP

PRISMA ACCESS

Cloud-delivered SASE for users and branches

WHAT TO TAKE FROM THIS

- 1** You already own an **AI enforcement point**. Your existing firewall sees GenAI traffic today.
- 2** Separate the two jobs. **Regulating** employee GenAI use and **protecting** the AI you run use different controls.
- 3** The first layer of AI control is **free with content**. Visibility and allow or block need no new SKU.
- 4** You buy up the stack by **risk and adoption**, from ACE to Enterprise DLP to AI Access Security.
- 5** Policy is **form-factor independent**. The same rule protects on-premises, virtual, and cloud workloads.

YOU CAN'T STOP THE BREACH. YOU CAN STOP THE SPREAD.

AI applications and AI workloads are not an exception to Zero Trust. They are simply new Protect Surfaces. The discipline is the same one ON2IT has run since 2005: define what matters, map the flows, write least-privilege policy, enforce it at the firewall, and watch it continuously. Every capability in this paper is an enforcement or inspection point. None of it governs itself.

That is the work ON2IT operates for you. The AUXO™ Zero Trust Platform turns these firewall controls into live, measurable posture, and our GSOC runs them around the clock so that an unsanctioned tool, a leaked prompt, or a poisoned model response is caught and contained, not discovered later. Turning the capability on is step one. Operating it as a managed Zero Trust service is where the breach stops spreading.

SEE WHAT YOUR FIREWALL ALREADY DOES FOR AI

Book a 30-minute working session and we'll map it against your own environment: what you already have switched on, what's one license away, and where AI traffic still moves unwatched. You leave with a prioritized next step, not a sales pitch.

BOOK A WORKING SESSION

or email us at marketing@on2it.net