AWS WAF AI Security: finding and protecting LLM, MCP and A2A endpoints

27 November 2026 · 4 min read · blogs.iictp.org/aws-waf-ai-security-mcp-a2a-llm-endpoint-protection

Short answer: AWS WAF AI Security is a managed rule group, made generally available by Amazon Web Services on 23 November 2026, that protects language model, Model Context Protocol (MCP) and Agent2Agent (A2A) endpoints. It arrives with automated discovery of those endpoints through the new API Gateway API Portal. It matters because an undocumented public MCP server is an exposed interface that many teams do not yet inventory.

TL;DR The short version

- AWS WAF AI Security, generally available from 23 November 2026, is a managed rule group protecting LLM, MCP and A2A endpoints.
- It comes with automated discovery of those endpoints through the new API Gateway API Portal.
- An undocumented public MCP server is an exposed interface, so inventory AI endpoints first and test the rules in count mode before blocking.

Read online, listen to it, or change the text size: https://blogs.iictp.org/aws-waf-ai-security-mcp-a2a-llm-endpoint-protection

What is AWS WAF AI Security?

AWS WAF is a web application firewall that filters requests before they reach an application. The AI Security rule group adds checks for AI traffic. MCP is a protocol that lets a model call tools and data sources, and A2A lets one agent talk to another. AWS says discovery covers endpoints behind API Gateway, Application Load Balancers and CloudFront across an AWS Organization.

How it works

- **Discovery.** The API Portal recognises AI endpoints among the APIs it finds from traffic logs and service configuration records. AWS says no agents need to be deployed and no logs forwarded.
- **Shadow APIs.** A public MCP server that nobody documented appears in the inventory as a shadow API.
- **Protocol checks.** The rule group recognises the protocols and blocks malformed MCP and A2A messages.
- **Content inspection.** When enabled, it calls the Amazon Bedrock Guardrails InvokeGuardrailsCheck API to detect prompt injection, jailbreak attempts and harmful content in requests, and sensitive information in responses.
- **Combination.** It can run alongside other managed rules and your own rules.

What is new compared with a generic web firewall

A general rule set sees HTTP paths and patterns. It does not know whether a message is a valid MCP exchange or whether a prompt is trying to override instructions. The new group adds protocol awareness at the edge and an optional model-based content check, and it ties both to an inventory so that unprotected AI endpoints can be listed and then covered.

Availability and pricing reported by AWS

- API Portal is listed as available in 32 AWS Regions.
- AI Security is listed for US East (N. Virginia and Ohio), US West (Oregon), Europe (London and Stockholm) and Asia Pacific (Tokyo and Sydney).
- Additional fees apply, and Amazon Bedrock charges separately for content inspection calls.
- In a separate notice the same day, AWS announced a new WAF pricing model from 1 February 2027, based on protected resources, inspected requests and protection level, with a premium request fee for traffic evaluated by advanced protections including AI Security.

Limitations and open questions

- The announcement gives no detection rates or false-positive rates for the content checks.
- Endpoints that are not behind the named AWS services are not described as discoverable.
- A firewall cannot decide whether an authenticated, well-formed tool call is something the user should be allowed to do. That stays with authorisation in the application.
- Sending content to an inspection API has cost, latency and data-handling effects you should measure.

Sources were read on 27 November 2026. Products, rules and data change, so check the linked originals before you rely on any detail.

How to adopt it

1. Run discovery and list every AI endpoint with an owner.
2. Remove or restrict public endpoints that have no business reason to be public.
3. Attach the rule group in count or monitoring mode where possible and review matches against real traffic.
4. Decide which routes need content inspection and estimate the extra Bedrock calls.
5. Keep authentication, per-tool authorisation and logging in place behind the firewall.
6. Recheck the bill model before 1 February 2027.

What it means for cloud and security engineers

Treat MCP and A2A servers as APIs: inventory them, authenticate them and rate limit them. A managed rule group is one layer. Pair it with the practices in what is Model Context Protocol and with least-privilege roles for whatever the tools can reach.

Frequently asked questions

- **What is AWS WAF AI Security?** A managed rule group that checks MCP and A2A messages and can inspect AI traffic for prompt injection and sensitive data.
- **What are shadow APIs here?** Endpoints seen in traffic or configuration that were not documented, including public MCP servers.
- **Does content inspection cost extra?** Yes. AWS says fees apply and Bedrock charges separately.
- **Is it available in every Region?** No. AWS lists a smaller set of Regions for AI Security than for the API Portal.

Related reading

- What is Model Context Protocol?
- What is prompt injection?
- What is least privilege?
- Cloud Architecture Practitioner certification

Related credential: Cloud Architecture Practitioner. Explore the exam facts and syllabus.

Sources and further reading

1. Amazon Web Services. AWS WAF announces automated discovery and protection for LLM, MCP, and A2A endpoints using Amazon Bedrock Guardrails. 23 November 2026. aws.amazon.com/about-aws/whats-new/2026/11/waf-ai-endpoint-discovery-protection
2. Amazon Web Services. AWS WAF introduces simplified pricing. 23 November 2026. aws.amazon.com/about-aws/whats-new/2026/11/aws-waf-introduces-simplified-pricing

Published by IICTP, an independent technical credential body, on 27 November 2026. The current version of this post is at https://blogs.iictp.org/aws-waf-ai-security-mcp-a2a-llm-endpoint-protection. More guides: blogs.iictp.org.