AI Log Agent

Deployment & Configuration Guide (AWS Marketplace)

This guide describes how to deploy and configure the AI Log Agent container image distributed via AWS Marketplace. The service exposes a REST API for log analysis with safe defaults (no outbound calls unless explicitly enabled).

1. What you deploy

AI Log Agent is a stateless FastAPI service packaged as a container image. By default it operates in offline mode (LLM disabled) and returns a structured response without contacting external services.

API endpoints:

1. **GET /health** – health probe; returns *{"status":"ok"}*.
2. **POST /v1/analyze** – accepts logs and returns a JSON analysis.

2. Prerequisites

1. A container runtime or orchestrator (Amazon ECS, Amazon EKS, or Kubernetes).
2. Permissions to pull images from the ECR repository associated with this product.
3. Networking configured according to least privilege (only the required inbound/outbound access).

3. Deployment

Deploy the image using your preferred platform (ECS, EKS, or Kubernetes). The container listens on a configurable port (default 8080). It is not publicly reachable unless you publish it via a Service / Load Balancer / port mapping.

Recommended deployment steps:

1. Pull the container image from the ECR repository linked to this AWS Marketplace listing.
2. Create an ECS task definition or Kubernetes Deployment/Pod spec.
3. Set CPU/memory limits based on expected log volume.
4. Expose the service only inside your VPC/cluster unless public access is required.

4. Configuration (Environment Variables)

Configuration is provided at runtime via environment variables. The image contains no embedded credentials or hard-coded secrets. If you enable external integrations, inject secrets using AWS Secrets Manager or Kubernetes Secrets.

Variable	Default	Allowed / Type	Purpose
PORT	8080	1–65535	Listening port inside the container

LLM_PROVIDER	none	none | openai | ollama | custom	Controls whether outbound analysis calls are ena
MASK_SECRETS	true	true/false	Redacts common secret patterns before analysis
MAX_CHARS	120000	integer	Safety cap for total input size processed

Safe default: With **LLM_PROVIDER=none**, the service performs no outbound network calls and returns an informational response indicating that LLM is disabled.

5. Security considerations

1. Do not bake credentials into images. Use Secrets Manager / Kubernetes Secrets for sensitive values.
2. Limit inbound access: expose the API only to trusted callers (private subnet, SG rules, or internal LB).
3. Limit outbound access: keep **LLM_PROVIDER=none** unless you intentionally enable an external provider.
4. Enable secret redaction by keeping **MASK_SECRETS=true** (default).

6. Example request

Send a JSON payload to **POST /v1/analyze** containing a log source label and a list of log lines. The service returns a JSON object with a severity/summary and metadata.

Example payload (structure): source, logs[], optional context{}.

7. Updates and maintenance

1. Update by deploying a newer image tag/version from the Marketplace-linked ECR repository.
2. Use rolling updates (ECS rolling deployment / Kubernetes rollout) to minimize downtime.
3. Monitor container health via your platform and use **GET /health** for readiness checks.

8. Support & responsibility

This product is provided as a self-managed container image. Customers are responsible for deployment, configuration, monitoring, scaling, and compliance with internal security and data-handling requirements. No SLA is provided.