生成式 AI 安全控制

Information Security for Workload Factory

NetApp
September 16, 2026

目录

生成式 AI 安全控制

了解 NetApp Workload Factory 中的 AI 控件

NetApp Workload Factory 包括生成式 AI 功能，可加速诊断，同时对模型访问、数据范围和运行时行为实施安全控制。

默认情况下，Workload Factory 启用 AI 诊断。您可以随时通过 Workload Factory **Administration** 设置禁用生成式 AI 功能。 "详细了解如何管理 GenAI 功能。"

安全范围的 AI 功能概述

Workload Factory 目前将生成式 AI 应用于以下功能：

- Ask Me 助手（RAG，包含精选 NetApp 文档，响应有来源支持）
- 延迟分析
- EMS 事件分析
- 错误日志分析

AWS Bedrock 概述（用于 AI 诊断）

Workload Factory 使用 Amazon Bedrock 进行 AI 推理。推理在您的 AWS 账户中运行，使用您配置的凭据和推理配置文件。

相关信息

- "Amazon Bedrock 选择加入 NetApp Workload Factory AI 诊断"
- "NetApp Workload Factory 的 AI 工具边界"
- "NetApp Workload Factory 的 AI 模型参数和计费"

Amazon Bedrock 选择加入 NetApp Workload Factory AI 诊断

NetApp Workload Factory AI 诊断使用 Amazon Bedrock 来分析事件，默认情况下该功能处于启用状态。在 Workload Factory 可以调用选定的 Bedrock 模型并收集诊断数据之前，您必须配置 IAM 权限、推理配置文件和具有 SSH 功能的已连接 Lambda 链接。如果这些组件中的任何一个缺失或 Bedrock 配置被删除，则 AI 诊断将不可用。

默认情况下，AI 诊断处于启用状态。

开始之前

- 确保您的帐户包含 AWS Bedrock 推理配置文件。
- 具有用于诊断数据收集的 SSH 功能的已连接 Lambda 链路。

步骤

1. 在您的 IAM 角色中授予 Bedrock 权限：
   - `bedrock:InvokeModel`
   - `bedrock:InvokeModelWithResponseStream`
   - `bedrock:ListInferenceProfiles`
   - `bedrock:GetInferenceProfile`
2. 在 Workload Factory 设置中配置推断配置文件 ARN。

如果缺少任何先决条件，系统将报告缺少的特定组件，AI 功能将保持非活动状态。

禁用 **AI** 诊断

要禁用 AI 诊断：

- 从您的 IAM 角色中删除 Bedrock 权限，或
- 在 Workload Factory 设置中删除 Bedrock 配置。

AI 诊断立即变为不可用，且不会发生任何残留数据处理。

区域和跨区域推理配置文件

Bedrock 推理在您的推理配置文件指定的 AWS 区域中执行。如果使用跨区域推理配置文件，AWS 可能会根据标准 AWS Bedrock 行为将请求路由到配置文件配置中的任何区域。数据不会离开 AWS 基础架构。

NetApp Workload Factory 的 AI 工具边界

NetApp Workload Factory 严格限制其 AI 功能与 AWS 和 ONTAP 环境的交互方式，因此您可以信任 AI 辅助诊断，而无需担心发生意外更改。Amazon Bedrock 使用只读的 AWS CLI 和 ONTAP CLI 诊断工具，这些工具会阻止变更命令并按大小和执行时间限制输出，所用的任何数据均为临时性数据，不会由 NetApp Workload Factory 或 AWS 保留。

工具安全控制

工具	功能	安全控制
AWS CLI（只读）	执行只读 AWS CLI 命令（ describe、list、get）	阻止所有变异动词（create、delete、modify 、update、put、remove）。每步最多 10 个命令。输出截断为 20 KB。
ONTAP CLI（只读）	使用 SSH 执行只读 ONTAP CLI 命令	阻止所有变异动词（DELETE、DESTROY 、CREATE、MODIFY、MOVE）。输出被截断。

代理无法修改、创建或删除资源。它只能观察和诊断。

数据保留和模型训练边界

根据 AWS 政策，Amazon Bedrock 不会存储或使用您的输入和输出来训练或改进基础模型。对于按需推理（由 Workload Factory 使用），AWS 会暂时处理您的数据，并在返回响应后不会保留数据。

NetApp Workload Factory 的 AI 模型参数和计费

NetApp Workload Factory 定义了与 AI 治理相关的模型配置、使用和计费边界。配置的模型使用确定性参数并返回结构化 JSON，以支持一致的分析结果。AWS 向您的账户收取 Amazon Bedrock 推理费用，而 Workload Factory 则跟踪 AI 使用情况并提供每个账户的每月成本可见性。这些部分描述了模型参数以及适用于 AI 消耗的限制。

型号和参数

参数	值
型号	Anthropic Claude（使用跨区域推理配置文件）
温度	0（确定性，无随机性）
最大输出令牌数	2,048
最大推理步骤	每次分析 15 次
输出格式	结构化 JSON（摘要、严重程度、根本原因、纠正措施）

账单和使用限制

- AWS 会将 Bedrock 推理费用计入您的账户（您的推理配置文件、您的凭据）。
- Workload Factory 在内部跟踪 AI 使用情况。
- Workload Factory 提供每个帐户的每月成本可见性。

  "在 NetApp Workload Factory 中跟踪存储成本"

Ask Me AI助手

适用于 NetApp Workload Factory 的 Ask Me 安全架构

Ask Me 是一款嵌入在 NetApp Workload Factory 中的生成式人工智能助手。它为 Amazon FSx for NetApp ONTAP 和 Workload Factory 主题提供实时对话支持。回复基于经过验证的 NetApp 文档，Ask Me 无法访问公共互联网或使用客户数据来训练模型。多个安全层有助于阻止恶意查询、限制对受支持域的响应，并防止用户输入覆盖系统指令。当 Ask Me 使用工具时，授权边界、只读查询强制执行、临时沙箱和审核日志记录有助于约束和监控执行。

您可以随时通过 Workload Factory 的 **Administration**（管理）设置退出 Ask Me，这将禁用针对您用户帐户的助手。"详细了解如何管理 GenAI 功能。"

Ask Me 中的 **RAG** 架构

Ask Me 使用检索增强生成（RAG）。

- 精心策划的知识库：回复基于经过验证的已发布 NetApp 文档的托管语料库。
- 检索评分和过滤：模型上下文中仅包含足够相关的内容。
- 无法访问互联网：模型无法获取、浏览或抓取外部内容。
- 源归因：回复包括对所用源文档的引用。

多层提示安全

- 第1层：安全分类器（LLM-as-a-judge）拦截恶意查询（提示注入、权限升级、数据泄露、社交工程、策略违规）。

  系统会记录被阻止的查询，生成模型永远不会看到这些查询。

- 第2层：系统提示强化可防止用户输入覆盖系统指令。
- 第 3 层：域范围限制将响应限定在 FSx ONTAP 和 Workload Factory 主题范围内。
- 第4层：工具使用治理在强化沙箱中验证参数并运行查询；模型无法执行任意操作。

模型数据隔离和隐私

- 不对客户数据进行模型训练
- 请求级隔离（无跨租户数据泄露）
- 具有有界历史记录的会话范围对话上下文
- 无持久模型内存
- 受管推理基础设施

工具使用和代理安全

- 授权范围的工具（强制执行用户权限边界）
- 严格的工具执行超时
- 用于检索数据的临时会话范围沙箱
- 使用允许列表实施只读查询
- 具有敏感参数清理功能的工具调用可审计性

适用于 **NetApp Workload Factory** 的 **Ask Me** 访问控制

在 NetApp Workload Factory 中，Ask Me 的访问安全核心在于运行时操作期间的身份验证会话、用户级问责制和受保护的密钥处理。身份验证控件会验证每个会话，并将交互与特定用户关联。敏感凭据在运行时从托管密钥保管库中检索，并从日志中清除。该服务还应用基础架构控制，包括非 root 容器执行和受限的 CORS 允许列表。

身份验证

- 带有加密验证的签名 JWT 身份验证（包括受众、颁发者和算法检查，如 RS256）
- 服务边界处中间件强制身份验证
- 用户身份范围界定，以便交互可归因于特定用户

凭据保护和密钥处理

- 安全保管库存储：服务使用的敏感凭据存储在托管密钥保管库中，并在运行时检索。应用程序代码、配置文件或容器映像中不存储任何机密。
- 日志清理：敏感值（凭据、令牌、密码、访问密钥、证书）在写入日志之前会从日志中隐去。

基础设施安全

- 非 root 容器执行
- CORS 限制为受信任的 NetApp 域的显式允许列表

适用于 **NetApp Workload Factory** 的 **Ask Me** 执行模式

NetApp Workload Factory 为 Ask Me 提供多种执行模式，每种执行模式具有不同的安全和隐私特性。当 Ask Me 使用工具集成时，例如查询客户自己的 Workload Factory 资源，工具执行通过分层保护进行控制。启用工具的执行在经过身份验证的用户权限内运行，并对每个操作的持续时间和范围施加限制。检索到的数据保留在阻止外部访问的临时会话沙箱中，并在会话结束时销毁。独立的提示和代码控制强制执行只读查询，而审核记录捕获工具活动并清除敏感值。

工具执行保障措施

- 授权范围的工具在经过身份验证的用户权限内运行。
- 工具执行超时限制长时间运行的操作。
- 临时数据沙箱在会话范围内存储工具检索的数据，在引擎级别阻止外部访问、文件 I/O、网络调用和扩展加载，并在会话结束时销毁。
- 只读查询强制执行通过严格的允许列表验证生成的查询。
- 提示执行和代码执行的安全性是独立应用的。
- 工具调用可审计性日志记录工具名称、参数、时间戳和结果状态，并对敏感值进行脱敏处理。

输出控制

- 令牌限制强制执行
- 分类/安全关键操作的确定性输出（温度 0）
- 提前终止和清理的流式传输

向我询问 **NetApp Workload Factory** 的隐私和可用性

在 NetApp Workload Factory 中，Ask Me 隐私控制可限制数据泄露、保持会话处理隔离，并维护用户交互的隐私边界。您的输入由托管 AI 服务处理，该服务不会使用这些输入来训练或改进基础模型。检索到的操作数据在会话期间保留在临时内存存储中，不会被持久化或共享。跨多个云区域的多模型回退链可在主模型受到限制或不可用时支持可用性，同时相同的安全控制也适用于回退模型。

数据处理和隐私

- 提示中的用户数据：由托管 AI 服务处理，不使用您的输入来训练或改进基础模型。
- 知识库内容：仅限精选、已发布的 NetApp 文档；您的数据不包括在内。
- 操作数据：仅在用户查询自己的资源时检索；在会话期间保存在临时内存存储中，不持久化或共享。
- 审核日志：记录查询元数据（匿名用户 ID、时间戳、持续时间），并清除敏感字段。
- 反馈数据：单独存储并链接到查询 ID，而非匿名用户 ID 以外的个人身份信息。

可用性和隔离控制

Ask Me 在多个云区域采用多模型后备链。

如果主模型受到限制或不可用，系统可以故障转移到其他模型。

所有模型均受相同的安全控制、系统提示加固和隔离保证。