Cloudflare AI Security Suite

管控数据和管理风险，确保 AI 全生命周期交互安全。

自信扩展 AI

AI 风险需要现代安全

您的团队正借助 AI 加速创新，但这也带来了重大安全风险。阻止 AI 或添加复杂单点解决方案的老套做法正在失效，因为这抑制创新且忽视了现实：

- **85% 的员工**使用未经 IT 审核的 IT 工具。[1]
- **93% 的员工承认**在未经批准的情况下将公司数据输入 AI。[2]
- **63% 发生数据泄露**的企业没有实施 AI 治理策略。[1]

了解如何赋能您的团队——充分利用 AI 带来的生产力优势，同时维持业务所需的安全性和控制。

保护智能体式和生成式 AI 的统一平台

Cloudflare 提供了一个统一平台，帮助企业自信拥抱 AI。我们赋能安全领导者管控风险，帮助技术团队释放生产力，并支持平台工程师安全构建，确保每个人都能安全地共同创新。

开始使用 Cloudflare AI Security Suite

Cloudflare AI Security Suite 通过统一平台交付，有效保护工作空间中的 AI 工具使用与面向公众的应用。发现影子 AI，保护模型免受滥用，保护智能体访问，并防止提示词中的数据泄露——让您的企业能够安全高效地创新，同时拥有更高的可见性和更强的控制力。

扩展可见性

覆盖 AI 应用、API 端点和 AI 智能体连接。

缓解风险

提供防护机制、态势控制和实时威胁缓解

保护数据

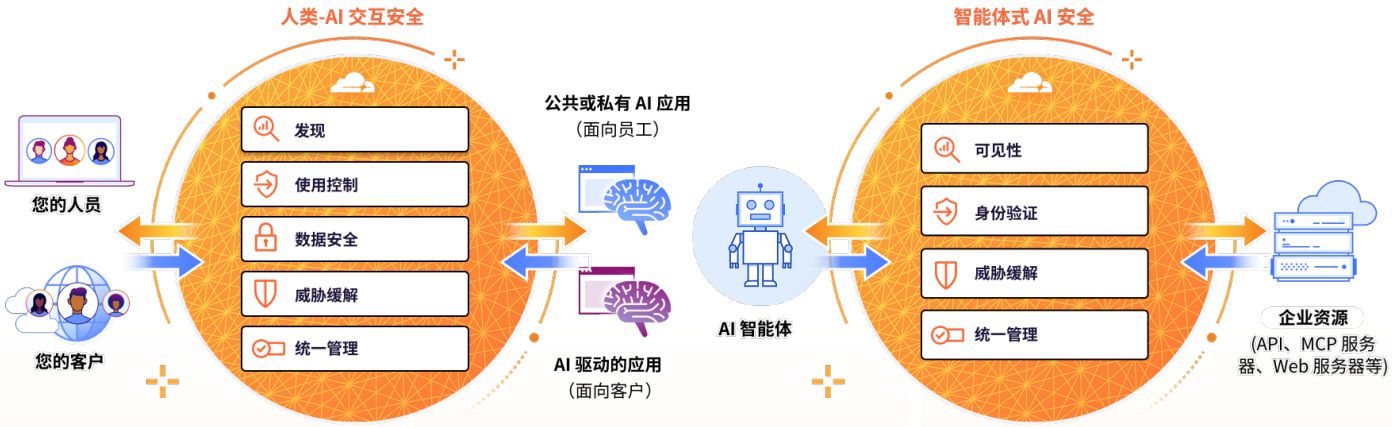
提供针对员工和 AI 智能体的提示词保护和访问控制。

享受内置安全防护
在 Cloudflare 上开发 AI。

 | 010 8524 1783 | enterprise@cloudflare.com | www.cloudflare.com/zh-cn/ REV:PMM-OCT2025

使用 Cloudflare AI Security Suite 保护生成式和智能体式 AI 通信

控制员工在生成式 AI 中使用的数据，管理自主智能体带来的安全风险，有效保护所有 AI 通信。

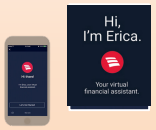


- **您的人员：**员工、承包商、开发人员等。
- **您的客户：**网站和应用的访问者

- **生成式 AI 应用**：独立或嵌入

Gemini Claude ChatGPT Copilot

- **AI 驱动应用**
示例：Bank of America 的 AI 聊天机器人 Erica



- 用于任何应用的 **API** 与 **MCP 服务器**

Google Workspace

- 用于推理和检索增强生成（RAG）的**Web 服务器**

通过贯穿 AI 生命周期的安全设计加速 AI 应用落地

确保员工安全使用 AI

实施 AI 使用控制和 AI 安全态势管理 (AI-SPM)，以降低风险并保护数据。

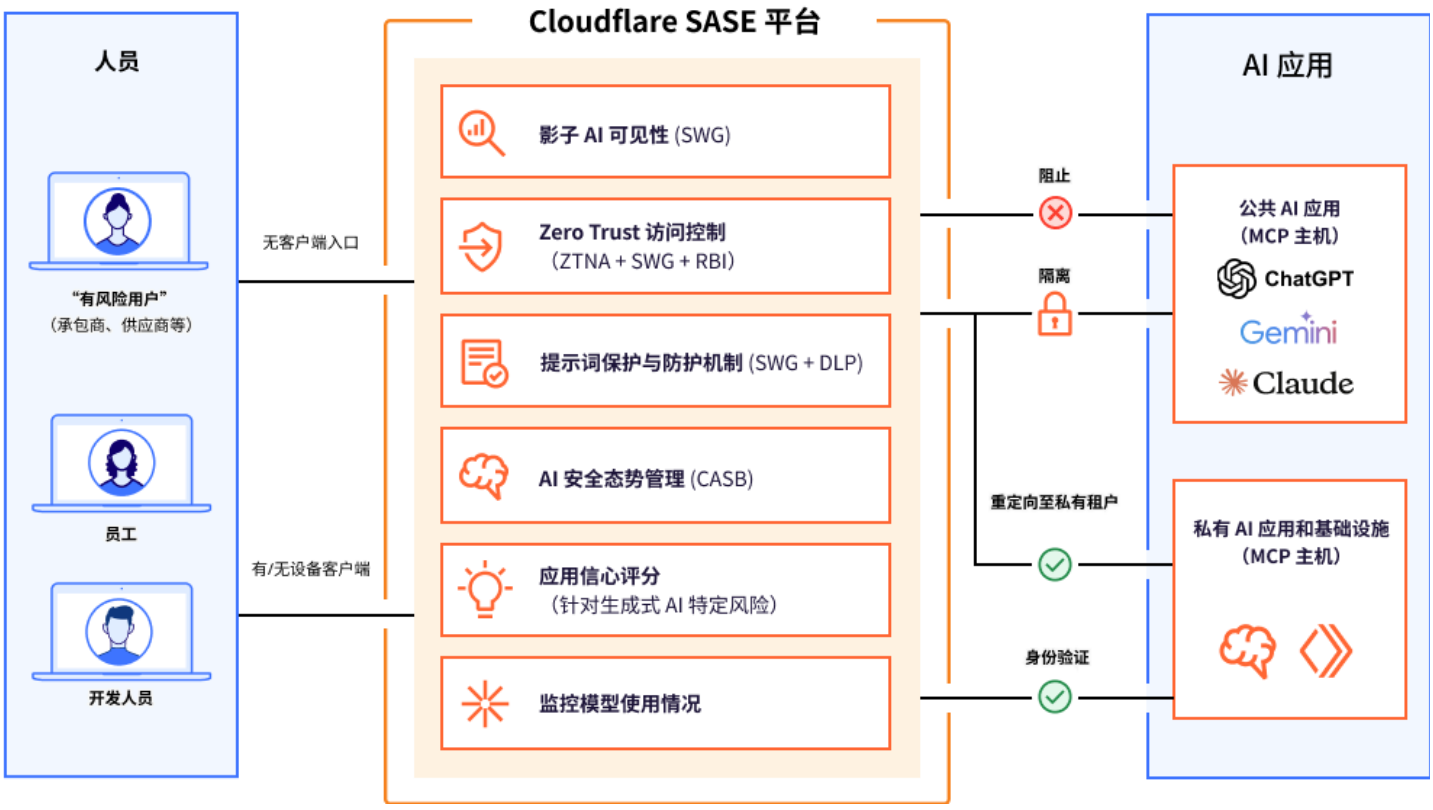
保护 AI 驱动的应用

实时保护您的 AI 应用和 API 免受提示词注入和数据泄露的侵害。

安全构建 AI

通过集成可观测性、速率限制和内联 AI 防护，助力开发人员保障 AI 应用的安全。

使用 Cloudflare 的 SASE 平台安全地保障员工使用 AI 应用和工作负载安全



SSE 保护人机通信

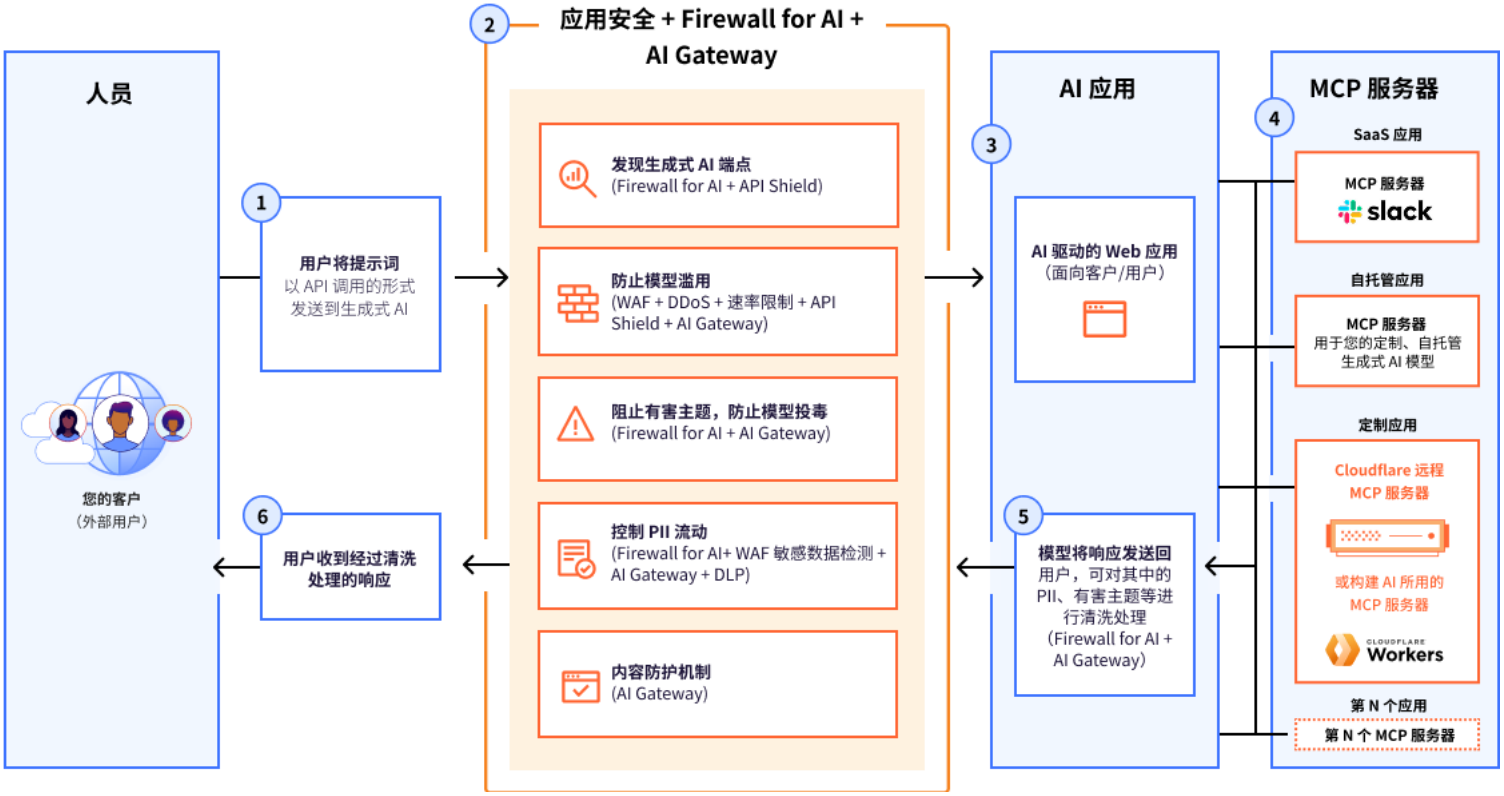
- **可见性：**通过内联流量检查，发现并分析影子 AI 使用。通过透明度评分评估这些 AI 应用带来的风险。
- **访问控制：**阻止、隔离、重定向或允许用户连接。按应用强制执行基于身份的 Zero Trust 规则。
- **提示词保护和防护措施：**根据意图（例如，越狱尝试、代码滥用、个人身份信息请求），检测并阻止用户提示词。
- **数据安全：**利用 AI 驱动的数据丢失防护（DLP）检测，阻止 PII、源代码等敏感数据泄露。
- **AI 安全态势管理：**通过 API 与 生成式 AI 工具集成，使用我们的云访问安全代理（CASB）扫描错误配置。现已支持 ChatGPT、Claude 和 Google Gemini。

用于保护 AI-资源之间通信的 MCP 服务器门户

- **可见性：**聚合所有模型上下文协议（MCP）请求日志，用于审计和分析。在添加到门户之前，审核并批准每个 MCP 服务器。
- **身份验证：**根据身份验证结果，允许用户访问门户。根据最小权限原则限制对 MCP 服务器的访问范围。
- **连接：**使用单个 URL 连接所有可访问的 MCP 服务器，无需单独配置每个 MCP 服务器。
- **统一管理：**对 AI 连接应用与人类用户相同的细粒度访问策略。

注意：MCP 服务器门户支持任何 MCP 服务器，包括（但不限于）在 Cloudflare 上构建或部署的远程 MCP 服务器。此功能作为 Zero Trust 网络访问（ZTNA）控制提供。

使用 Cloudflare 的模型中立内联安全功能保护 AI 应用和工作负载



通过应用安全和 Firewall for AI 保护面向公众的 AI：

- **发现生成式 AI 端点：**自动发现企业 Web 资产中的所有 AI 模型和 API。
- **保护 AI 模型免受滥用：** 使用我们专门打造的 Firewall for AI 来防范提示词注入、模型投毒、过度使用以及可能绕过传统安全防护手段的其他威胁。
- **控制 PII 流：**扫描用户提示词和模型响应，以防止敏感数据暴露，帮助维持合规性。
- **内容防护：**使用 Llama Guard 等集成模型，阻止不安全或有害的提示词。创建自定义 WAF 规则，轻松阻止或记录可疑 AI 交互。

通过 Developer Platform 和 AI Gateway 保护您构建的 AI

- **统一 AI 控制平面：**通过单一仪表板管理所有 AI 应用。路由请求、缓存响应、控制成本并监控性能。
- **在边缘保护凭据：**在边缘安全地存储 API 密钥和机密，防止客户端泄露，并简化跨供应商的密钥轮换。
- **实施内容安全防护措施：**自动识别并阻止/删除提示词和响应中的有害内容和 PII。

Cloudflare 是唯一能够同时保护企业公共和私有 AI 环境的供应商

实施恰当的防护机制，安全地采用 AI，确保安全能够加速而非阻碍创新。

- **统一的 AI 生态系统保护：**复杂的安全架构会增加风险。使用单一平台保护数据，确保 AI 全生命周期的合规性。
- **面向未来的全球架构：**通过一个后量子安全网络预防未来的挑战，其能够扩展以支持任何流量规模，持续适应新威胁，并可编程以满足新的使用场景。
- **AI 驱动的安全：**我们的 AI 驱动防御系统检查提示词和响应，以进行实时威胁检测。
- **经验证的 AI 领导力：**借助一个受到 80% 的 50 强生成式 AI 公司信赖的平台，自信地进行创新。
- **模型中立部署：**安全控制适用于组织环境中的所有 AI 模型，为 AI 部署治理提供统一方法。

客户感言

indeed

世界第一大求职网站
阅读案例研究

识别并控制影子 AI

同步推进 VPN 替代项目

APPLIED

保险科技
阅读案例研究

隔离 ChatGPT 等公共生成式 AI 工具

阻止复制粘贴敏感数据

AI 赋能的 SaaS 公司

保护 PII

防止客户向公共生成式 AI 端点提交敏感信息

Cyndx

AI 驱动的金融科技

推理成本降低 95%

通过采用 Cloudflare 来缓存和运行来自 AI 模型提供商的响应

准备好讨论您的 AI 安全需求了吗？

咨询专家

1. 2025 Manage Engine 研究：来源
2. 2025 IBM，《数据泄露成本报告》：来源

 | 010 8524 1783 | enterprise@cloudflare.com | www.cloudflare.com/zh-cn/