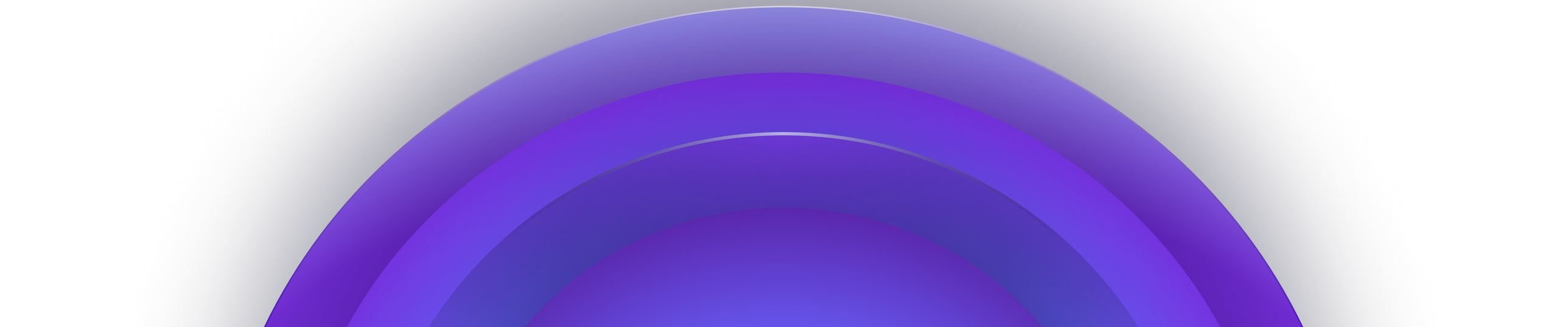
Agentsflare
下一代企业级 AI 统一网关解决方案

新一代私有化、自主可控、零信任合规的 AI 模型统一网关解决方案



我们的愿景：
成为您值得信赖的下一代主权级 AI 基础设施

产品名称

Agentsflare AI Gateway

命名寓意

“Agent”代表智能体，是AI应用的核心；“Flare”象征着光芒与网络边缘的加速能力，寓意为AI应用提供清晰的洞察与极致的性能。

产品口号

“Your Sovereign AI Infrastructure（您的主权AI基础设施）”

核心定位

“主权”－这一关键词体现了Agentsflare赋予客户的完全控制权、数据所有权和治理权。针对目标客户而言，“主权”意味着其AI应用和数据再也不用受制于外部第三方，一切尽在掌握。

目标用户

Agentsflare 专注服务于那些对安全性、合规性、可用性和全球性能有严格要求的大型组织。我们将以 “**企业主权级AI基础设施**” 的以及顶级咨询方案设计为企业量身定制专业的一站式大模型接入的解决方案。

为您的核心 AI 团队赋能

企业管理者

核心挑战：失控与风险

- **成本失控**：各部门自行采购，预算超支频发，集团总账成“黑箱”。
- **管理混乱**：大模型 API Key 散落各处，安全风险高，无法统一管控。
- **风险高悬**：密钥泄露风险高，缺乏审计日志，合规追责难。

AI 项目管理者

核心挑战：低效与掣肘

- **授权两难**：共享主密钥风险大，独立申请流程慢，团队创新受阻。
- **成本不透明**：无法实时查看本部门账单，成本难以分摊和管理。
- **协同低效**：缺少独立管理空间，内部资源调配与项目监控困难。

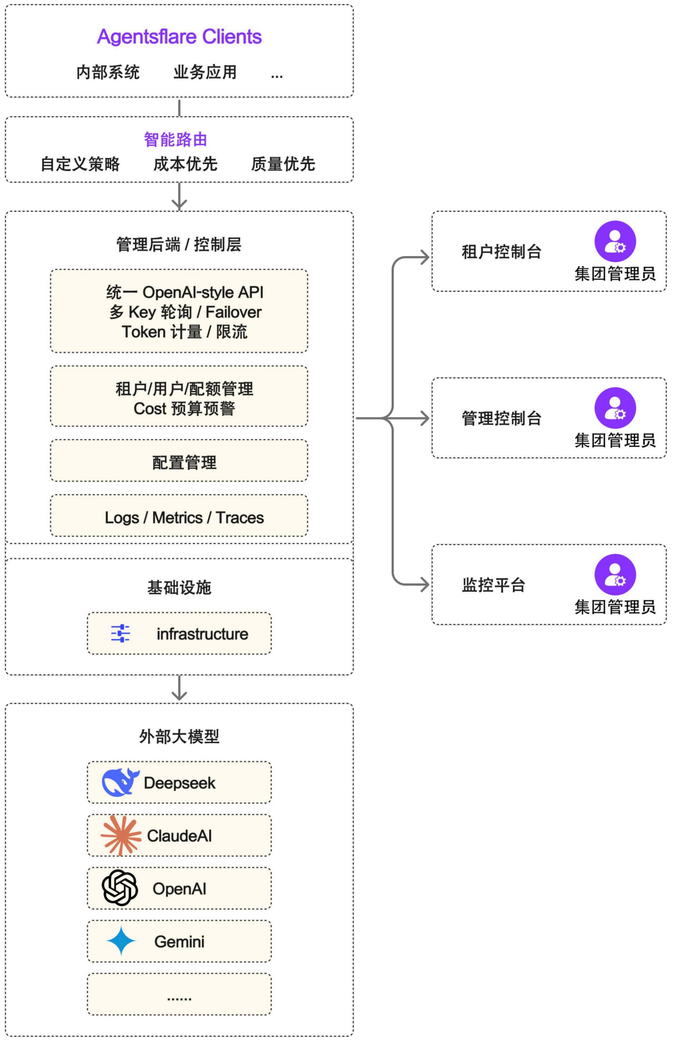
AI 开发者

核心挑战：繁琐与不稳

- **开发繁琐**：需适配多家模型API，接口不一，浪费大量开发时间。
- **服务中断**：高并发时轻易触发厂商速率限制，切换繁琐，导致应用频繁卡顿。
- **选型困难**：难以动态智能路由最具性价比的模型，导致不必要的成本浪费。

我们的差异化优势：从“工具”到“专业基础设施”

企业若选择自建 AI 网关，通常会面临一个由众多开源组件和云服务拼凑而成的复杂架构，不仅搭建和维护成本高昂，更难以形成统一的治理和安全标准；Agentsflare 深度洞察大型企业在 AI 时代的核心痛点，全方位为企业不同场景提供统一管控、性能、安全保障。



统一 API 接入

提供与 OpenAI完全兼容的 API，开发者零成本迁移一次集成即可调用所有模型。

多租户与统一计费

强大的组织与项目管理能力，支持分层级的 API Key 生成与权限控制。并实时进行账单，支持限额预警。

智能故障转移与负载均衡

当首选模型或节点不可用时，系统自动无感切换至备用选项，保障业务连续性。

金融级数据安全赋能

支持企业级数据安全沙箱集成，隐私计算能力集成，全球零信任安全网络集成。（需第三方集成）

灵活的部署模式

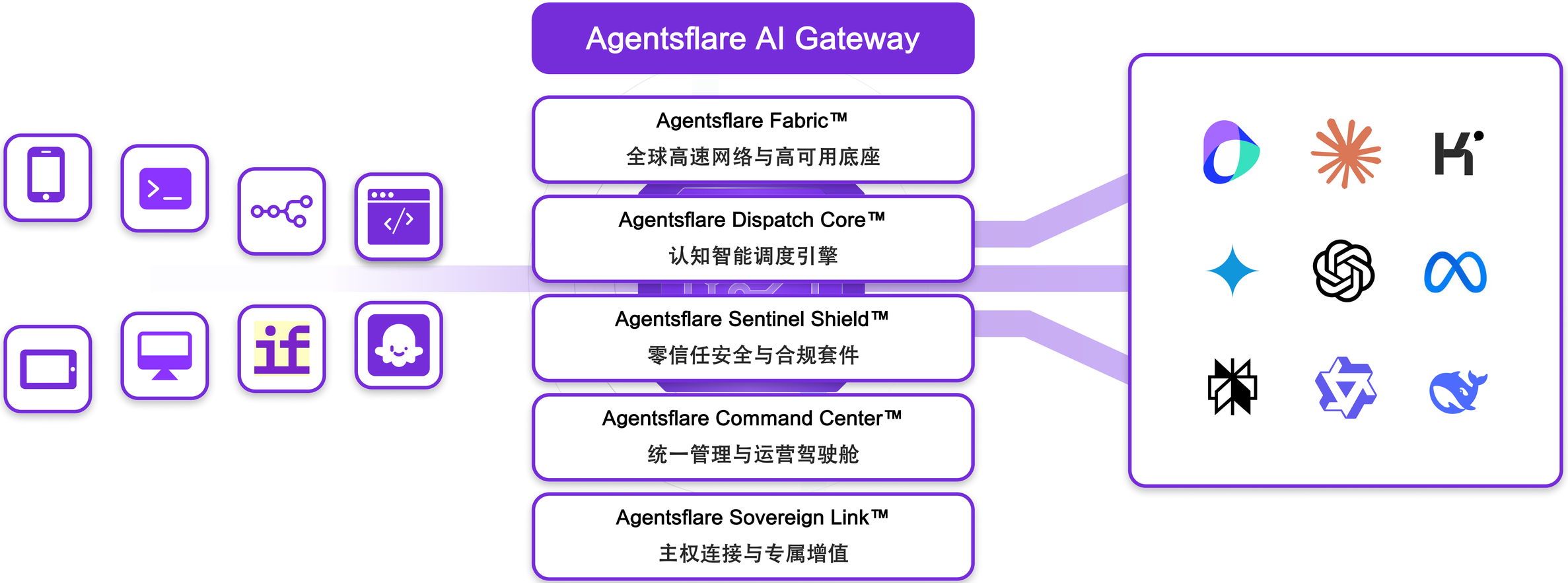
支持公有云 SaaS、多云私有化以及本地混合部署，满足最严苛的数据合规要求。

亚太区专属服务

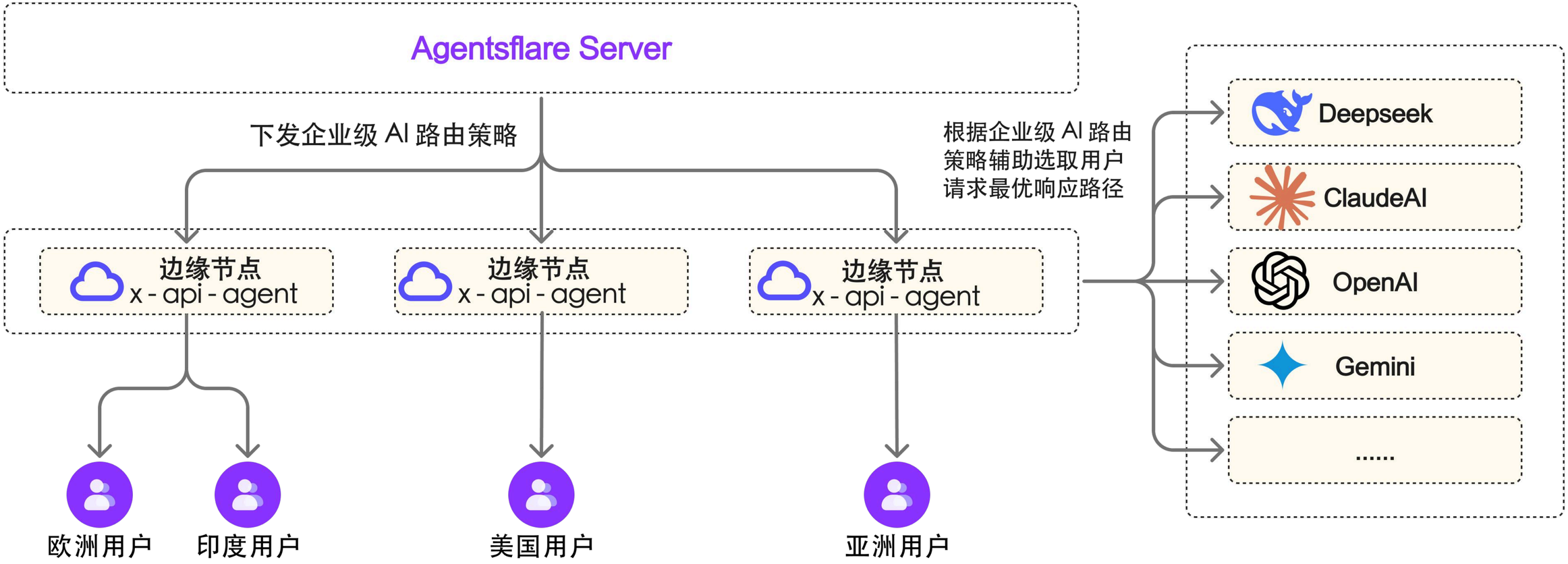
提供企业级现场部署、技术支持与定制开发服务，亚太区 7x24 现场服务保障。

核心架构 • 五大支柱

Agentsflare由五大核心模块组成，它们有机协作，共同构筑起平台的强大能力。



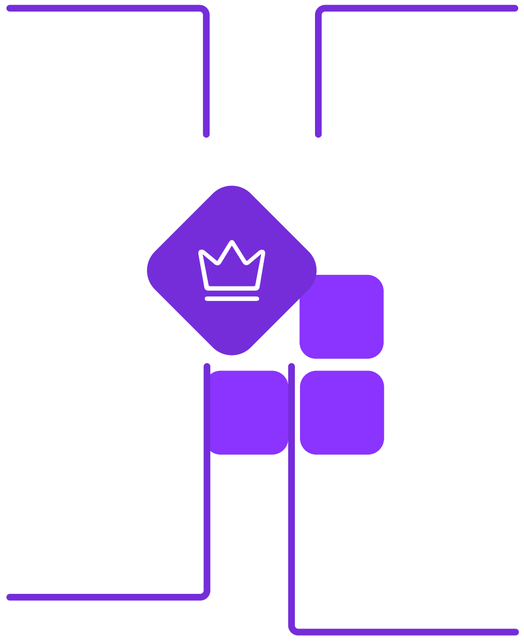
Agentsflare 技术架构



Agentsflare Sovereign Link™ 主权连接与专属增值

定位

Agentsflare Sovereign Link™ 代表我们为顶级客户提供的专属权益，是“绿色通道”理念的延伸。



专属模型访问权

通过与模型厂商的深度合作，为顶级客户争取到最新模型的抢先使用资格，甚至未公开发布版本的试用权，帮助客户保持技术领先。

预留算力容量

与供应商协议预留算力配额。在服务尖峰时段，客户的请求将优先处理，不排队、不降速，如同高速路的“专用车道”。

全能力专家服务

为每位顶级客户配备专属专家团队，提供贯穿全生命周期的全能力专家式服务，从战略研讨到持续运营优化，犹如客户的延伸部门。

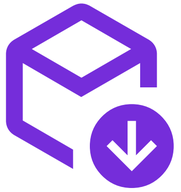
为您的核心 AI 团队赋能

Agentsflare 公有云

- 数据经互联网加密传输至 Agentsflare 云服务，再由平台转发至各模型厂商。
- 您无需关心服务器、扩容、版本更新等任何运维工作。
- 极速上线，成本最低 (无运维人力)，自动扩展

多云私有化部署

- 部署在您自己的云账号 (如 AWS/Azure/阿里云) 内。
- 核心网关和私有模型调用在您内网完成。仅在访问公有云模型时，数据从您的VPC统一、可控地流出。
- 希望利用自有云折扣 (BYOK) 的集团
- 复用现有云资源和安全体系



本地化部署

- 数据与应用100%不出企业内网，满足金融、军工级的最高合规要求。
- 所有数据流、模型调用均在企业内网闭环，与公网完全隔离。
- 最高级别的数据安全和隐私保障。完全的自主可控，不受外部影响
- 网络性能稳定，需自行维护私有化大模型算力基础设施及端点。

功能/方案详细对比

我们不追求短暂的速度优势，而是构建一个围绕 **信任、安全、性能** 的“重装甲”平台。我们的目标是让客户一旦选择，就再也无需考虑其他方案。

功能/方案	Agentflare	云服务商原生工具	传统API网关	传统API网关
AI特定威胁防御	高 (AI WAF集成)	中	低	极高复杂性/成本
零信任访问 (ZTNA)	高 (原生集成)	中	无	极高复杂性/成本
统一多云策略	高 (核心价值)	低 (锁定)	无	极高复杂性/成本
不可变审计日志	高 (类QLDB架构)	低	低	极高复杂性/成本
数据主权与区域管控	高 (全球网络)	中	无	极高复杂性/成本
低延迟边缘推理	高 (Workers AI)	有限	无	极高复杂性/成本
模型与云无关性	高	低	高	高复杂性
运营与集成成本	低	中	高	极高

案例：某全球金融服务巨头如何通过Agentsflare重获AI控制权

客户行业：财富500强金融服务公司

面临挑战：超过15个业务部门独立采购和使用不同的LLM服务，导致成本失控、安全审计困难、新应用集成周期长达数月。

解决方案：部署Agentsflare企业版网关，作为集团所有AI调用的唯一出口。

实施前

- API密钥散落在各部门代码库中
- 每月收到多张LLM账单，无法归因
- 关键应用因OpenAI宕机而中断
- 新模型集成需3周开发测试

实施后

- ✓ 统一密钥保险库，动态注入统一
- ✓ 成本中心，按部门/项目分账
- ✓ 自动故障切换至Azure备用模型
- ✓ 通过网关配置，1小时内上线