

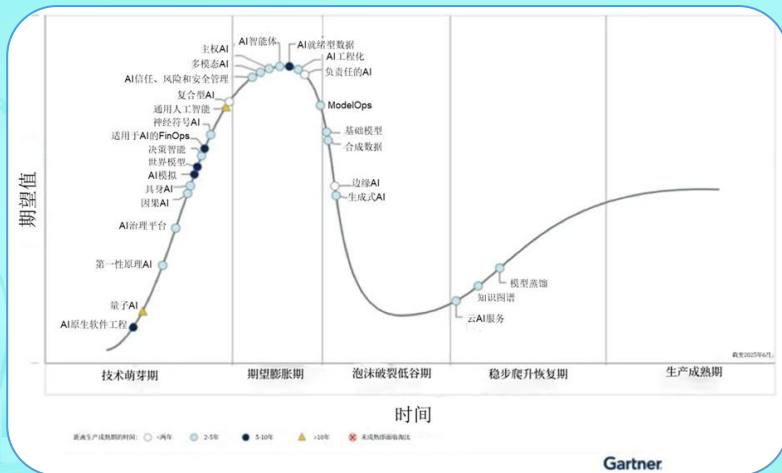
AI智能体全生命周期保护与Token治理

张铁峰 | 深信服科技股份有限公司 金融行业运营总监

AI各场景能力在快速发展，Agent转向生产实践

到2030年，大部分IT工作都将AI化

到2030年人类不借助AI独立完成的IT工作将为零，而在AI辅助下完成和AI独立完成的IT工作将分别占到75%和25%。这意味着企业需同时关注AI就绪度与人员就绪度，通过平衡二者才能从AI中获取价值。



上述所有技术成熟度曲线以及观点内容来源：Gartner



AI从个人级走到生产级



Agent数量爆炸式增长



自开发自建是AI应用明显的趋势

AI的快速发展给传统IT也带来新的挑战

Mythos的出现，高水平安全攻击快速平民化

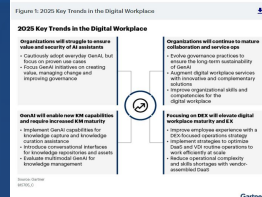
对比维度	Mythos	Opus 4.6	GPT-5	Gemini Ultra 4
零日漏洞挖掘能力	★★★★★	★☆☆☆☆	★☆☆☆☆	★☆☆☆☆
沙箱 / 安全护栏绕过	★★★★★	★☆☆☆☆	★☆☆☆☆	★☆☆☆☆
黑客武器化能力	★★★★★	★☆☆☆☆	★☆☆☆☆	★☆☆☆☆
编程 / 底层系统推理	★★★★★	★★★★☆	★★★★☆	★★★★☆
自主规划 / 目标执行	★★★★★	★★★★☆	★★★★☆	★★★★☆

美国财长 + 美联储主席 + SEC 主席 + 五大银行 CEO 紧急闭门会议

核心结论: Mythos 具备摧毁全球支付清算系统、银行核心交易系统、证券交易所基础设施的潜在能力，纳入美国国家金融安全应急预案

输入 25 美元 / 百万 token，输出 75 美元 / 百万 token

办公安全现状愈发严峻（数据泄露）



2024年全年监测到37575起有效数据泄露事件，同比23年增加**92%**；

共涉及2598家企业，金融、电商、快递、汽车、本地生活等多个关键行业。

—Gartner

Openclaw引起AI应用爆发



Openclaw的爆发也是行业发展的必然，预示着AGI产业化的第一个拐点出现了；越来越多用户以更低成本拥抱大模型，会催生用户端**大模型应用爆发**。

Agent部署数量和存储容量爆炸式增长

IDC中国企业级存储市场规模预测	2024	2025	2026	2027	2028
厂商收入（百万美元）	5096	5481	5798	6078	6289
YOY Growth %		7.6%	5.8%	4.8%	3.5%

大模型的数据类型包含文本、图片、音频、视频等多模态数据，呈现爆炸式增长，Gartner预测，到2025年AI的内容生成会占到整个数据量10%。

AI从初期转型到大规模深度使用，看不清，管不住，成本高

早期启动AI推广阶段



看不清，导致AI转型慢

- 开发了几个Agent，如何推动转型？
- 都说在转型，实际各部门tokens和GPU根本没用起来
- 本地、云端、租赁各种算力买了很多，无法统一衡量各部门算力使用情况，推动各部门AI转型无凭无据

多部门开启AI转型后



管不住，导致AI资源浪费

- 管理全靠“人+制度”！各部门都说GPU/Token不够用，怎么管理算力都用在价值场景不清晰
- 用不稳，单算力供应商和模型服务故障或流量突发，业务不可用
- 安全差，数据非法、泄密导致生产经营受阻

大规模深度使用AI后



成本高，AI投入远超预期

- 效果好了，成本指数级飙升！
- AI Coding、数字员工等明星AI场景，都用最好效果的模型，要么tokens调用费用极高（1W/人/月起步）！要么本地显卡成本极高（5W/人起步）！

从“问答知识库”到“主动智能体”：AI驱动产业效能升级



第一级 问答类智能体

- 单次推理问答
- 基于预训练知识被动响应
- 内容生成与问答为主

主要风险

- 输出合规性、敏感信息、虚假事实风险
- 输入隐蔽性、语义化渗透、静态过滤失效风险



第二级 工作流类智能体

- 短程预配置业务流
- 涉及完整工作链路
- 业务流固定、自动化执行

主要风险

- 业务连续性风险
- AI供应链风险、AI组件脆弱性
- 传统信任边界失效、授权边界模糊



第三级 数字员工类智能体

- 长程自主规划业务流
- 具备核心规划与决策能力
- 自主理解、规划、执行
- 环境感知能力

主要风险

- 模型内生性局限、逻辑黑盒（看不清）
- 非受控物理执行、决策主权丧失（管不住）
- 治理失能、权责对等失效（追不到）

智能体安全风险的战术：攻击面分析

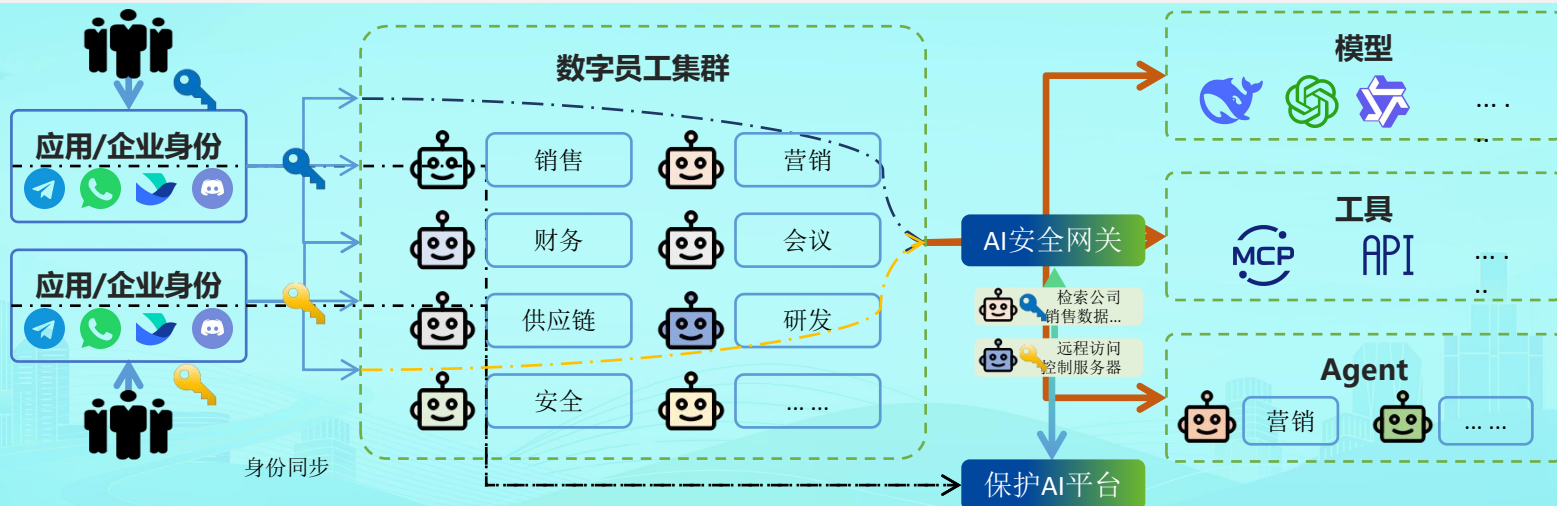


深信服认为 像保护数字员工一样保护企业AI应用

“数字员工”逐步上岗的安全挑战，需要延展新的治理模式

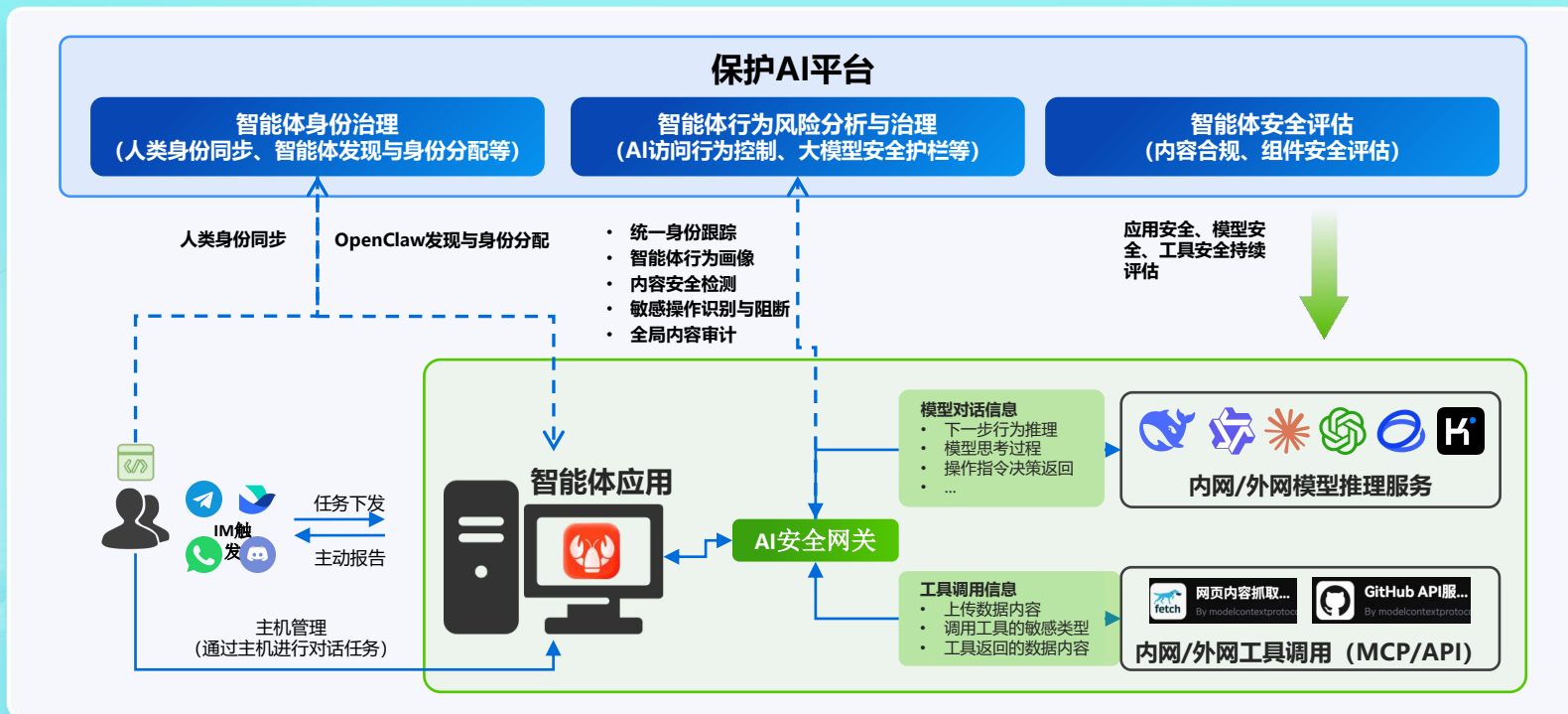
智能体作为“数字员工”具有身份的“双重属性”：① 作为“应用”或“程序”具有横向的静态可见性或持久化链接能力；② 作为人类的“执行代理”可以继承人类的权利和限制。

因此，人->智能体的访问应该由传统的IAM或零信任系统统一管理。智能体->模型、工具、其他智能体的访问，由符合身份人类凭据的权限和智能体横向可见性和持久化权限的交集决定。



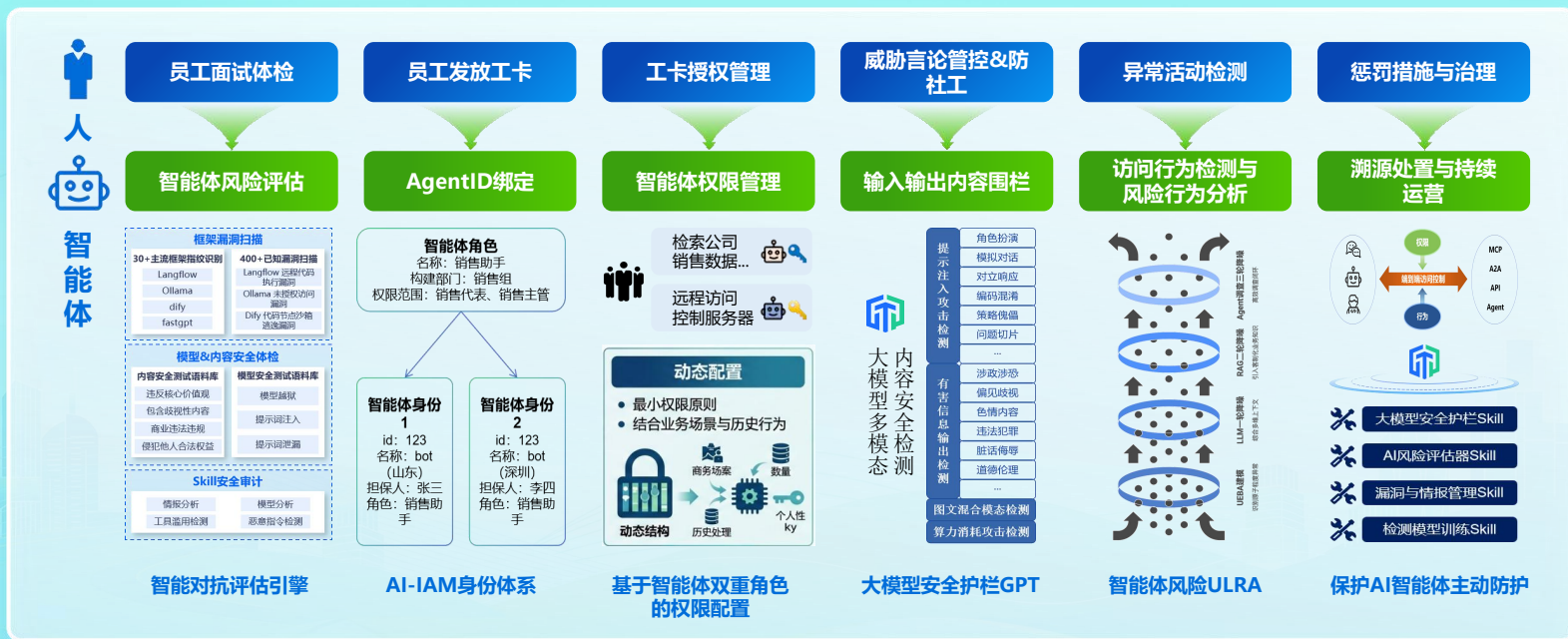
人的访问链路：——> 智能体的访问链路：——>

深信服AI安全平台保护架构



数字员工的治理&防护旅程

保护AI的治理的本质在于**以身份为基础，聚焦智能体行为意图**。它不再是对 AI 能力的限制，而是以AI保护AI，构建全链路治理体系，**让AI可见可控可控，助用户安全放心用AI。**



深信服认为，通过统一治理，让每一份算力
看得清、管得住、省得下、用得稳、更安全

深信服AI算力网关价值主张



强大的可视能力：各角色Token用量、行为可视，各部门AI转型推动不再难

强大的可视能力

多维度可视

基于部门/个人/应用级Token用量和成本结算：推动各部门AI转型，发现AI转型明星员工，加速优秀AI应用推广



深度内容和行为可视

调用供应商、调用状态、调用行为、输入输出token和内容、全面可视

AI资源滥用可追溯，效能提升有依据



精细管理：Token级智能限流、配额管理，让每一分算力 用有所值

管理员基于Token费用进行部门级成本控制

业务组管理

管理员为下属组织设置Token费用配额

名称	描述	组管理员	API Key 数	可用Token数	当月配额 (已用/剩余)	操作
默认分组	系统默认分组	-	8	-	无限制	默认分组不支持操作
测试部门	//部门	张三	必 10	10	¥200.00/¥400.00	查看用量 设置价格 更多
运营部门	//部门	李四	-	-	¥200.00/¥400.00	查看用量 设置价格 更多

编辑 删除

业务组管理

下属组织的员工和AI应用只能在费用配额内消费，超出配额无法消费

名称	描述	关联 API Key 数	关联Token数	配额 (已用/总量)	操作
测试部门	//部门	8	-	无限制	查看用量
运营部门	//部门	10	10	¥200.00/¥400.00	查看用量

保障重要业务、关键人员AI体验，为非关键需求设置限流策略

Qwen2.5-7B-Instruct

路由统计 模型列表 API Key

开发一箱 开发二箱

无限制 无限制 无限制

API Key 数 8 3

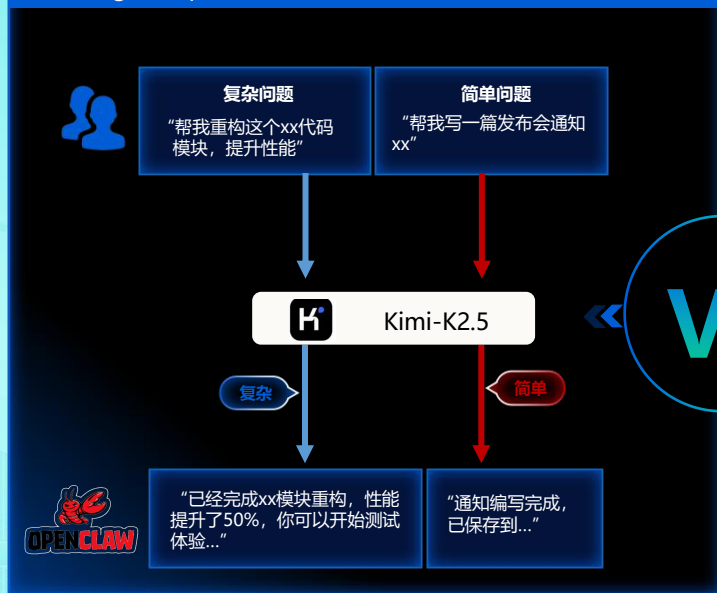
使用有效期 2019-12-08 1... 2019-12-08 1... 2019-12-08 1...

操作 删除 删除 删除

基于请求频率、输入输出token、并发连接数、访问限制进行精准限流

创新自研的智能路由引擎越用越省钱

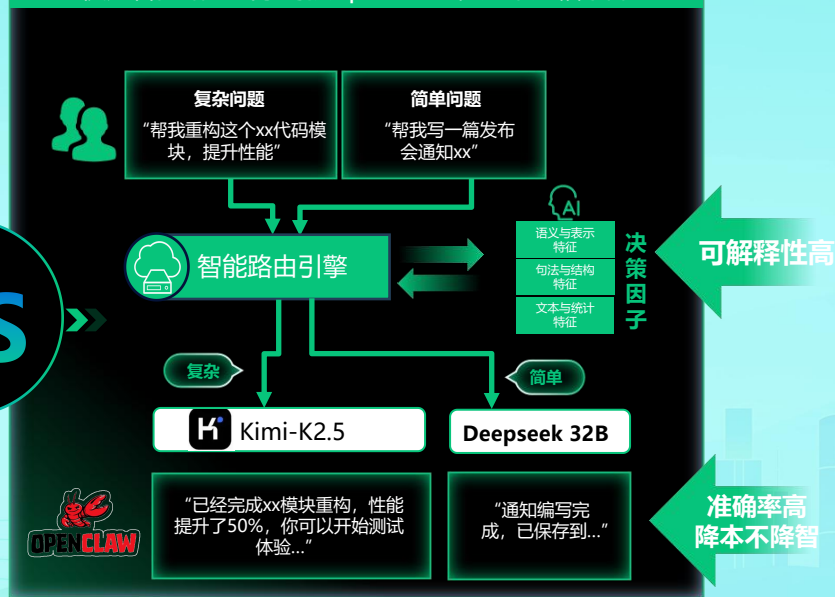
AlCoding使用openclaw，所有问题都使用顶尖模型，成本非常高



贵! 花费¥80万

*3000人研发团队一个月调用外部模型费用

使用智能路由引擎对接openclaw，成本大幅下降



可解释性高

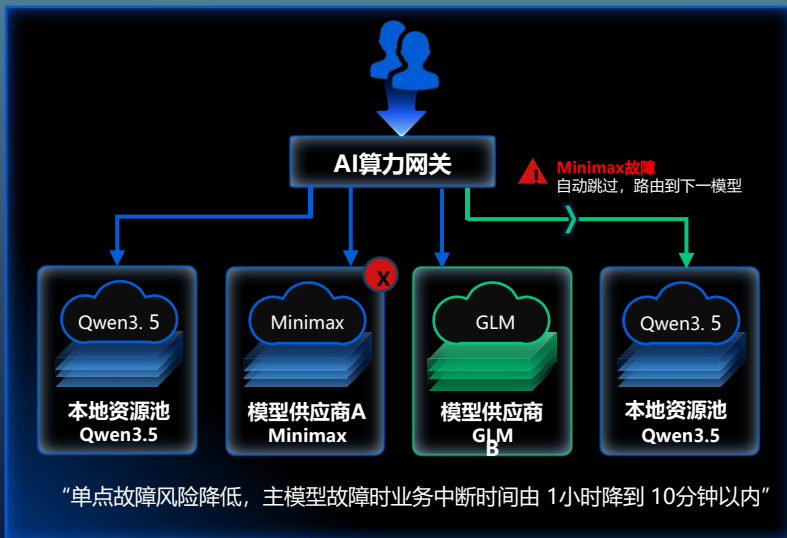
准确率高
降本不降智

省! 仅花费¥40万

*使用智能路由引擎后，3000人研发团队一个月调用外部模型费用

多维度模型聚合路由 – 避免多来源模型单点故障，云上云下模型弹性切换

避免多来源模型单点故障，保障服务流畅稳定



云上云下模型弹性切换，提升AI业务健壮性



深信服AI全栈解决方案

让每一个用户的生产级AI更简单、更安全





SANGFOR
深信服科技

谢谢

让每一个用户的数智化更简单 更安全！

