东吴证券 SOOCHOW SECURITIES

计算机行业深度报告

AI 安全深度报告：智能体拓展安全边界，传统网安有望迎来新一轮产品升级

2026 年 10 月 21 日

增持（维持）

证券分析师 王紫敬
执业证书：S0600521080005
021-60199781
wangzj@dwzq.com.cn

投资要点

- **AI 安全正在成为企业智能化转型的刚性必要投入，应用防护开辟出网络安全行业全新增量市场。**随着大模型从内容生成演进为智能体执行业务任务，安全需求由输入输出审核拓展至完整业务流程控制。早期大模型主要承担问答、检索和文本生成，模型厂商通过安全训练、接口限制与内容过滤，可以覆盖相当一部分通用风险；智能体接入数据库、业务账号和执行工具后，错误判断的影响扩大至企业生产与数据资产，企业需要独立设置数据访问边界、操作权限与审批流程。安全需求将沿着办公辅助、业务协同、核心生产逐级深化，防护范围由内容合规、敏感信息保护扩展至跨系统身份认证、运行隔离、异常中止和恢复机制，推动单客户采购价值提升。AI 为网安行业带来新增防护需求与既有业务提效两条增长路径：大模型安全网关、权限治理、数据防泄露等产品直接创造收入增量；AI 赋能告警研判、漏洞分析和辅助响应，降低人工分析成本、缩短处置时间。行业商业化将首先体现为独立 AI 安全产品销售，随后通过平台扩容、续费及持续运营服务扩大客户价值。

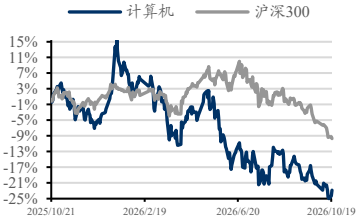
- **AI 安全产业链呈现分层协作格局，企业部署入口决定第三方安全厂商的市场竞争地位。**模型厂商掌握权重、训练过程和部署策略，主导安全训练、模型对齐及危险能力评估；第三方厂商聚焦企业侧应用与基础设施防护，将通用安全能力与客户的数据边界、权限结构和业务流程结合，专业评估机构参与独立验证。企业同时使用外部模型接口、内部部署模型及不同部门开发的智能体，容易形成策略不一致、日志分散和责任边界不清，进一步提高跨模型统一安全管理的价值。第三方厂商通过联动身份认证、数据治理和安全运营系统，形成覆盖访问、执行与审计的产品体系。海外头部模型企业联合安全厂商开展防御项目，验证了模型能力与安全交付体系结合的产业路径：发现漏洞后，仍需完成资产定位、修复协调、补丁验证和持续跟踪。传统网安厂商的核心壁垒来自长期部署形成的数据连接、客户授予的处置权限及行业交付经验，能够结合业务上下文识别风险，并将处置嵌入生产流程。

- **国内 AI 安全商业化已经落地，行业盈利兑现要看收入结构和交付效率。**国内厂商已披露 AI 安全相关收入增长，新增防护需求正转化为实际收入贡献。各家公司统计范围存在差异，需要区分独立 AI 产品收入与 AI 赋能传统产品收入，并结合合同扩容、续费及交付成本评价实际增量。商业化将经历单点防护、统一治理、持续运营三个阶段：初期围绕具体应用采购安全网关，随后随模型和智能体数量增加，扩展统一身份、策略与审计能力，最终形成持续监测、攻击测试和安全运营服务。标准化程度越高，厂商越容易降低重复交付成本、提升单客户价值。行业产品重心将由单纯检测向独立执行控制延伸，模型防护难以完全消除提示注入风险，应结合最小权限、高风险操作审批及下游系统授权校验，限制检测失效后的影响范围。叠加国内外监管对安全评估、风险记录及事件报告的要求，能够将权限、审计和持续监测沉淀为标准产品的厂商，有望随客户 AI 应用扩张实现持续增长。

- **投资建议：**围绕产品落地、客户拓展、盈利修复筛选标的。**相关公司包括：深信服、绿盟科技、亚信安全、中新赛克、启明星辰、奇安信、安恒信息及天融信。**

- **风险提示：**企业 AI 部署与安全采购不及预期、传统网络安全需求持续承压、竞争加剧导致产品价格及毛利率下降、产品防护效果和可靠性不及预期、盈利修复及现金流改善不及预期、估值波动风险。

行业走势



相关研究

《下一个 AI coding，AI Developing 》
2026-09-14

《商业航天深度报告：航天发动机》
2026-09-07

内容目录

1. AI 安全成为企业智能化的必要投入，应用防护打开新增市场 4
1.1. 大模型从生成内容走向执行任务，安全需求由内容审核延伸至业务控制 4
1.2. AI 防护形成新增收入，安全运营智能化改善交付效率 6
2. 模型厂商与安全厂商分层协作，企业入口决定第三方竞争位置 8
2.1. 模型原生安全由厂商主导，应用与基础设施防护具备广泛第三方空间 8
2.2. 海外合作进入产品交付阶段，先进模型能力通过安全生态落地 9
2.3. 数据连接、执行权限和行业交付构成传统厂商核心壁垒 10
3. 产品商业化已经启动，盈利兑现取决于收入结构与交付效率 11
3.1. AI 安全收入开始兑现，行业进入结构性增长阶段 11
3.2. 权限与审计控制外置，推动安全产品从检测向执行延伸 11
4. 投资建议 13
5. 风险提示 13



图表目录

图 1：企业应用从传统软件向 AI 应用和智能体扩展 4
图 2：企业智能体的独立运行治理体系 5
图 3：智能体间接提示注入风险与防护建议 6
图 4：大模型应用防火墙部署于应用与模型之间 7
图 5：Prisma AIRS 企业 AI 安全防护体系 8
图 6：安全平台统一连接模型、智能体与工具 9
图 7：Charlotte AI 告警分析报告界面 10
图 8：AI 辅助研发的外置控制与人工审核流程 12

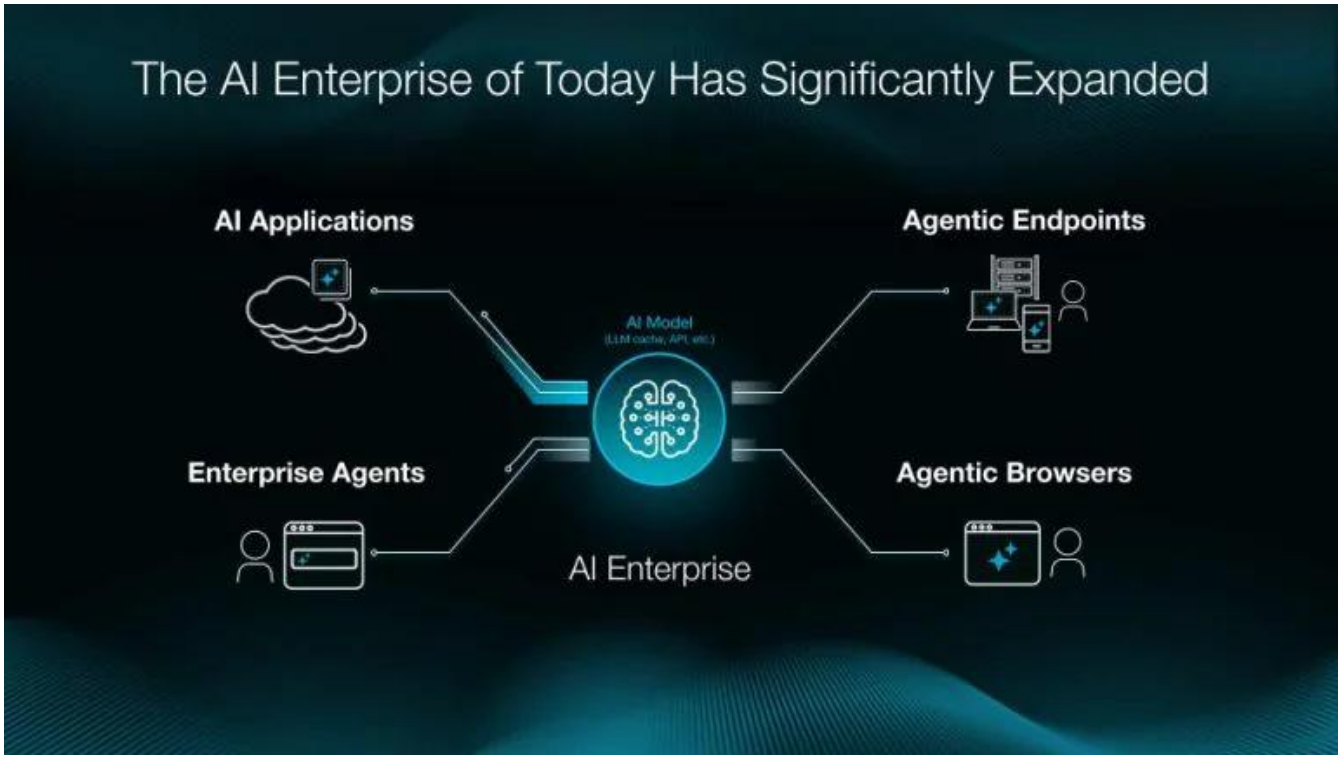
1. AI 安全成为企业智能化的必要投入，应用防护打开新增市场

1.1. 大模型从生成内容走向执行任务，安全需求由内容审核延伸至业务控制

AI 安全的需求拐点来自智能体进入生产系统。随着大模型接入企业数据、业务账号和执行工具，安全防护对象由输入输出内容扩展至完整业务流程，独立安全产品的采购必要性明显提高。

早期大模型主要承担问答、检索和文本生成任务，风险集中于内容违规、信息泄露和回答错误。模型厂商通过安全训练、接口限制和内容过滤，可以覆盖相当一部分通用风险。企业部署智能体后，模型开始调用数据库、运行代码、修改文件，并通过业务账号执行操作，错误判断的影响相应扩大至数据资产和生产运行。

图1：企业应用从传统软件向 AI 应用和智能体扩展



数据来源：发改委，东吴证券研究所

以财务智能体为例，读取发票、核验合同、生成付款建议和提交付款，分别对应不同的数据权限与审批要求。模型准确理解任务，只解决了业务执行中的一个环节。企业仍需明确允许读取哪些合同、能否访问其他部门的数据、付款操作由谁审批，以及异常任务如何终止。这些规则需要与组织架构、业务系统和内部控制共同落实。

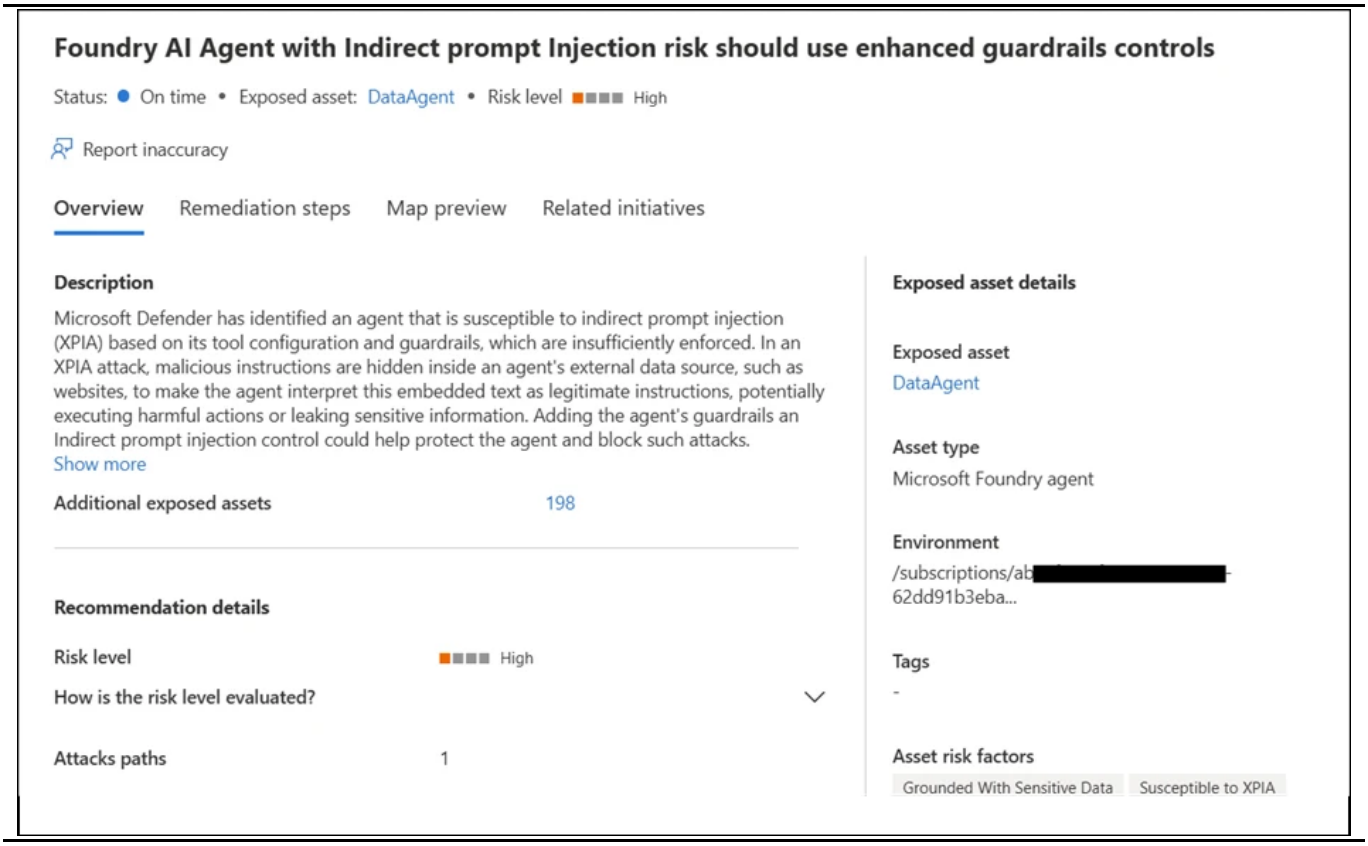
图2：企业智能体的独立运行治理体系



数据来源：亚信安全官网，东吴证券研究所

模型能力越强、接入系统越多，企业对独立权限管理和运行审计的需求越高。 提示注入也由对话窗口向邮件、网页、知识库和工具返回结果扩散，攻击者可能通过智能体读取的外部内容影响其后续行为。OWASP 已将提示注入和过度代理权限列为大模型应用的重要风险，强调限制工具功能、执行权限和自主操作范围。

图3：智能体间接提示注入风险与防护建议



数据来源：微软官网，东吴证券研究所

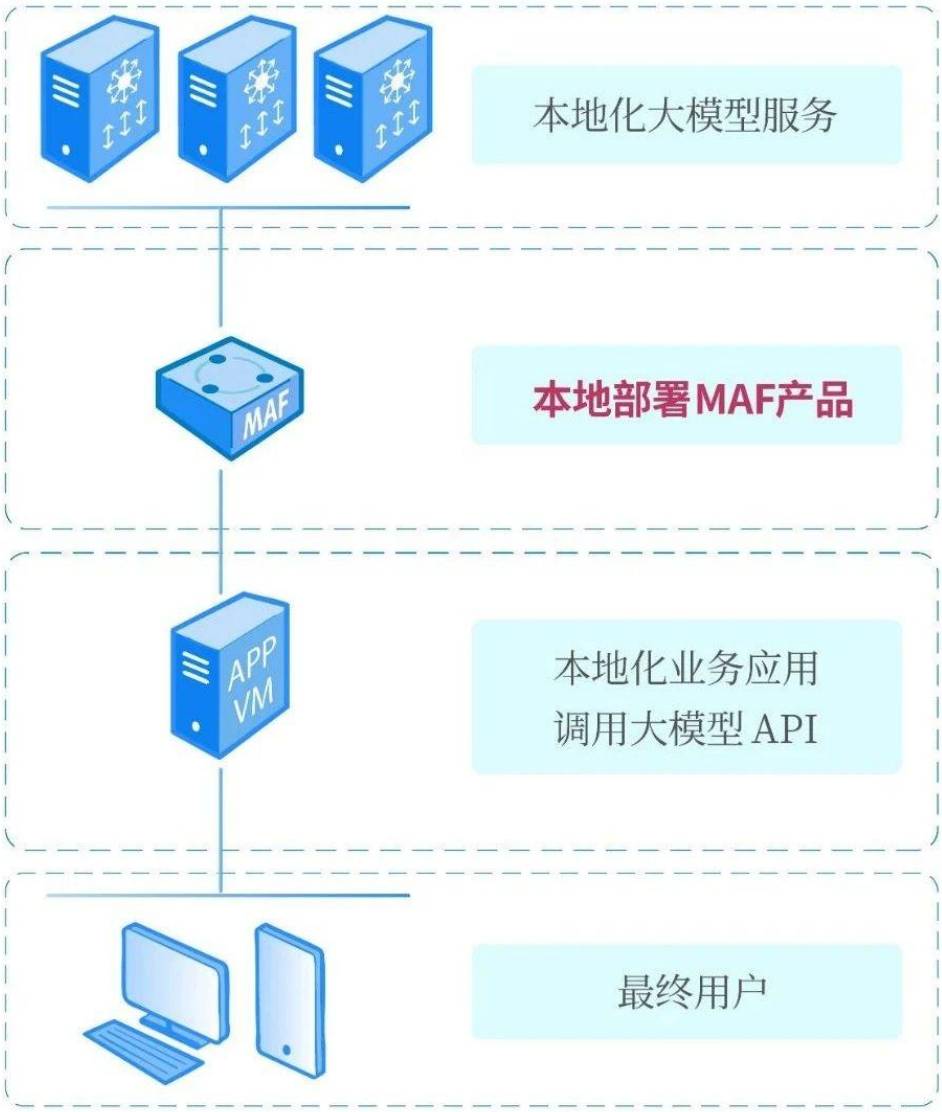
我们判断，AI 安全需求将沿着办公辅助、业务协同、核心生产三个阶段逐步加深。办公辅助主要需要内容合规和敏感信息保护；业务协同需要跨系统身份认证、权限控制和操作审计；核心生产进一步要求高风险操作审批、运行隔离、异常中止和恢复机制。随着应用进入更深的业务环节，安全产品覆盖范围与单客户价值量均有提升空间。

1.2. AI 防护形成新增收入，安全运营智能化改善交付效率

AI 对网络安全产业的贡献包括新增防护需求和既有业务提效，两条路径分别作用于收入与成本。

保护 AI 应用本身，主要对应大模型安全网关、数据防泄露、智能体权限治理和运行审计等产品。企业引入新的模型和智能体，需要增加相应控制能力，有望形成独立采购或原有合同扩容。

图4：大模型应用防火墙部署于应用与模型之间

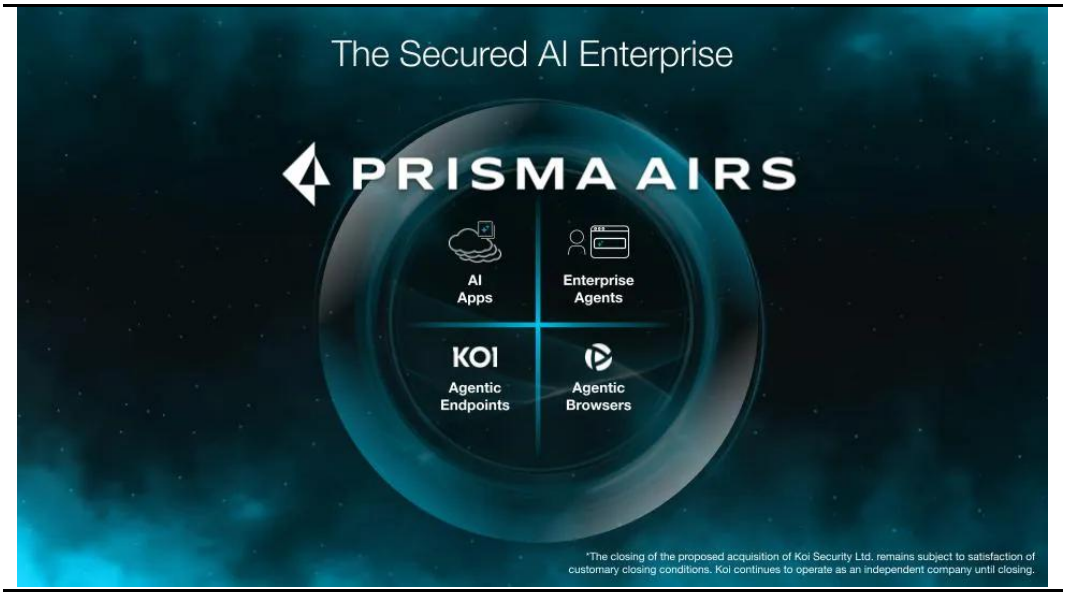


数据来源：启明星辰官网，东吴证券研究所

利用 AI 提升安全效率，主要对应告警研判、漏洞分析、威胁调查和辅助响应。其价值首先体现在减少人工分析、提高服务覆盖和缩短处置时间，随后通过产品升级、续费和托管服务体现为收入。单纯接入大模型不等于新增收入，客户付费和交付成本变化才是商业化的最终结果。

Palo Alto Networks 的 Prisma AIRS 已覆盖 AI 应用、模型、数据及智能体防护，并提供运行时防火墙、模型安全和红队测试等功能。产品覆盖范围表明，企业 AI 安全正在由单点内容过滤向完整运行环境保护延伸。

图5：Prisma AIRS 企业 AI 安全防护体系



数据来源：Palo Alto，东吴证券研究所

我们认为，本轮行业增长将首先体现为 AI 安全产品销售，随后体现为平台扩容和持续服务收入。拥有独立收费产品的厂商更容易获得早期收入增量，拥有企业数据入口和安全运营体系的厂商则更容易实现后续交叉销售与客户扩容。

2. 模型厂商与安全厂商分层协作，企业入口决定第三方竞争位置

2.1. 模型原生安全由厂商主导，应用与基础设施防护具备广泛第三方空间

AI 安全产业链具有明确的分层特征。模型厂商掌握底层训练与安全策略，第三方厂商围绕企业部署环境承担持续防护，专业评估机构参与独立验证。

模型原生安全主要包括安全训练、对齐、危险能力评估，以及欺骗性行为和失控风险研究。模型厂商掌握权重、训练过程、内部实验和部署策略，在这一层具有明显的信息优势。外部评估机构通过独立方法、对抗测试和更深入的访问权限，提高风险识别和验证的有效性。

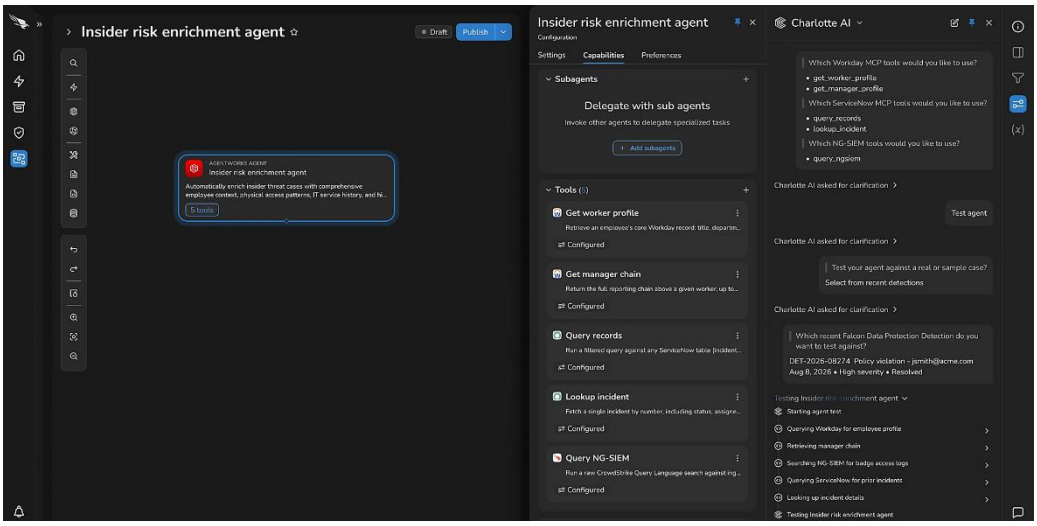
应用与智能体安全直接对应企业生产环境。不同客户具有不同的数据边界、权限结构和审批流程，模型厂商提供的通用策略需要与企业实际系统结合。第三方厂商能够跨模型、跨应用执行统一安全规则，并与身份认证、数据治理和安全运营系统联动，因而具备清晰的产品位置。

基础设施安全则覆盖模型权重、训练数据、访问凭证和训练推理环境。相关风险与

既有云安全、身份安全、数据安全及供应链安全存在较强关联，传统网安厂商可以依托现有产品向 AI 环境延伸。

企业多模型部署进一步提高了统一安全管理的价值。 同一企业可能同时使用外部模型接口、内部部署模型和不同部门开发的智能体，分别配置安全策略容易造成权限不一致、日志分散和责任边界不清。能够跨模型执行统一策略的厂商，有望获得更稳定的客户位置。

图6：安全平台统一连接模型、智能体与工具



数据来源：Charlotte AI，东吴证券研究所

2.2. 海外合作进入产品交付阶段，先进模型能力通过安全生态落地

OpenAI 的 Daybreak 设置 Blue 与 Red 两类访问权限，分别服务经批准的防御工作和更高阶的授权安全研究，将模型能力开放与访问控制结合。2026 年 9 月，OpenAI 宣布投入 10 亿美元，支持一线防御机构获得补贴性访问、培训、技术支持及合作服务。

Anthropic 的 Project Glasswing 将 Claude Mythos Preview 提供给合作机构开展防御性安全工作，参与方包括 AWS、微软、CrowdStrike 及 Palo Alto Networks 等。Anthropic 承诺提供最高 1 亿美元使用额度，并于 2026 年 6 月宣布扩大项目覆盖范围。

上述合作验证了模型能力与安全交付体系结合的产业路径。 漏洞发现只是安全工作的一部分，企业还需要判断受影响资产、协调修复窗口、验证补丁效果并持续跟踪风险。模型提高分析能力，安全厂商负责将能力嵌入实际业务环境，客户最终购买的是风险处置结果。

第三方评估也在向更深入的参与方式演进。Anthropic 首席执行官 Dario Amodei 于 2026 年 9 月提出嵌入式外部评估机制，并承诺给予评估人员接近员工级别的访问权限及独立发布结论的空间。这一安排体现了扩大外部验证的方向，现阶段属于已宣布的治

理承诺。

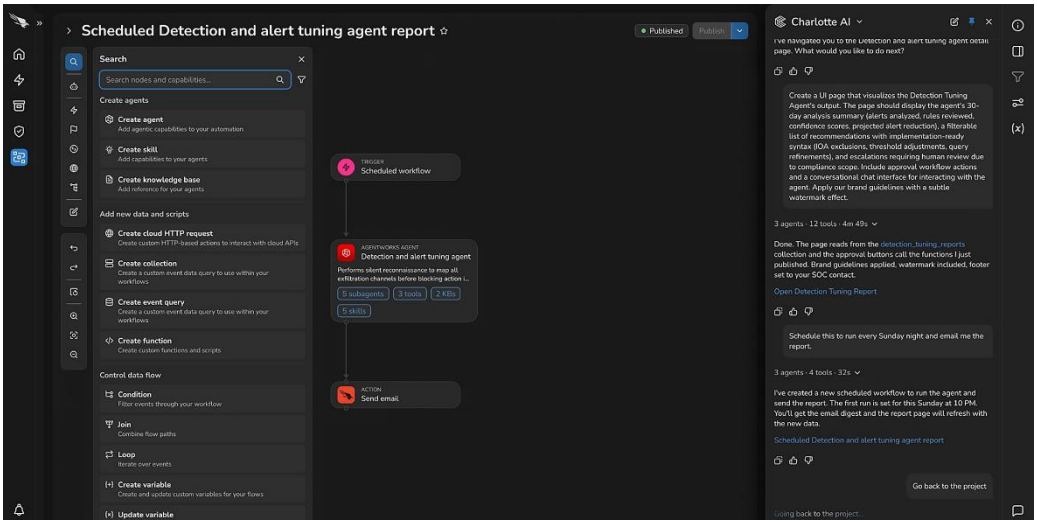
我们判断，安全产品与独立评估将形成不同的商业市场。前者服务广泛的企业客户，具备平台扩容和持续运营空间；后者服务对象更集中，对研究能力和独立性要求更高。国内上市公司的近期增量主要来自企业应用防护和安全运营升级。

2.3. 数据连接、执行权限和行业交付构成传统厂商核心壁垒

传统网安厂商的竞争优势集中于企业环境中的长期部署位置。模型可以提升分析效率，持续获得高质量数据并可靠执行处置，仍需要已有产品和客户授权。

安全分析需要结合终端行为、身份信息、网络连接与历史事件。同一次数据库访问，在不同账号、设备和业务背景下，可能分别对应正常操作、权限错误或数据窃取。长期部署形成的数据连接和业务上下文，能够提高模型分析结果的实用性。CrowdStrike 将 Charlotte AI 嵌入 Falcon 平台，正是依托既有安全数据和工作流发挥模型能力。

图7：Charlotte AI 告警分析报告界面



数据来源：Charlotte AI，东吴证券研究所

处置权限进一步提高客户黏性。隔离主机、撤销凭证、终止进程和修改网络策略，均可能影响生产运行。客户需要验证误报控制、审批流程、操作回滚和故障恢复，相关能力依赖长期接口适配与现场交付积累。

行业经验影响产品复制效率。金融机构侧重客户数据与交易权限，政务客户重视部署边界和责任追溯，工业企业强调生产连续性。拥有行业客户基础的厂商，可以将既有权限、数据和运营体系延伸至 AI 环境，降低新产品导入成本。

本轮竞争将更多向具备运行防护、身份权限和数据治理能力的厂商集中。 通用文

本检测容易面临功能集成和价格竞争，能够控制关键访问路径、接入执行系统并承担长期运营的厂商，更有机会保留产品价值和利润。

3. 产品商业化已经启动，盈利兑现取决于收入结构与交付效率

3.1. AI 安全收入开始兑现，行业进入结构性增长阶段

国内 AI 安全已出现实际收入增长，传统业务与新业务的景气分化仍然显著。投资判断需要同时考察新增产品放量和公司整体经营变化。

启明星辰 2026 年上半年大模型应用防火墙收入同比增长 238.15%；天融信同期 AI 安全防护业务收入同比增长超过 130%。两项披露反映出相关产品已形成客户采购。

各公司 AI 收入的统计范围也存在差异。部分披露对应独立 AI 产品，部分包括 AI 赋能后的传统产品收入，两者需要分别评价。前者更接近新增产品商业化，后者还需要结合合同金额、续费变化和交付效率判断实际增量。

我们预计，商业化将经历单点防护、统一治理和持续运营三个阶段。 初期客户围绕具体应用采购安全网关；随着模型与智能体增加，统一身份、策略和审计需求提高；进入长期运行后，持续监测、攻击测试和安全运营服务进一步扩大采购范围。产品标准化程度越高，厂商越容易提高单客户收入并降低重复交付成本。

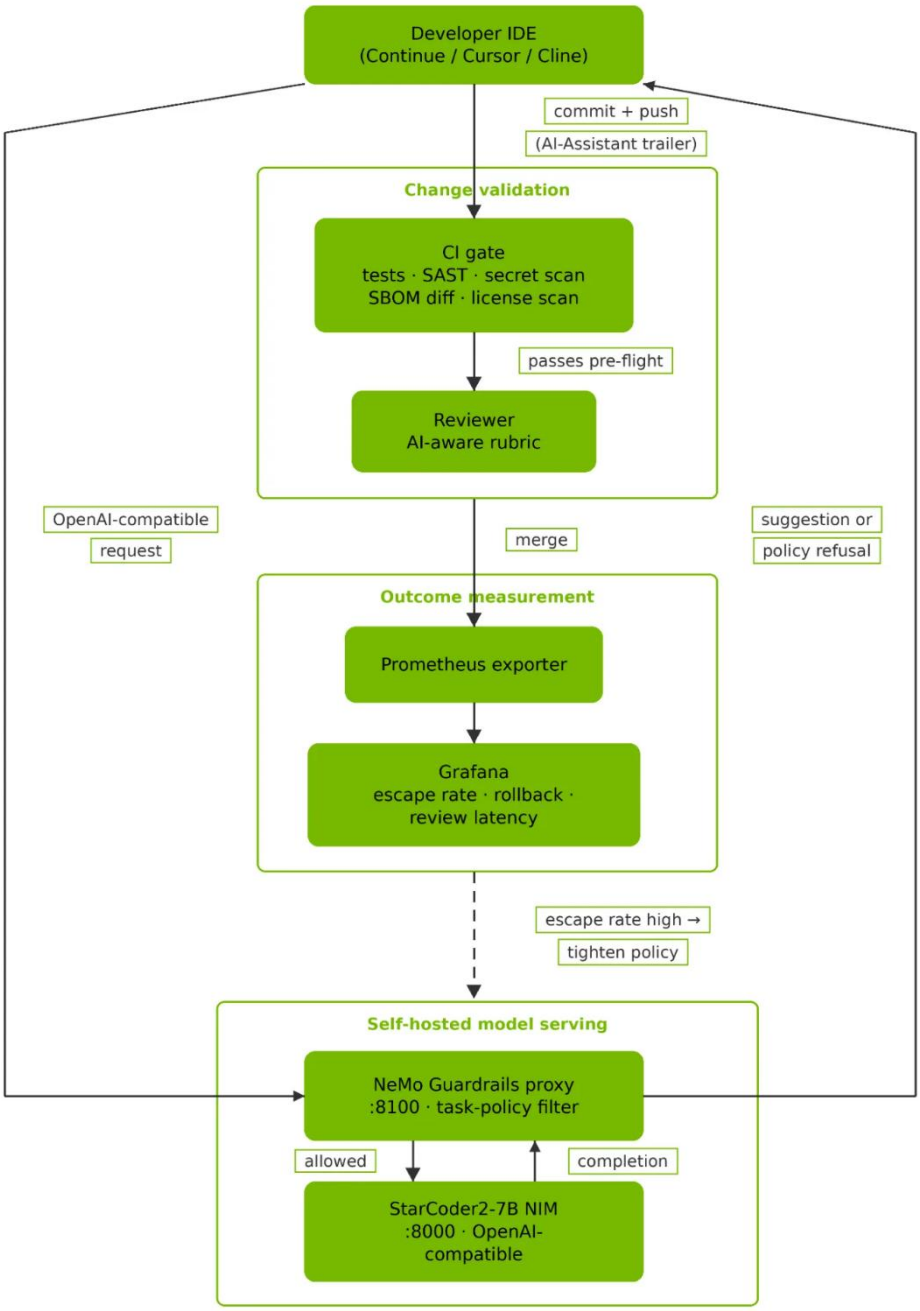
3.2. 权限与审计控制外置，推动安全产品从检测向执行延伸

模型检测并不能替代独立的执行控制，权限管理和运行审计将成为 AI 安全产品的重要增量。

使用 AI 识别攻击，可以提高复杂语义和多轮交互的检测覆盖，也会引入漏检、误报和判断依据难以验证的问题。检测模型给出的解释需要与原始输入、工具调用和实际执行结果共同核验。企业需要在模型之外设置明确的资源访问与操作边界，使检测失效时仍有其他控制环节限制影响。以读取客户资料并发送邮件的智能体为例，资料访问、收件人范围和发送操作应分别受控。模型即使受到恶意文档诱导，权限系统仍应限制其读取未授权数据，执行端仍应拒绝越权发送。OWASP 的防护建议也强调分层防御、最小权限和对高风险操作设置审批。

面向递归自我改进等更高阶风险，控制范围还需要延伸至代码修改、训练任务提交、算力申请和部署发布。模型参与研发的风险，取决于其获得的实际资源与执行权限。相应产品需求包括受控运行环境、资源配额、凭证隔离、独立审批及紧急停用机制。

图8：AI 辅助研发的外置控制与人工审核流程



数据来源：Charlotte AI，东吴证券研究所

监管要求也将提高风险记录和责任追溯的重要性。国内《生成式人工智能服务管理暂行办法》对具有舆论属性或者社会动员能力的服务提出安全评估和算法备案要求；欧盟对具有系统性风险的通用 AI 模型提出模型评估、风险缓解、严重事件报告和网络安全保护义务。具体要求应按适用对象理解，不能一概表述为所有 AI 应用均须购买第三方认证。

我们看好能够将权限、审计和持续监测沉淀为标准产品的厂商。 这类产品可以随客户应用数量增加而扩容，持续服务属性也有利于提高收入稳定性。

4. 投资建议

国内标的选择应围绕产品商业化、客户扩展与盈利修复展开。

相关公司包括深信服、绿盟科技、亚信安全、中新赛克、启明星辰、奇安信、安恒信息及天融信。

5. 风险提示

企业 AI 部署与安全采购不及预期。 若企业大模型应用停留在试点阶段，生产系统接入和独立安全预算释放较慢，相关产品收入增长可能低于预期。

传统网络安全需求持续承压。AI 安全业务仍处于扩张初期，若传统产品采购下降、项目延期或验收放缓，新增业务可能难以抵消存量业务压力。

竞争加剧导致产品价格及毛利率下降。 模型厂商、云平台和应用平台持续完善内置安全功能，通用检测和基础网关可能面临集成替代，厂商需要持续投入差异化能力。

产品防护效果和可靠性不及预期。 提示注入、越权调用等攻击持续演化，漏检、误报和检测时延可能影响客户部署，安全事故也可能造成交付、赔偿和声誉损失。

盈利修复及现金流改善不及预期。 研发、推理和定制交付成本增加，应收账款回收放缓，可能使收入增长难以转化为利润和经营现金流。

估值波动风险。 部分公司尚未形成稳定盈利，市场定价较依赖新业务增长和扭亏预期。若订单、收入或利润兑现低于预期，估值可能出现较大调整。

仅供内部参考，请勿外传

东吴证券 SOOCHOW SECURITIES

免责声明

东吴证券股份有限公司(以下简称“本公司”)经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本研究报告仅供符合中国证监会《证券期货投资者适当性管理办法》规定的专业投资者使用，且接收人须为本公司认定并建立服务关系的专业机构投资者使用，本公司不会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。本报告不面向普通投资者发布或传播。任何非专业机构投资者收到本报告后，应立即退回并销毁，不得依据本报告内容做出任何投资决策。

在法律许可的情况下，东吴证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

市场有风险，投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息，本公司力求但不保证这些信息的准确性和完整性，也不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。

本报告的版权归本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的，应当注明出处为东吴证券研究所，并注明本报告发布人和发布日期，提示使用本报告的风险，且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数，新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的），北交所基准指数为北证 50 指数），具体如下：

公司投资评级：

买入：预期未来 6 个月个股涨跌幅相对基准在 15%以上；

增持：预期未来 6 个月个股涨跌幅相对基准介于 5%与 15%之间；

中性：预期未来 6 个月个股涨跌幅相对基准介于-5%与 5%之间；

减持：预期未来 6 个月个股涨跌幅相对基准介于-15%与-5%之间；

卖出：预期未来 6 个月个股涨跌幅相对基准在-15%以下。

行业投资评级：

增持： 预期未来 6 个月内，行业指数相对强于基准 5%以上；

中性： 预期未来 6 个月内，行业指数相对基准-5%与 5%；

减持： 预期未来 6 个月内，行业指数相对弱于基准 5%以上。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况，如具体投资目的、财务状况以及特定需求等，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

东吴证券研究所
苏州工业园区星阳街 5 号
邮政编码：215021
传真：（0512）62938527
公司网址： http://www.dwzq.com.cn