

AI 技术每日分析：Google 把零信任治理推进到 Agent 运行时

中国高技术产业发展促进会新质生产力工作委员会

博雅云创 & 中科创新驱动

2026 年 9 月 16 日

摘要

今天 AI 技术方向的新动态集中在“运行时治理、垂直模型、Agent 审计和数据基础设施”四条线上。Google 发布零信任 Agent 运行时治理方案，把安全边界从提示词和代码检查进一步推进到工具调用前的语义策略判断、跨轮次异常检测与闭环修复；Salesforce 与 NVIDIA 推出 Koa，把近 30 年的 CRM 业务知识通过合成工作流后训练到专门的推理模型中，说明企业软件厂商开始把行业流程直接做进模型，而不是只做 RAG；AIUC 完成 4000 万美元 A 轮融资，把 Agent 审计、认证和保险结合为新的“可信基础设施”；G5 Labs 带着 1400 万美元种子轮从隐身状态走出，尝试用“自然语言意图 + 系统本体”管理多个 Coding Agent；Keewano 则正式推出 KeewanoDB 并融资 1200 万美元，专门为 Agent 保存连续事件序列和实时上下文。AI 产业正在从“谁有更强模型”继续转向“谁能让 Agent 在真实业务里长期、受控、可审计地完成任务”。

Contents

- 一、Google 把零信任安全推进到 Agent 运行时：先判断“意图”，再允许工具执行

2

二、Salesforce 与 NVIDIA 推出 Koa：行业知识开始直接做进推理模型	2
三、AIUC 融资 4000 万美元：Agent 审计、认证和保险开始变成独立产业	3
四、G5 Labs 从隐身状态走出：自然语言意图想成为 Coding Agent 之上的“控制层”	4
五、KeewanoDB 融资 1200 万美元：Agent 需要的不只是向量库，还需要完整事件历史	5
趋势观察	5
参考文献	6

一、Google 把零信任安全推进到 Agent 运行时：先判断“意图”，再允许工具执行

9 月 15 日，Google Developers Blog 发布零信任 Agent 系列第二部分，把安全治理从构建阶段推进到运行阶段。Google 给出的架构不再只依赖正则表达式、SQL 解析器或一次性的单轮测试，而是通过 Agent Gateway 统一拦截用户、Agent、模型和工具之间的交互。

其中包括三层控制：Model Armor 负责在入口和出口检测提示注入、越狱、恶意链接和敏感数据泄露；Semantic Governance Policies 在工具真正执行前，用自然语言策略判断用户意图、会话历史和业务规则是否允许本次动作；Agent Anomaly Detection 则分析整段会话甚至 Agent 群体的遥测数据，识别单次都“合法”、累计起来却异常的多轮操作。

Google 用退款 Agent 举例说明问题：攻击者可以把一笔违规的大额退款拆成多次小额退款，每一次都符合单次规则，但总额最终超过订单金额。此时传统规则看不出问题，只有跨轮次异常检测才能发现累计行为。Google 还展示了异常发现后动态新增策略、无需重新部署 Agent 即可阻断下一次工具调用的闭环。

这条技术路线很值得企业 Agent 借鉴。真正的生产安全不能只检查“模型说了什么”，而要控制“模型准备做什么”。未来 Agent 治理的关键中间层，很可能就是工具调用前的语义授权、跨会话行为检测、硬件身份签名和可动态更新的运行时策略。

二、Salesforce 与 NVIDIA 推出 Koa：行业知识开始直接做进推理模型

9 月 15 日，Salesforce 与 NVIDIA 在 Dreamforce 发布 Koa，这是 Salesforce 首个面向 CRM 的专用推理模型，基础来自 NVIDIA Nemotron

3 Super。Salesforce 没有使用客户真实数据训练 Koa，而是根据近 30 年 CRM 部署经验构造合成 workflow，覆盖销售线索、商机推进、服务工单等多步骤任务。

与通用模型加 RAG 不同，Koa 试图把“企业流程怎么走、什么工具应该什么时候调用、业务规则如何约束动作”直接训练到模型行为中。Salesforce 称，在其 CRM Benchmark 上，Koa 在部分任务中达到或超过领先通用模型，同时错误率降低约 3 倍。这个结果属于企业自测，仍需要外部评测进一步验证。

更重要的是部署方式。Salesforce 掌握模型权重，并把后训练与推理放在自己的信任边界内；面向政府和强监管客户，Nemotron 模型还会进入 Missionforce，可部署到私有云、隔离网络甚至完全气隙环境。这说明垂直 AI 下一阶段并不一定是“行业大模型越大越好”，而是模型是否真正理解流程、工具和权限，并能够在客户自己的边界里运行。

Koa 代表一种很可能扩散的产品路线：通用基础模型提供底座，软件厂商利用多年行业经验完成后训练，把原来沉淀在产品逻辑、实施顾问和业务规则里的知识，转化成领域推理模型。

三、AIUC 融资 4000 万美元：Agent 审计、认证和保险开始变成独立产业

9 月 15 日，Artificial Intelligence Underwriting Company (AIUC) 宣布完成 4000 万美元 A 轮融资，由 Ribbit Capital 领投，并有 First Harmonic 和 Terrain 参与。加上此前 1500 万美元种子轮，公司累计融资达到 5500 万美元。

AIUC 的产品思路很特别，它不是直接卖一个安全扫描器，而是试图复制传统工业里的“认证 + 保险”体系。AIUC-1 标准会针对越狱、幻觉、数据泄露等风险测试企业 Agent，并配合独立审计机构和保险机制形成可

信标志。公司称，目前 Cursor、ElevenLabs、Harvey、UiPath、Lovable 等产品已经参与相关认证。

这类公司出现，说明企业 Agent 进入真实生产后，采购方需要的不只是厂商自己说“安全”。金融、法律、客服、代码等领域都需要第三方证明：这个 Agent 经过哪些攻击测试、风险边界是什么、发生事故由谁承担。

未来 Agent 市场可能出现类似云安全、支付合规和工业认证的第三方服务层。模型能力越强，企业越需要独立评测、标准化审计与保险来降低采用风险。AI 安全也由“技术团队内部问题”逐渐变成采购、法律、保险和公司治理共同参与的问题。

四、G5 Labs 从隐身状态走出：自然语言意图想成为 Coding Agent 之上的“控制层”

9 月 15 日，MIT 教授 Tim Kraska 创办的 G5 Labs 公开亮相，并披露 1400 万美元种子轮融资。G5 的核心观点是：当越来越多代码由 Agent 生成后，企业真正的问题不再是“能不能写代码”，而是多个 Agent 是否理解同一套业务意图、架构约束和安全规则。

G5 把业务需求、架构决定、合规政策和编码标准组织成所谓 System Ontology，并尝试让 Claude Code、Codex 等不同 Coding Agent 都在这一共享语义层下工作。平台希望自然语言意图不只是 Prompt，而是可以被比较、合并、追踪和治理的“上层源代码”。

这种做法与传统规范驱动开发有相似之处，但它更强调“规范与实现持续同步”。如果一个 Agent 修改了底层代码，系统本体也要能够反映变化；如果两个 Agent 提出冲突修改，则先在意图层处理冲突，而不是等代码合并时才发现问题。

这类中间层是否能够稳定维护大型软件系统仍需时间验证，但方向

很清楚：Coding Agent 越强，企业越需要一个比代码更高层的控制面，统一管理业务目标、架构规则、权限与多个 Agent 之间的协同。

五、KeewanoDB 融资 1200 万美元：Agent 需要的不只是向量库，还需要完整事件历史

同一天，以色列创业公司 Keewano 发布 KeewanoDB 并披露 1200 万美元种子轮融资。它认为传统数据库和数据仓库主要面向人类预先定义的查询，而 Agent 经常需要回答“为什么发生了这个结果”，这要求系统保存完整事件顺序，而不是只留下聚合结果。

KeewanoDB 按照用户、设备、交易等实体保存连续事件序列，让 Agent 能够直接查看某个对象从开始到现在的完整行为轨迹，并把部分计算放到数据库内部完成，再只把精简上下文返回模型。公司称这样可以减少 Agent 需要读取的 Token，并提升实时分析效率；相关性能数字目前主要来自企业自身测试。

这一方向说明 Agent 数据基础设施正在与传统 RAG 分化。RAG 擅长从文档里找到相关知识，但很多经营分析、风控、用户行为和运维调查需要时间顺序、状态变化和因果上下文。未来 Agent 可能同时需要文档知识库、事务数据库和事件序列数据库，不同数据结构服务不同推理任务。

趋势观察

今天几条新闻共同指向一个更成熟的 Agent 技术栈：Google 解决运行时意图治理，Salesforce 把行业流程做进模型，AIUC 提供第三方可信认证，G5 管理多个 Coding Agent 的系统意图，Keewano 重新设计 Agent 读取事件历史的数据层。AI 产业的竞争单位正在从“一个模型”变成“模型 + 上下文 + 运行时 + 治理 + 审计”的完整系统。真正进入生产以后，

任务完成率、风险可解释性、恢复能力和数据边界会越来越重要。

参考文献

- 1. Google Developers Blog | 《Build zero-trust AI agents that judge intent, not just syntax》 | 2026-09-15 | 核验 Model Armor、Semantic Governance Policies、Agent Anomaly Detection 与闭环修复。<https://developers.googleblog.com/build-zero-trust-ai-agents-that-judge-intent-not-just-syntax/>
- 2. Salesforce | 《Announcing Koa: Salesforce's First CRM Reasoning Model, Built on NVIDIA Nemotron》 | 2026-09-15 | 核验 Koa 模型、合成训练数据、信任边界与试点。<https://investor.salesforce.com/news/news-details/2026/Announcing-Koa-Salesforces-First-CRM-Reasoning-Model-Built-on-NVIDIA-Nemotron/default.aspx>
- 3. TechCrunch | Salesforce 与 NVIDIA Koa 报道 | 2026-09-15 | 交叉核验领域模型趋势。<https://techcrunch.com/2026/09/15/salesforce-and-nvidias-new-reasoning-model-is-everything-the-ai-labs-should-fear/>
- 4. AIUC | 《Series A to build the confidence infrastructure for frontier AI》 | 2026-09-15 | 核验 4000 万美元融资、AIUC-1 与审计保险思路。<https://aiuc.com/updates/series-a-announcement>
- 5. TechCrunch | AIUC 融资与 Agent 审计报道 | 2026-09-15 | 补充团队、融资与市场背景。<https://techcrunch.com/2026/09/15/early-anthropic-hire-former-metr-coo-have-found-a-way-to-rein-in-rogue-ai-agents/>
- 6. VentureBeat | G5 Labs 平台与 1400 万美元融资 | 2026-09-15 | 核验 System Ontology 与自然语言意图控制层。<https://venturebeat.com/>

technology/should-all-enterprise-code-and-workflows-become-natural-language-g5-labs-thinks-so-and-its-new-g5-platform-does-it-for-you

- 7. Pillar VC | 《G5 Labs (Why We Invested)》 | 2026-09-15 | 补充 G5 团队和产品方向。<https://www.pillar.vc/investing/g5-labs-why-we-invested/>
- 8. VentureBeat | KeewanoDB 发布与 1200 万美元融资 | 2026-09-15 | 核验 Agent 事件序列数据库与融资。<https://venturebeat.com/data/every-user-device-and-agent-gets-its-own-table-in-keewanodb-and-queries-never-join-across-them>
- 9. Keewano | KeewanoDB 产品资料 | 2026-09 访问 | 补充事件序列、Agent 接口与实时分析架构。<https://keewano.com/keewanodb/>
- 10. Salesforce | Air India 扩展 Agentforce 应用 | 2026-09-15 | 背景资料, 观察 Agent 从问答向复杂业务执行迁移。<https://www.salesforce.com/in/news/press-releases/2026/09/15/air-india-accelerates-customer-service-transformation-with-agentforce/>

联系我们，请扫描二维码



新质生产力工作委员会
官方公众号



工业智能算网
gyznsw.cn

新质生产力工作委员会：

中国高技术产业发展促进会新质生产力工作委员会，专注于推动工业人工智能、智能制造、数字化转型等前沿技术发展，为企业提供政策解读、技术咨询和产业对接服务。

工业智能算网：

专注于工业人工智能、新质生产力、工业软件 CAE、智能制造等前沿技术。提供每日动态分析、技术趋势解读、解决方案分享，推动工业智能化转型。

网站地址：<https://gyznsw.cn>