cloud security alliance®

zenity

Survey Report

Enterprise AI Security Starts **With AI Agents**

致谢

主要作者

希拉里·巴伦

贡献者

玛丽娜·布雷戈库 Josh·布克 Ryan·吉福德

设计

Stephen Lumpe 斯蒂芬·卢姆佩
Stephen Smith 斯蒂芬·史密斯

关于赞助商

Zenity 是专为 AI 代理构建的首个安全与治理平台，覆盖 SaaS、自研平台（云）和终端用户设备（端点）。该平台深受财富 500 企业的信赖，通过提供全生命周期覆盖的纵深防御，帮助安全团队自信地采用 AI，从代理发现和姿态管理到实时检测、内联防护和响应。Zenity 以代理为中心，优先考虑代理的行为、访问内容以及调用的工具，消除盲点，并在各环境中执行一致的政策和控制，使组织能够在不牺牲安全的前提下创新 AI 应用。了解更多信息请访问 zenity.io .

目录

执行摘要

与人工智能代理相关的风险已不再是理论上的。调查发现，自主系统在日常工作中已超出预期权限、在预期范围外运行。在许多环境中，这些行为至少偶尔会发生，表明范围违规并非孤立事件，而是一种常规的运行状况。近半数组织报告称经历过涉及人工智能代理的安全事件，且当事件发生时，检测和响应通常需要数小时甚至数天，而非数分钟，从而扩大了互联系统中的潜在暴露窗口。这些模式表明，造成实质性暴露并非源于假设性的未来采用，而是由于自主系统已实现规模化运行——通常是在实时代理清单、持续运行时授权控制和全面行动可追溯性仍需成熟的环境中。

关键发现1：AI代理的采用已在各组织中广泛普及并投入运营。

人工智能代理已日常被大量劳动力使用，43%的组织报告称，超过一半的员工定期使用人工智能代理。采用情况很少集中：只有5%使用单一代理平台，而44%使用两到三个平台，43%使用四个或更多平台，这增加了环境复杂性。

关键发现2：影子人工智能代理正在早期出现。

未经授权的AI代理在采用初期就出现，54%的组织报告称拥有1至100个未经授权的AI代理，即便整体代理数量仍然相对较少。所有权往往不明确：只有15%的报告称76%至100%的代理具有明确定义的所有权，而最常见的所有权范围是26%至50%（占34%），这导致在发生事件时出现责任追究空白。

关键发现3：AI代理范围违规是一种常见的运行状态。

范围违规现象很普遍而非罕见，53% 的组织报告称 AI 代理偶尔或有时会超出预期权限。这些行为会产生实际影响：47% 的报告涉及涉及 AI 代理的安全事件，58% 指出检测和响应耗时五小时或更久，从而延长了暴露窗口期。

关键发现4：缺乏AI代理安全策略，合规将成为默认选项。

治理深受现有监管框架的影响，然而，仅有13%的组织表示已为即将出台的AI相关法规做好了充分准备，这凸显了合规性对齐与运营准备度之间的差距。

外卖

人工智能代理已作为企业数字化工作队伍的一部分在运行，但安全和治理实践尚未完全适应其自主性。采用速度已超过可见性、所有权和控制力，导致无意行为正变得操作上常态化而非异常。应对这些挑战将需要为自主系统量身定制的治理和安全方法，这些方法应与采用规模同步扩展。对于安全领导者而言，这些发现表明需要进行结构性转变：有效的风险面正越来越多地由基础设施和访问权限定义，同时也由自主行为和行动定义。

关键发现

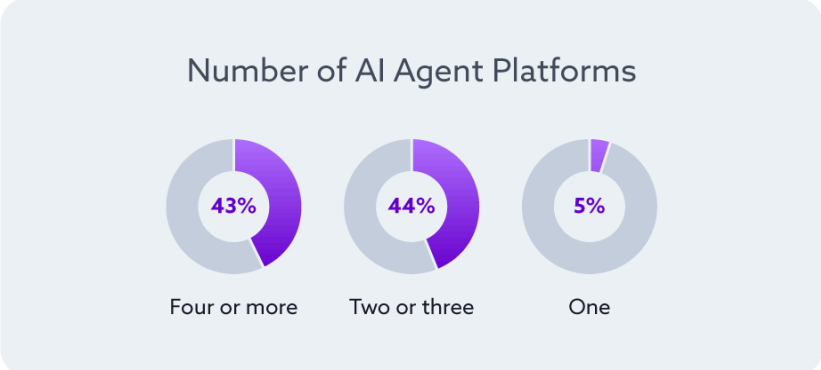
人工智能代理正迅速成为企业运作的有机组成部分，重塑着工作执行方式、决策制定过程以及系统间的交互模式。自主系统在可见性、控制力与问责制方面引入了新的动态，对传统的安全与治理模式构成了挑战。对于那些既要管理风险又要推动规模化创新的组织领导者而言，理解各组织如何应对这一转型以及哪些环节存在空白至关重要。

关键发现1：

人工智能代理的采用已在各组织中广泛普及并投入运营

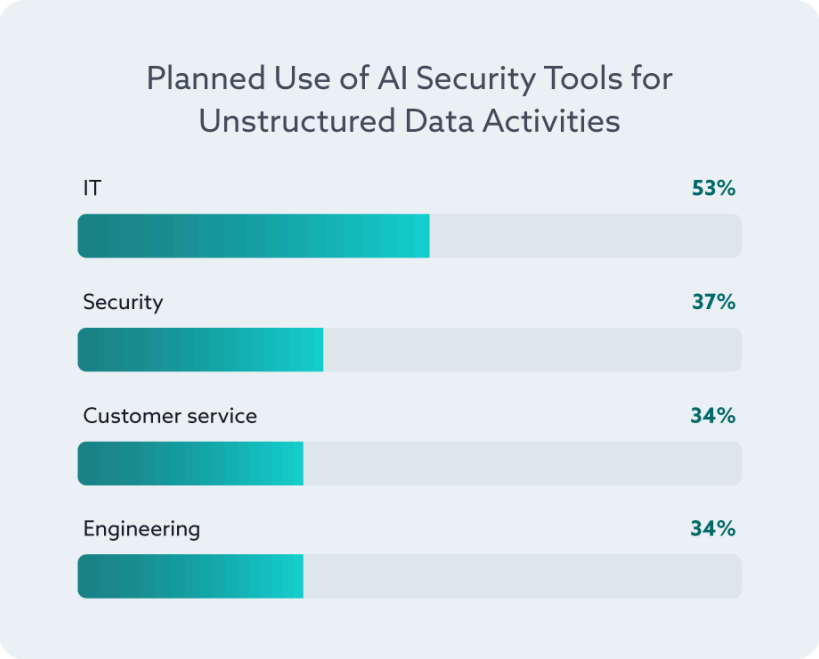
人工智能代理在跨多个垂直领域的众多组织中已达到有意义的规模。相当一部分受访者表示，人工智能代理是大量员工日常工作的组成部分， **43%的人报告说，他们组织中超过一半的员工经常使用AI代理。** 最常见的使用范围落在26-50%的员工之间（占30%），其次是51-75%（占24%）。这种模式反映出采用范围已远超孤立的使用案例，表明AI代理正日益嵌入生产工作流程，而非局限于早期实验。

随着收养（或：采用）变得更加实际化，AI代理的部署方式带来了额外的安全考量。大多数组织跨多个AI代理平台运作，而非单一、集中的环境： **43% 的组织使用四个或更多平台，44% 使用两个或三个，只有 5% 报告使用一个。** 在多平台环境中，应用政策可能更加困难。



始终如一地保持对代理行为的统一可见性，并以标准化的方式收集遥测数据或审计证据。虽然执行效果没有直接衡量，但碎片化平台环境的普遍性表明，在采用生命周期的早期，安全和治理工作必须应对日益增加的复杂性。

随着劳动力使用的扩展，AI代理的应用也扩展到广泛的组织职能中。**使用率报告最多的是IT领域（53%），其次是安全领域（37%）、客户服务（34%）和工程领域（34%）。** 在技术和客户服务职能中都存在AI代理，这表明其应用并不仅限于创新团队。



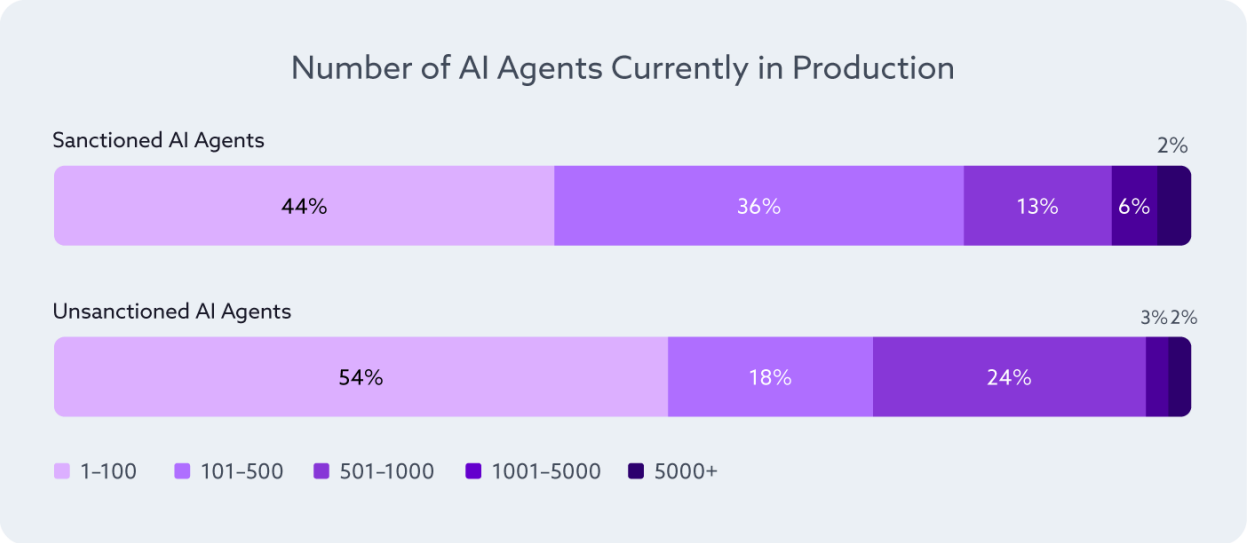
这一模式总体反映了这样一个环境：AI代理已经融入了跨职能的日常工作。广泛的劳动力

使用和跨职能部署表明，AI代理不再仅仅是外围工具；它们已嵌入到核心工作流程中。这种扩展不仅代表技术多样化，也意味着自主决策系统在与敏感数据和工具交互方面的增加。随着代理的持续采用，有效的安全边界正越来越多地从基础设施转向行为。

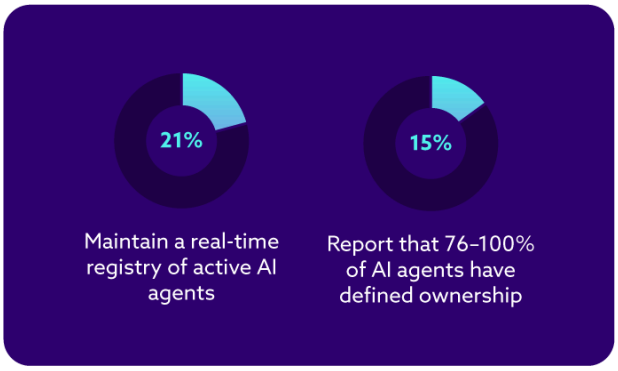
关键发现2：

阴影AI代理正初现端倪

未经授权的AI代理人的存在已显而易见，即便许多组织报告其生产环境中的代理总数相对较少。 **近半数组织（47%）报告拥有101个或更多的非授权AI代理。** 这表明，影子AI活动在采用初期就已出现，而不仅仅是在部署达到大规模时才出现。对于获准的AI代理，最常报告的范围是**1-100名经批准的AI代理机构，被44%的组织引用** 这种分布表明，正式批准的部署仍然集中在较低的业务量范围内，这反映了治理结构可能仍在采用过程中同步发展。在某些环境中，未经批准的人工智能代理的扩展速度几乎与正式批准的相当，这可能反映了去中心化的采用模式，即团队独立于中央监督来部署人工智能代理。



代理数量也显示出与组织规模的高度关联性，规模较大的组织报告的生产环境中部署了更多AI代理。这种关系表明，随着组织规模的扩大，既经批准又未获批准的AI代理群体都在扩张，从而增加了在治理和可见性机制完全建立之前，未获批准的使用现象出现的可能性。



这一挑战因基础可见性机制仍存在实施不均的情况而加剧；另一份云安全联盟（CSA）报告的数据显示， *确保自主AI实体* ，表明 **只有21%的组织维护着活跃AI代理的实时注册记录。** 与此同时，近三分之一的人报告根本没有任何集中库存。当代理数量在没有全面库存和监督的环境中增长时，部署工作超出可见性和集中控制范围的风险会更加突出。

随着代理部署的扩大，所有权清晰度仍然参差不齐。 **只有15%的组织报告称，76%至100%的人工智能代理已明确界定所有权。** 而最常见的回应表明，AI代理物的所有权覆盖范围仅为26-50%（占34%），其次是1-25%（占20%）。这些分布情况表明，AI代理物的明确问责制往往落后于其部署。

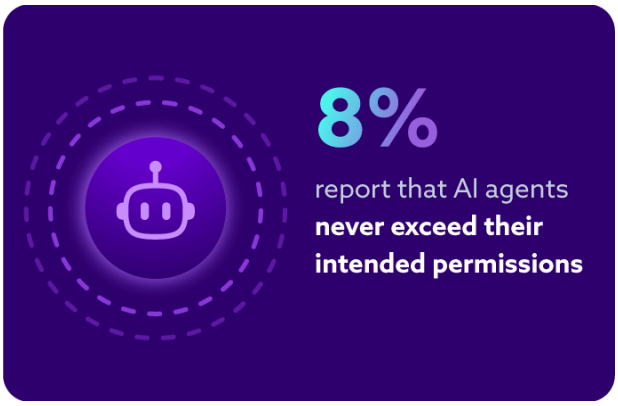
然而，问责不仅取决于明确的归属，还取决于在事件发生时重建代理行为的能力。最近的 *确保自主AI代理* CSA的报告显示，只有28%的组织能够在所有环境中将代理行为追溯到初始人员或系统，而46%的报告称仅在部分环境中具有可追溯性。另外9%报告称没有可追溯性，而16%则不确定此类可见性是否存在。行动可追溯性有限会加剧所有权空白，使得调查事件、归因责任以及控制非预期代理活动更加困难。对于受监管的组织而言，当AI代理与敏感数据或受监管的工作流程交互时，不完整的可追溯性还可能引入额外的法律和审计风险。

综合来看，影子AI代理的存在、所有权定义不均以及部分行动可追溯性，都指向了随着AI代理应用而日益显现的结构性可见性差距。随着组织规模扩大和代理群体增多，治理与问责机制必须以同等速度成熟，否则将落后于实际运营现实。

关键发现三：

AI代理越权行为正变得运营中常见

AI代理范围违规行为，包括执行超出预期结果的操作，似乎是一种常规的运营现象，而非边缘情况。例如，一个旨在建议供应商的采购AI代理，可能会转而自动向供应商发送报价请求邮件。 **只有极少数组织报告称，人工智能代理从未超出其预期权限，仅占8%。** 相比之下，多数人表示人工智能代理偶尔或有时会超出权限（53%），这表明范围违规是日常事务中反复出现的一部分。



而非孤立异常，人工智能代理会定期超出其设计边界运行。在不允许人工智能代理的环境下，不完整的库存和监管不足进一步限制了组织全面评估此类行为范围的能力。

随着AI代理频繁超出预期范围，出现这种情况也就不足为奇了。 **近半数组织（47%）表示在过去12个月内经历过涉及人工智能代理的安全事件。** 据报告，检测和响应的时间线通常在数小时或数天范围内，其中38%表示为5至24小时，另外20%报告称持续一天或更长时间。即使这些时间段也代表着有意义的暴露，因为在不解决事件的情况下，非预期的代理行为可能会在连接的系统中持续存在。来自IBM的行业研究显示 *数据泄露报告成本* 表明识别和遏制漏洞的平均时间以月为单位而非小时，突显了感知到的AI代理事件响应与漏洞检测和遏制更漫长、更复杂的现实之间的潜在差距。这些发现共同表明，尽管组织可能认为他们正快速响应AI代理事件，但有限的可见性和不完善的监控可能掩盖了暴露的全面范围和持续时间。在自主系统跨互联平台运行的环境中，即使是短暂的遏制延迟也可能在修复措施实施前，允许非预期行为得以蔓延。

这些运营现实与CSA报告相符。 *确保自主AI代理* 这表明基础运行时控制仍然有限。只有11%的组织报告完全实施了运行时授权策略执行，9%的组织报告完全采用了零信任原则用于AI代理，而且仅仅

23%的人报告已完全实施会话录制或审计日志记录。细粒度访问控制（12%）和持续身份验证（13%）也仍然不常见。这些控制的有限采用为所观察到的检测和遏制挑战提供了结构性背景，这进一步强化了事实：许多组织在缺乏全面保障措施的情况下管理自主系统。

在基础运行控制受限的环境中，检测和监控能力本身就受到限制。这体现在当前用于检测人工智能代理特定威胁的工具所报告的置信度水平上。 **只有16%的组织表示高度自信，而44%则表示他们只是稍微自信，或者根本不自信。** 这种缺乏信心表明许多人

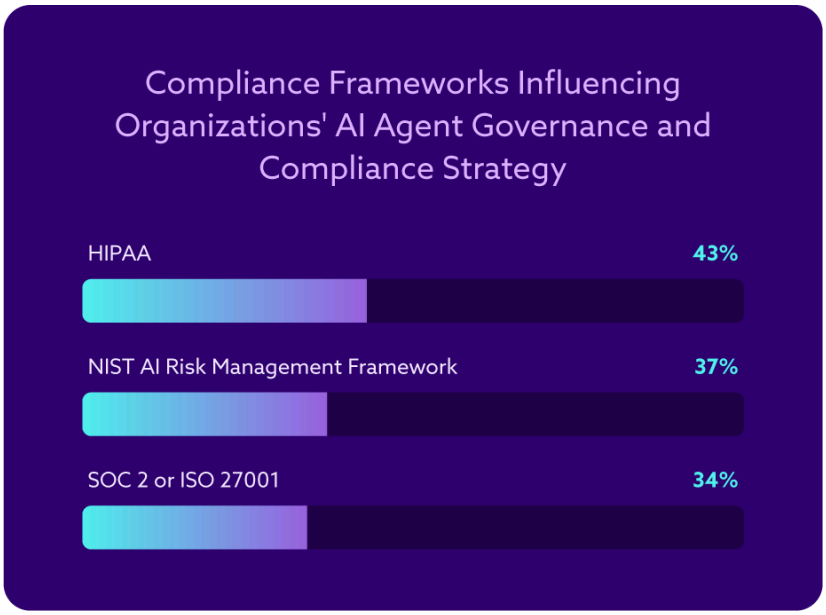
环境在代理行为偏离预期时，依赖人工调查和事后分析，而非主动或自动检测。对于安全团队而言，这可能意味着调查负担增加，以及大规模监控自主系统能力的下降。

这些模式共同表明，AI代理超出其范围是预期的运营现实，而非异常的故障模式。AI代理经常超出预期权限运行，事件屡见不鲜，而应对措施则发生在可见性、运行时控制和可追溯性仍不成熟的环境中。这种组合使安全团队倾向于被动调查和人工监督，这并非仅因对抗性活动，而是因为动态、自主系统正被采用更适合静态软件的控制和流程进行管理。数据表明，AI代理的失当行为并非边缘情况；它需要专门的行为边界和监督。

关键发现4：

缺乏AI代理安全策略，合规将成为默认选项

人工智能代理的正式治理在许多组织中存在，但通常不完善且应用不均。 **不到三分之一的组织（31%）报告已正式采用人工智能代理治理政策** 而另一半则表示政策仅部分记录或执行不一致，这代表了治理成熟度最常见的状态。另外12%的人报告说根本没有正式政策。这种分布表明，治理工作通常仍处于进行中，而非完全建立或实施到位。

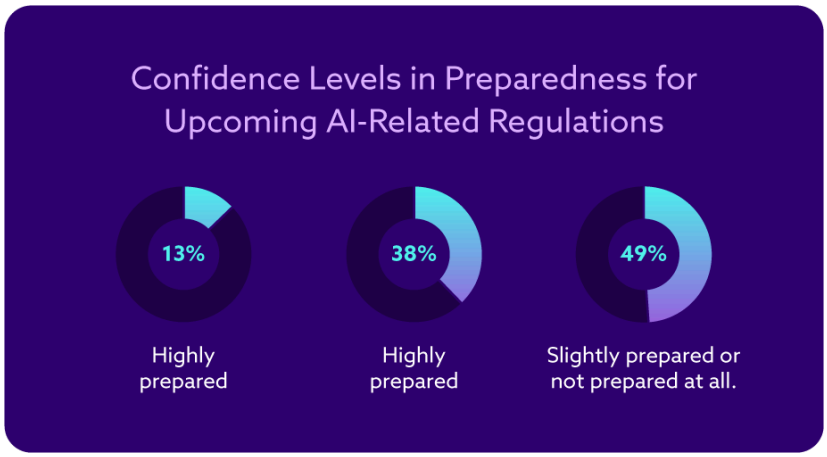


缺乏全面的内部治理框架，组织似乎严重依赖现有的监管与合规结构来指导人工智能代理的监督。 **当被问及哪些框架对人工智能代理治理影响最大时，受访者最常提及的是《健康保险流通与责任法案》（HIPAA，占43%），其次是《国家 institute of standards and technology人工智能风险管理框架》（NIST AI Risk Management Framework，占37%），以及SOC 2或ISO 27001（占34%）。** 医疗保健、隐私和通用安全框架的突出地位表明，治理方法

通常，这些规范会借鉴既有的合规体系，即便这些框架并非专门为自主系统或基于代理的系统而设计。这种依赖性进一步加剧了复杂性，因为针对代理的合规标准仍在发展中且存在碎片化，对于如何在生产环境中管理自主行为，它们提供的指导有限。

尽管依赖监管指导，但准备情况的信心仍然有限。 **只有13%的组织表示已为即将出台的AI相关法规做好了充分准备，而近半数则表示他们准备程度一般或完全没有准备（49%）** 监管之间的这种差距

影响和感知的准备状态表明，合规性的一致性并不必然转化为对实践中如何治理人工智能代理的信心或清晰度。

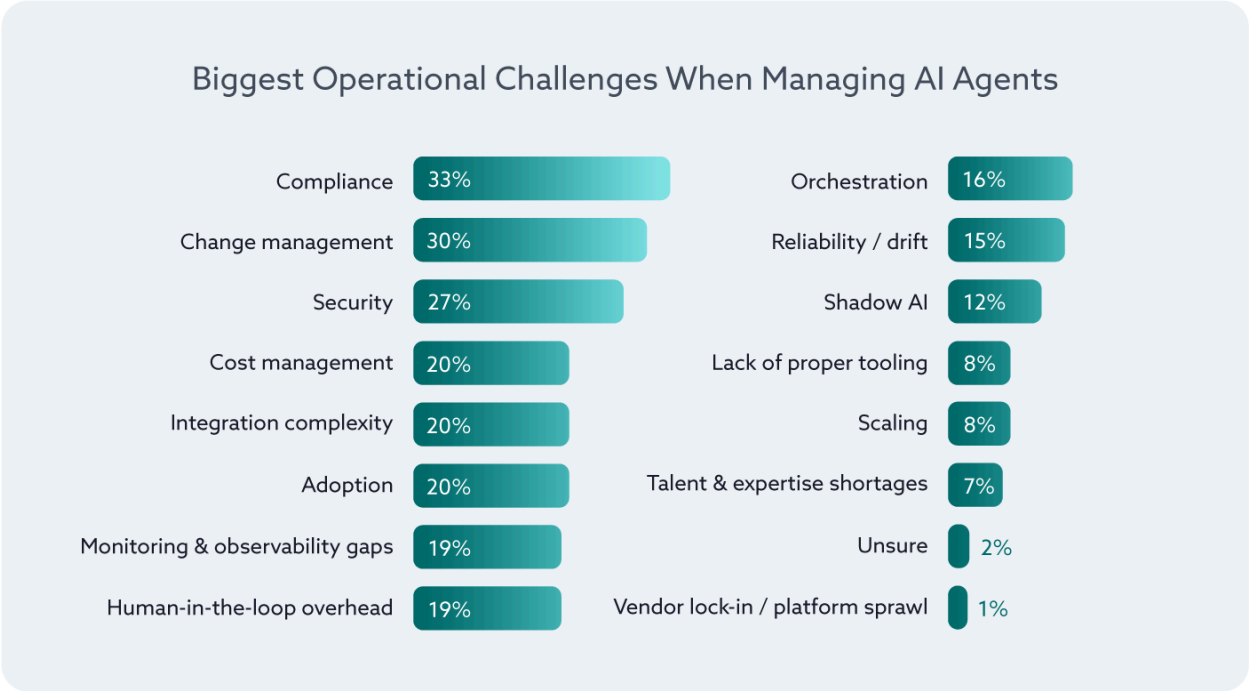


运营挑战强化了这一模式。合规性被列为与AI代理最相关的最常见的运营挑战，有33%的受访者选择了此项。

超过与安全相关的担忧（27%）。这种侧重表明，组织正投入巨大精力来解读和与监管预期保持一致，即便在治理结构、问责制和控制等更广泛的问题仍悬而未决的情况下。

总的来说，人工智能代理的治理似乎更多地受到外部监管压力的影响，而非内部定义的战略。政策部分采纳、强大的监管影响以及低信任度，

准备状态指向这样一种环境：合规性充当了凝聚治理的替代品。



尽管这种方法提供了一个起点，但它也反映了在如何将监管要求转化为适用于大规模运行的人工智能代理的组织范围治理模型方面持续存在的 Uncertainty。

结论

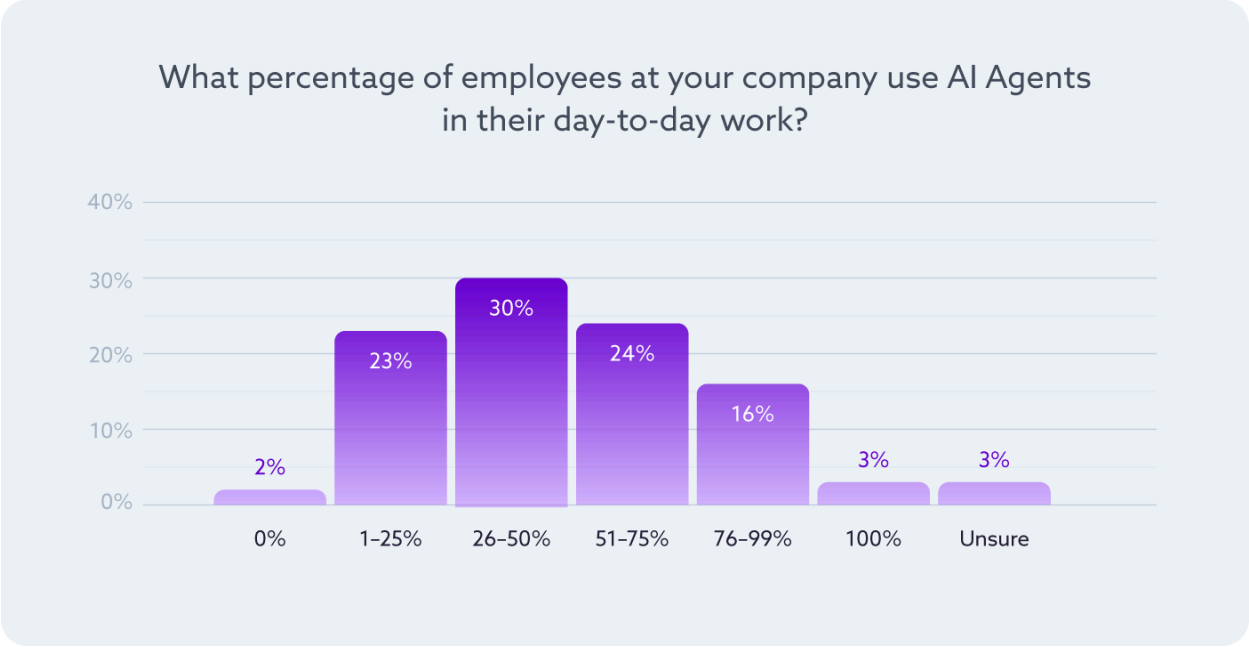
人工智能代理正日益嵌入各组织的日常工作中，跨越职能、平台和用例。其采用正以务实和快速的方式推进，动力源于即时的运营价值而非集中式设计。因此，人工智能代理正成为核心数字劳动力的一部分。

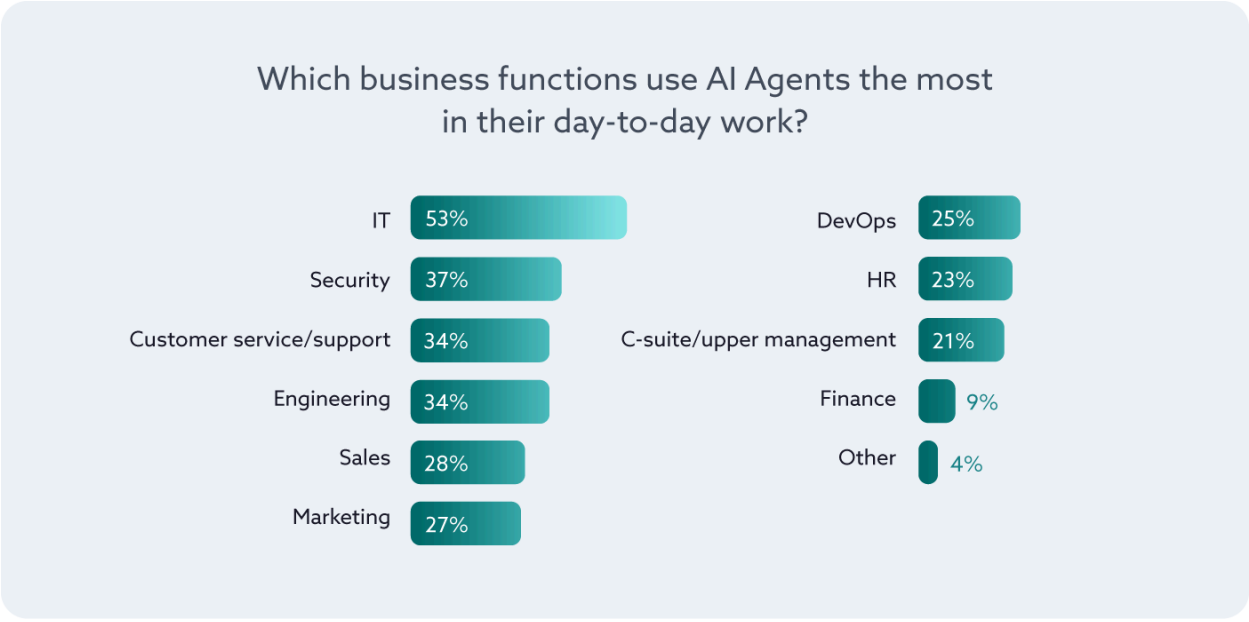
与此同时，调查数据显示部署与管控之间存在持续性的差距。未经授权的AI代理会过早出现，所有权和问责制往往不明确，代理行为常常超出预期边界，这表明安全策略仍在发展中。这些情况导致出现难以检测、处理缓慢的安全事件，增加了对人工调查和被动响应的依赖。为静态软件环境设计的工具和流程，往往不适用于自主运行且随时间变化的系统。

治理工作反映出运营现实与控制之间日益增长的张力。AI代理已被嵌入核心工作流程，但监督仍主要受现有监管和合规框架的影响，而非特定于代理的治理模型。因此，在可见性、所有权和行为控制方面的差距呈现出系统性特征，而非过渡性特征，即使采用程度成熟，这些差距依然存在。解决这些差距将需要这些要素协同发展，使治理和安全实践与AI代理的自主性和互联性相一致。虽然由代理驱动的系统扩张不可避免，但相应调整其控制模型的组织有机会以增强而非削弱安全性和弹性的方式来驾驭这一转变。

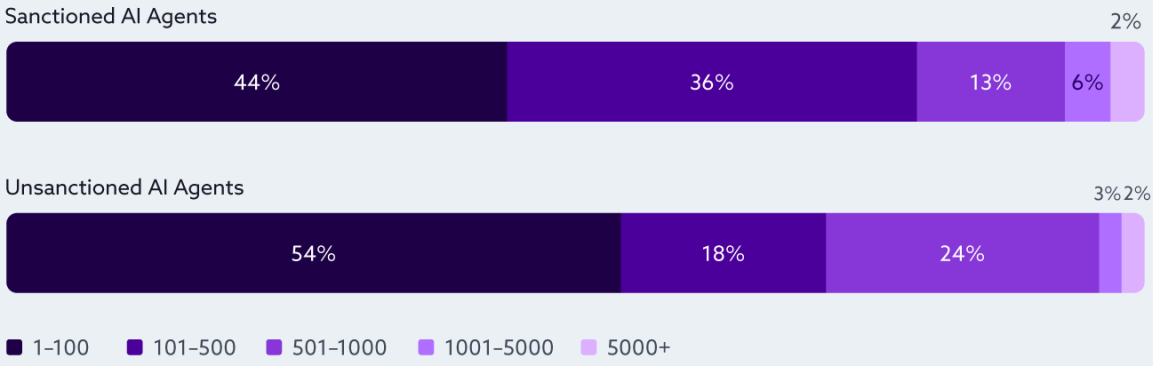
完整结果

收养与使用

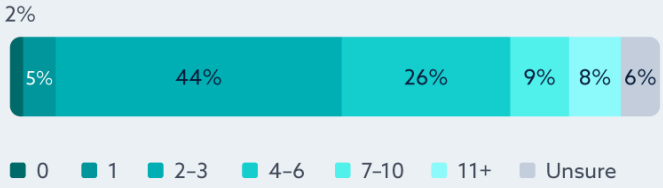


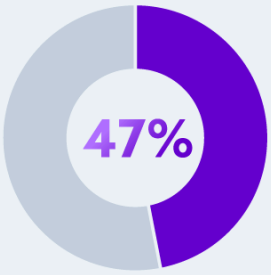


How many AI Agents does your organization currently have in production?



How many different AI Agent platforms are in use across your enterprise?

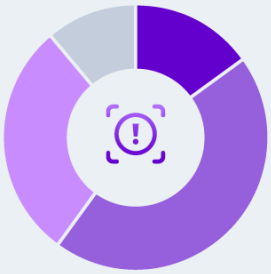




What percentage of your AI/security budget will be allocated to AI agent security specifically?

安全和风险

How confident are you in your ability to detect and respond to AI Agent-related incidents today?

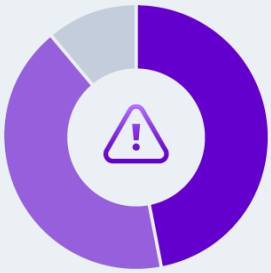


- 15% Highly confident
- 45% Moderately confident
- 29% Slightly confident
- 11% Not at all confident

What types of AI Agent-specific security risks concern you most?

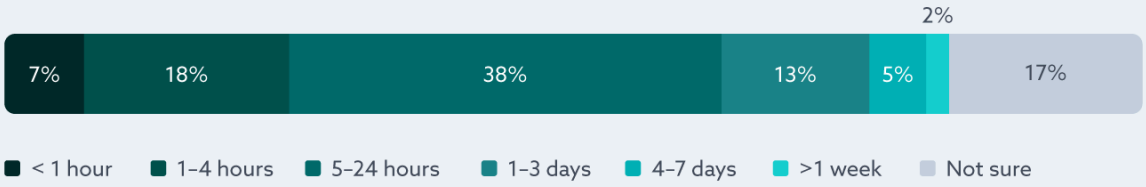
- 29% Prompt injection
- 28% Cost abuse / runaway calls
- 24% Secrets exposure
- 21% Jailbreaks / policy bypass
- 20% Data exfiltration
- 19% Over-privileged agents
- 18% Unsafe browsing/scraping tools
- 17% MCP server vulnerabilities
- 16% Unsafe tool use / function-calling
- 12% Multi-agent risks
- 11% Insecure memory
- 8% Workflow/orchestrator vulnerabilities
- 7% Unsafe code execution
- 7% Impersonation / spoofed agents
- 7% Insufficient monitoring & anomaly detection
- 6% RAG poisoning
- 6% Lateral movement
- 6% Inadequate human-in-the-loop
- 5% Supplychain risks
- 2% Unsure
- 1% Other

Have you experienced a security incident involving an AI Agent in the past 12 months?

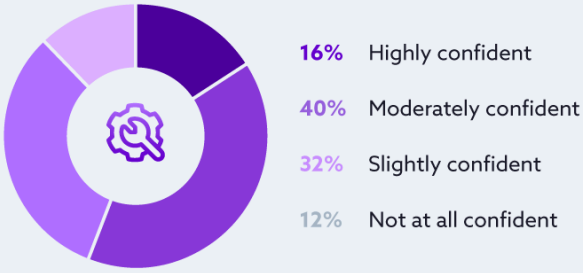


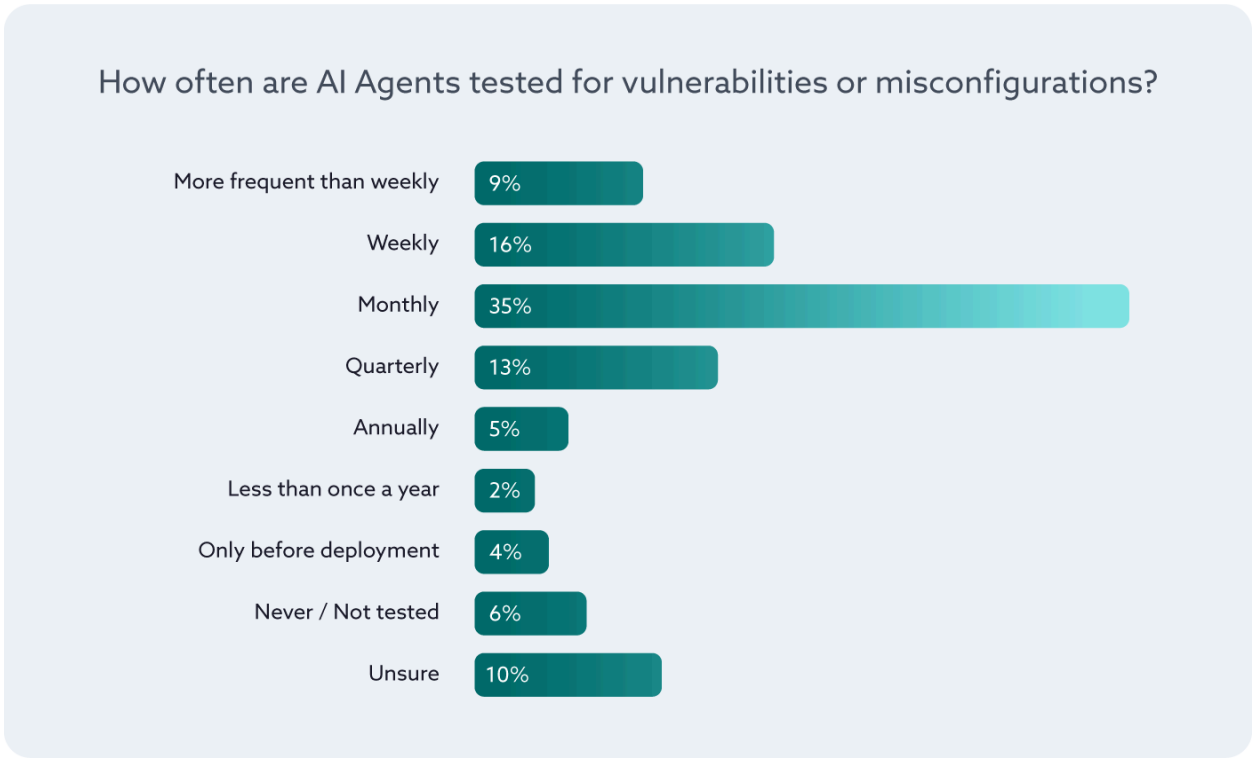
- 47% Yes
- 42% No
- 11% Not sure

On average, how long does it take to detect and respond to a security incident involving a compromised AI Agent (from initial compromise to full remediation)?

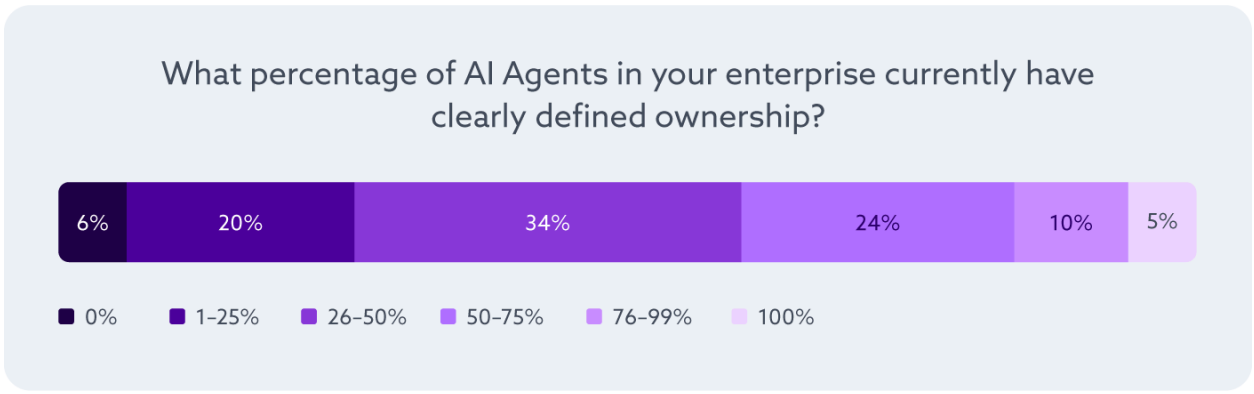


How confident are you that your organization's current security tools can detect AI Agent-specific threats?

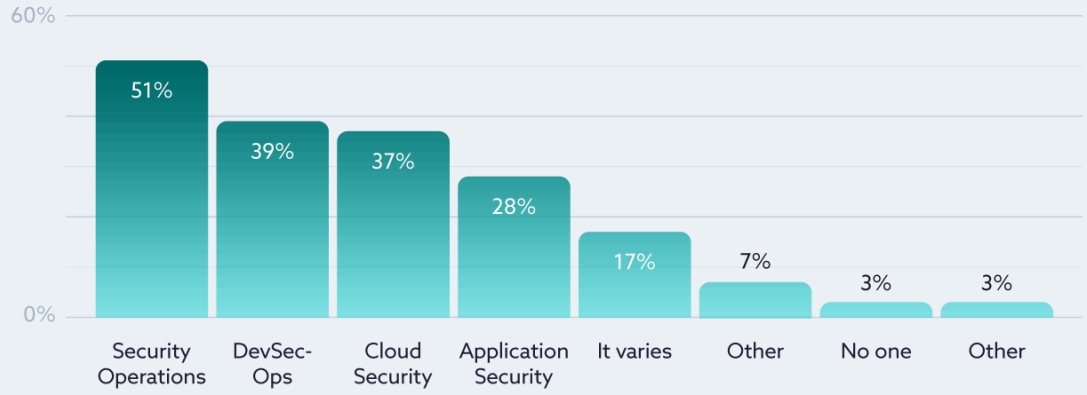




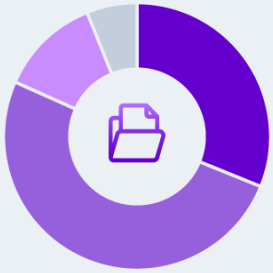
治理与合规



Who is responsible for AI Agent security in your organization?

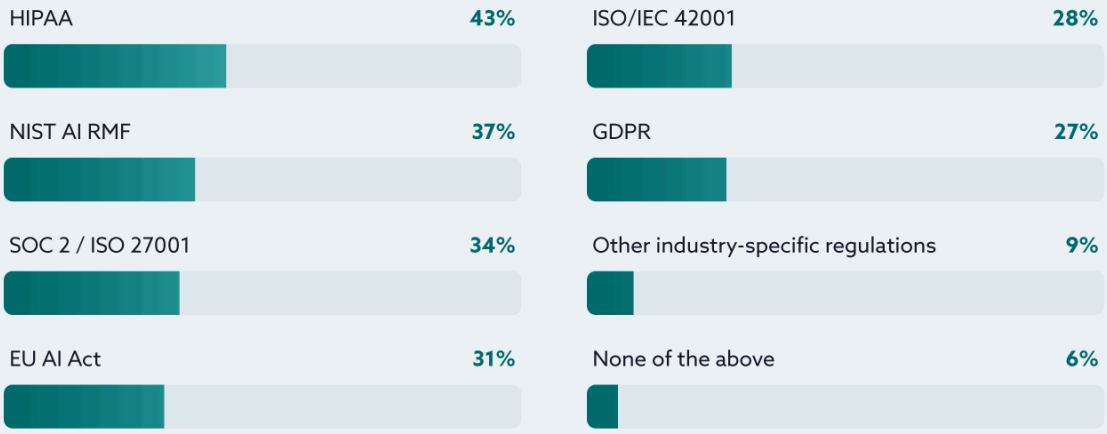


Do you have formal governance policies for AI Agent usage?

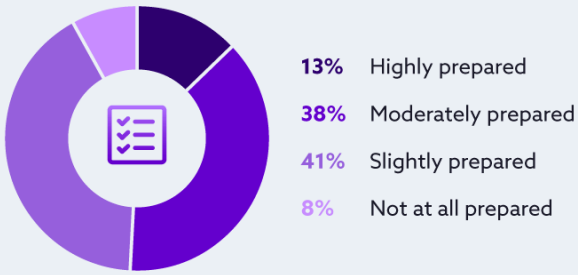


- **31%** Yes — documented and fully adopted
- **50%** Somewhat — partially documented or inconsistent across teams
- **12%** No — there are no formal policies in place
- **6%** Not sure

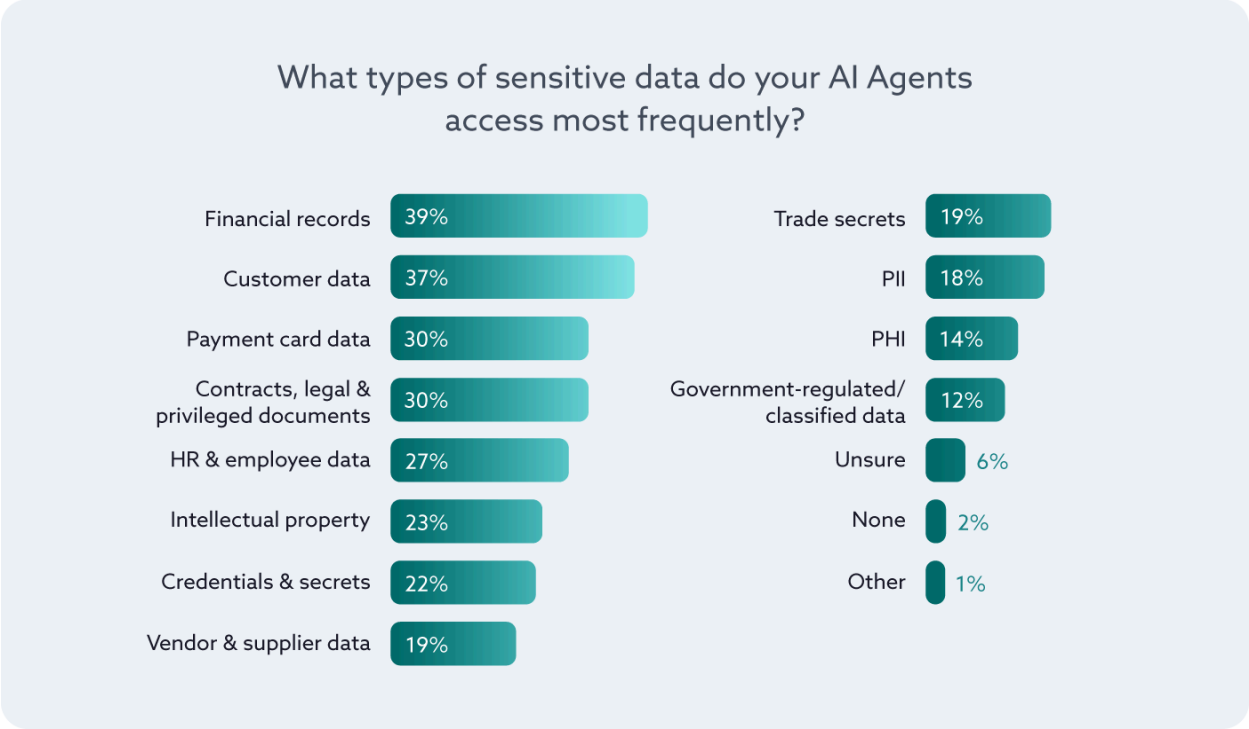
Which compliance frameworks most influence your AI agent governance and compliance strategy? (select all that apply)

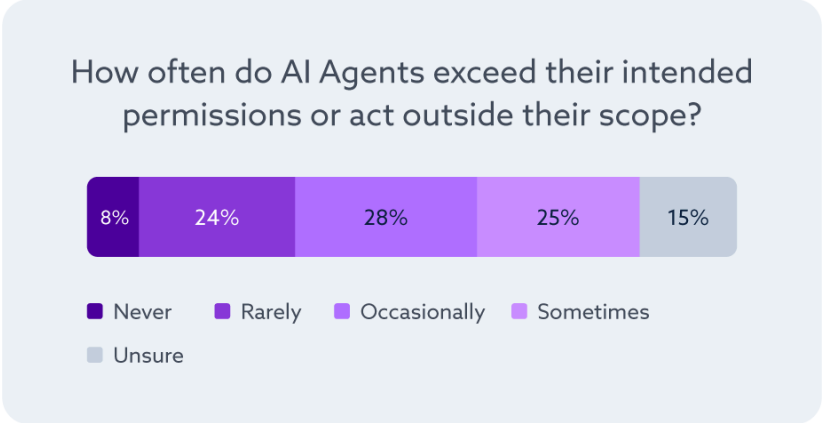


How prepared do you feel for upcoming AI Agent-specific regulations in your industry?

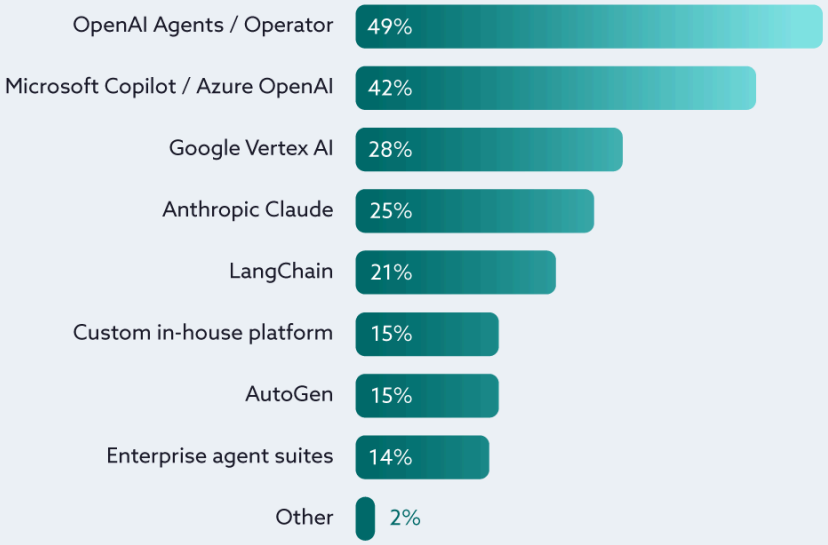


建筑

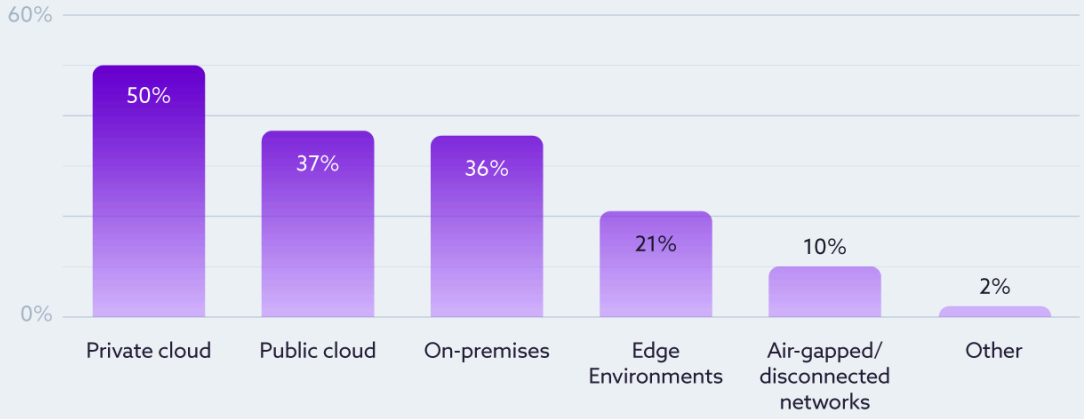




What are your organizations most used Agentic Platforms?

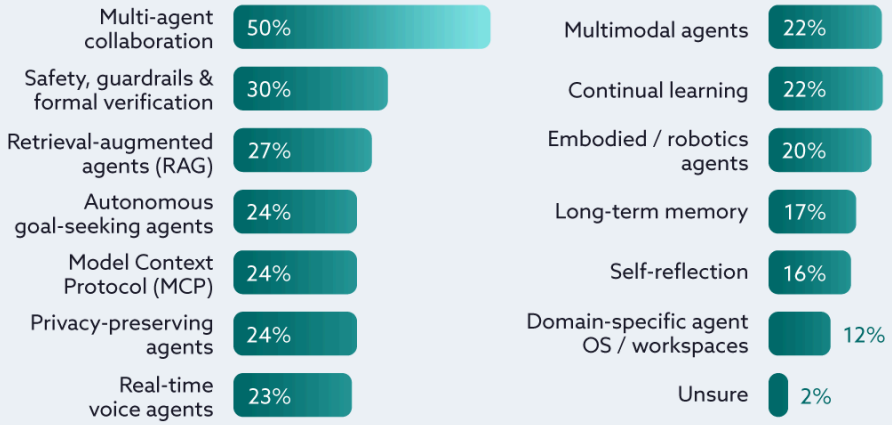


Where are AI Agents typically deployed in your organization?

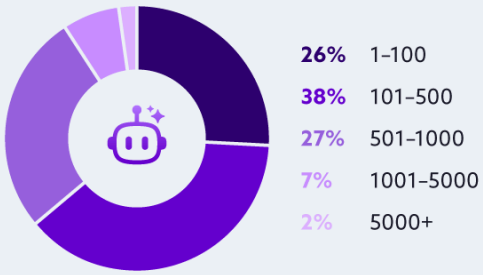


未来展望

Which emerging AI Agent technologies are you most interested in?



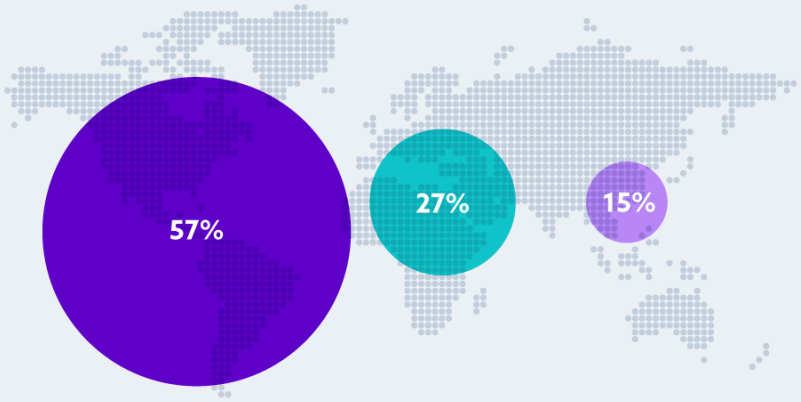
How many AI agents do you expect your organization will have live by the end of 2026?



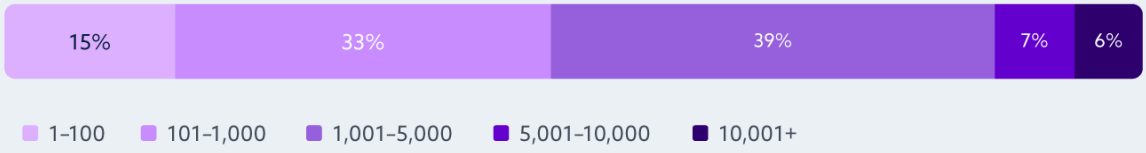
人口统计

Location

- Americas
- Europe, Middle East, Africa (EMEA)
- Asia Pacific (APAC)



Organization Size



1–100 101–1,000 1,001–5,000 5,001–10,000 10,001+

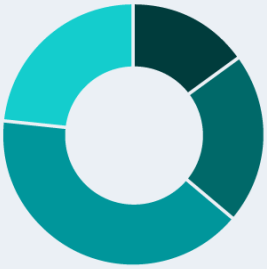
Organization Industry

- **27%** Telecommunications, Technology, Internet & Electronics
- **11%** Education
- **10%** Finance & Financial Services
- **10%** Entertainment & Leisure
- **7%** Business Support & Logistics
- **6%** Construction, Machinery, and Homes
- **4%** Healthcare & Pharmaceuticals
- **4%** Manufacturing
- **3%** Automotive
- **2%** Government
- **2%** Insurance
- **2%** Prefer not to answer
- **2%** Advertising & Marketing
- **2%** Utilities, Energy, and Extraction
- **2%** Retail & Consumer Durables
- **1%** Agriculture
- **1%** Airlines & Aerospace (including Defense)
- **1%** Food & Beverages
- **1%** Health & Fitness
- **1%** Real Estate
- **1%** Transportation & Delivery
- **1%** Nonprofit

Primary Functional Area of Role

- **22%** Security
- **9%** IAM
- **41%** IT Operations
- **11%** DevOps / Platform engineering
- **4%** Data / ML
- **7%** Business Unit
- **6%** Executive leadership

Job Level



- **15%** C-level or Executive
- **21%** Director
- **40%** Manager
- **23%** Staff

调查方法学

云安全联盟（CSA）是一个非营利组织，其使命是广泛推广最佳实践，确保云计算和IT技术中的网络安全。CSA还教育这些行业内的各种利益相关者关于所有其他形式计算中的安全问题。CSA的成员是由行业从业者、企业和专业协会组成的广泛联盟。CSA的主要目标之一是进行调查，以评估信息安全趋势。这些调查提供了关于组织当前成熟度、观点、兴趣和关于信息安全与技术的意图的信息。

禅提（Zenity）委托CSA开发一项调查和报告，以更好地了解行业对自主AI代理的认知、态度和观点。禅提资助了该项目，并与CSA研究分析师共同开发了问卷。该调查由CSA于2025年9月和11月在线进行，收到了来自不同规模和地点的IT和安全专业人士的445份回复。CSA的研究分析师为该报告执行了数据分析和解读。

研究目标

本研究旨在探讨AI代理在企业环境中如何被采纳、管理和保障。其目标是明确了解随着AI代理从实验阶段转向核心运营，当前的使用模式、新兴风险和组织准备情况。

具体而言，该调查旨在：

- 衡量AI代理人在各职能、技术平台和劳动力队伍中的采用与使用情况
- 评估与人工智能代理行为、事件、检测和响应相关的安全态势
- 了解治理与合规方法，包括所有权、政策成熟度和监管影响
- 探索与AI代理的可视性、控制权及日常管理相关的运营挑战
- 在组织规划扩大人工智能代理使用时，规划未来的投资和创新优先事项