

2025-2026 AI Agent 行业深度分析报告

术语与分析前提：AI Agent 与 Agentic Workflow 的边界

行业对“Agent/Agentic AI”的叫法非常混乱：既有人把“多步提示词”称为 Agent，也有人把“能调用工具”就算 Agent。若不先统一定义，后续的市场分层、产品比较与成熟度评估都会失去可比性。Gartner 在其 2025 年十大战略技术趋势中给出的“*agentic AI*”核心特征是：系统能够自主规划并采取行动以达成用户定义目标，并被视作推动“虚拟劳动力（virtual workforce）”的关键路径。¹ McKinsey 在 2025 年全球调研报告中则把 AI agents 描述为：基于基础模型、能够在真实世界中行动，并能在工作流中规划并执行多步的系统。² Microsoft Copilot Studio 的官方指导也强调：自主 Agent 能感知事件、做出决策并独立执行任务，通过触发器、指令与护栏运行，并在企业场景中受“权限范围与可审计流程”约束。³ AWS Bedrock Agents 的定义同样指向一个“编排基础模型、数据源与软件应用，并自动调用 API/知识库去完成动作”的自治系统。⁴

基于上述一手材料，本报告采用一个“功能性、可操作”的行业通用定义，用于后续分层与对比：

本报告中的 AI Agent 定义（Functional Definition）

一个系统只有在同时满足以下条件时，才被归类为“AI Agent”：
- **持续目标**：目标不是一次性文本输出，而是可持续推进的任务状态（例如“完成报销”“跑通测试”“提交表单”）。⁵
- **自主决策权**：系统能在运行中决定“下一步做什么”，而非仅按预设步骤执行。⁶
- **可执行行动**：能对外部环境（网页/桌面软件/API/企业系统）执行动作，而不仅是生成建议。⁷
- **反馈闭环**：具备“观察→决策→行动→评估→再决策”的循环，并能根据失败调整策略或请求信息补全。⁸
- **受控犯错与治理**：系统在设计上允许在边界内失败（重试/回滚/人工接管/拒绝策略），并通过权限、审计、确认等机制保证可控。⁹

在此定义之下，“**Agentic Workflow**”被单独归类：它通常具备多步、工具调用与一定的自动化，但流程推进权更偏向外部编排（系统/规则/人），而非由 Agent 自主决定。Microsoft 将“生成式编排（generative orchestration）”描述为：系统会基于工具/主题/知识源元数据构建计划、串联调用，并可响应事件触发；这非常接近“*agentic workflow*（系统主导的多步编排）”。¹⁰ AWS 也在 Bedrock Agents 中明确区分“默认 ReAct 式编排”与“自定义 orchestrator”，后者本质是把编排权交给你写的状态机（Lambda）来决定每步调用。¹¹

一个实务上好用的边界判别法是：

如果去掉外部编排器/步骤脚本后，系统仍能围绕目标自主推进并自我纠错，它更像 Agent；如果必须依赖“预定 step 顺序”才能跑通，它更像 agentic workflow。¹²

市场现状与趋势：2025-2026 从概念热潮走向平台化治理

从数据看，2025-2026 年的 AI Agent 市场处于“高试点率、低规模化、平台化加速”的阶段。McKinsey 2025 年全球调研显示：62% 的受访者所在组织“至少在尝试 AI agents”，但只有 23% 报告其组织正在“在至少一个职能中扩展部署/采纳 agentic AI 系统（scaling）”，且多数只在 1-2 个职能落地。² McKinsey 2026 年初进一步指出：AI agents 在“规模化采用”上最领先的行业集中于科技行业，职能上以软件工程与 IT 的“scaled use”更突出。¹³

与此同时，行业开始显著“去泡沫化”。Gartner 在 2025 年 6 月的新闻稿中警告：到 2027 年末，超过 40% 的 **agentic AI** 项目将被取消（主要原因是成本上升、业务价值不清晰或风险控制不足），并指出大量供应商在

进行“agent washing”（把传统助手/RPA/聊天机器人换皮成 agent），Gartner 估计“成千上万”所谓 agentic AI 供应商中，真正具备实质 agent 能力的约 130 家。¹⁴

从企业管理视角看，BCG 基于其与 MIT Sloan 的研究指出：agentic AI 既像软件又像同事，企业必须重新设计工作流、治理与决策权才能释放价值；其文中引用的调研结果显示 35% 的企业已开始使用 agentic AI，另有 44% 计划近期采用。¹⁵ 这与 McKinsey 的“高探索、低规模化”并不矛盾：探索很普遍，但达到可复制、可审计、可控的规模化运行仍然困难。

从产品形态上，2025–2026 的竞争主线可以归纳为四条“形态赛道”，并呈现从“能力突破”到“治理平台”的演化：

- **Browser/Computer-use Agents**: 把“GUI 作为通用行动接口”，在无 API 场景中也能自动化（OpenAI Operator/ChatGPT agent、Anthropic computer use、Copilot Studio computer use）。¹⁶
- **Enterprise Agent Platform**: 围绕数据连接、权限、审计、可观测性与生态互操作，构建可规模化的“企业智能体平台”（Salesforce Agentforce 3/360、Microsoft Copilot Studio、Google Agentspace→Gemini Enterprise）。¹⁷
- **Engineering Agents**: 围绕“计划-执行-测试-PR”的闭环，把软件工程任务变成可自动推进的状态目标（Devin）。¹⁸
- **Cloud Orchestrated Agents**: 以云厂商提供的“托管编排运行时”承接企业系统动作面（action groups/knowledge bases/guardrails/custom orchestrator），降低落地门槛（Amazon Bedrock Agents）。¹⁹

在互操作标准上，2024–2025 出现两个关键协议，正在把“单体 agent”推向“多 agent 生态”：

Anthropic 在 2024-11 发布并开源 **MCP (Model Context Protocol)**，试图用统一标准把 AI 应用连接到数据源与工具，降低 N×M 集成复杂度。²⁰ 2025-12，Anthropic 宣布把 MCP 捐赠给 Linux Foundation 体系下的 Agentic AI Foundation，并披露生态上已有“超过 10,000 个活跃的公共 MCP servers”，且 MCP 已被 ChatGPT、Gemini、Microsoft Copilot 等采用。²¹ Google 在 2025-04 发布 **A2A (Agent2Agent)**，强调让不同厂商/框架的 agent 进行安全协作，明确其与 MCP 互补（MCP 提供工具与上下文，A2A 提供 agent-to-agent 协作）。²² OpenAI 也在开发者文档中提供对远程 MCP servers 的连接与工具导入机制，并强调对 MCP 工具调用默认启用审批与日志建议，以应对数据共享与提示注入风险。²³

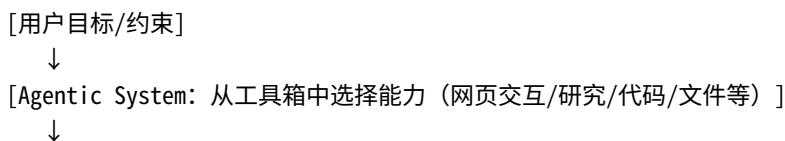
典型产品案例：运行架构、核心能力与 Agent 属性拆解

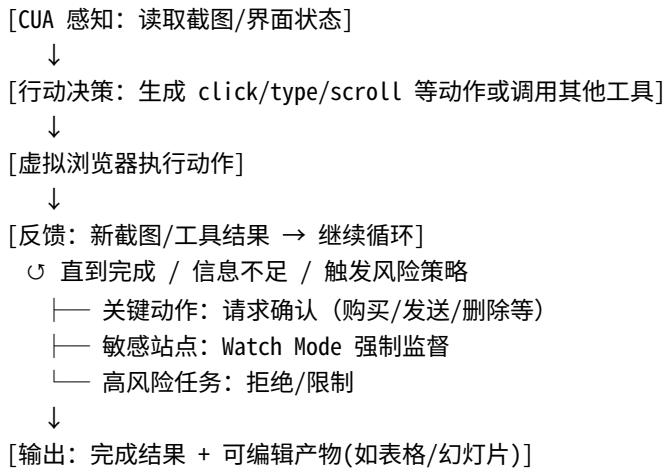
本节按“运行架构（文字流程图）→核心能力→关键治理点→Agent 属性判定”拆解七个代表性产品。为保持可比性，所有判定均基于前述功能性定义。

OpenAI Operator / ChatGPT agent (Browser Agent)

定位是“能在网页上完成任务的通用行动型 agent”。Operator 在 2025-07-17 更新中被整合为 ChatGPT 的“agent mode”，并逐步替代独立 Operator 站点。²⁴ ChatGPT agent 官方描述强调：系统将 Operator 的网页交互、deep research 的信息综合与 ChatGPT 的对话能力合并为“统一 agentic system”，并在执行关键动作前请求许可、支持随时中断/接管。²⁵

运行流程图（文字化）：





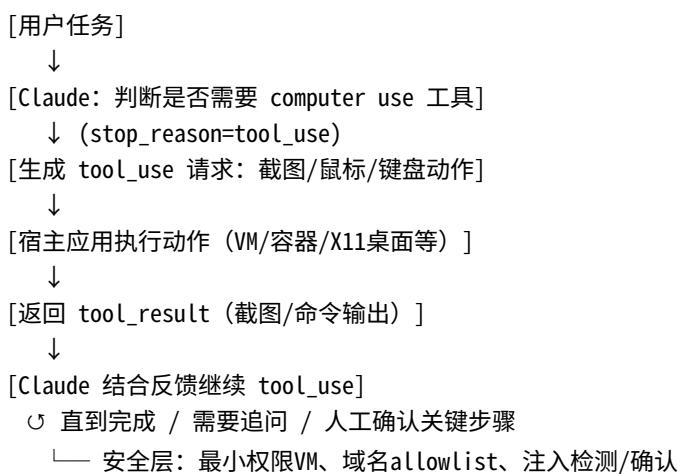
支撑证据包括：Operator 以 CUA (Computer-Using Agent) 驱动，通过截图与鼠标键盘动作完成网页任务，并在登录/支付/CAPTCHA 等场景要求用户接管；此外系统卡强调通过拒绝高风险任务、关键动作确认与监控来减轻提示注入等风险。²⁶ OpenAI 的 API “computer use” 文档进一步将其描述为“连续循环：发送动作→执行→回传截图→模型决定下一步”，并明确在高风险/完全认证环境中不建议盲目信任。²⁷

Agent 属性判定：满足持续目标（完成任务）、自主决策（下一步动作）、可执行行动（浏览器操作）、反馈闭环与受控犯错（确认/Watch Mode/接管/拒绝），因此属于 **L3 级自治 Agent**（见后文成熟度模型）。²⁸

Anthropic Computer Use (Computer-use Capability + Reference Loop)

Anthropic 将 computer use 定位为“下一代通用行动接口”，让模型像人一样操作任意软件界面。其研究说明强调：Claude 在适当软件设置下可移动光标、点击与输入，并能把用户文本提示转化为步骤序列，遇到障碍会自我纠错重试。²⁹ API 文档进一步给出“agent loop”的工程实现：模型输出 tool_use→应用执行→返回 tool_result→直到任务完成；并强调需要沙箱环境、最小权限、域名 allowlist、关键决策人工确认等安全措施，同时提供对提示注入的分类器与“检测到疑似注入时要求确认”的额外防线。³⁰

运行流程图（文字化）：

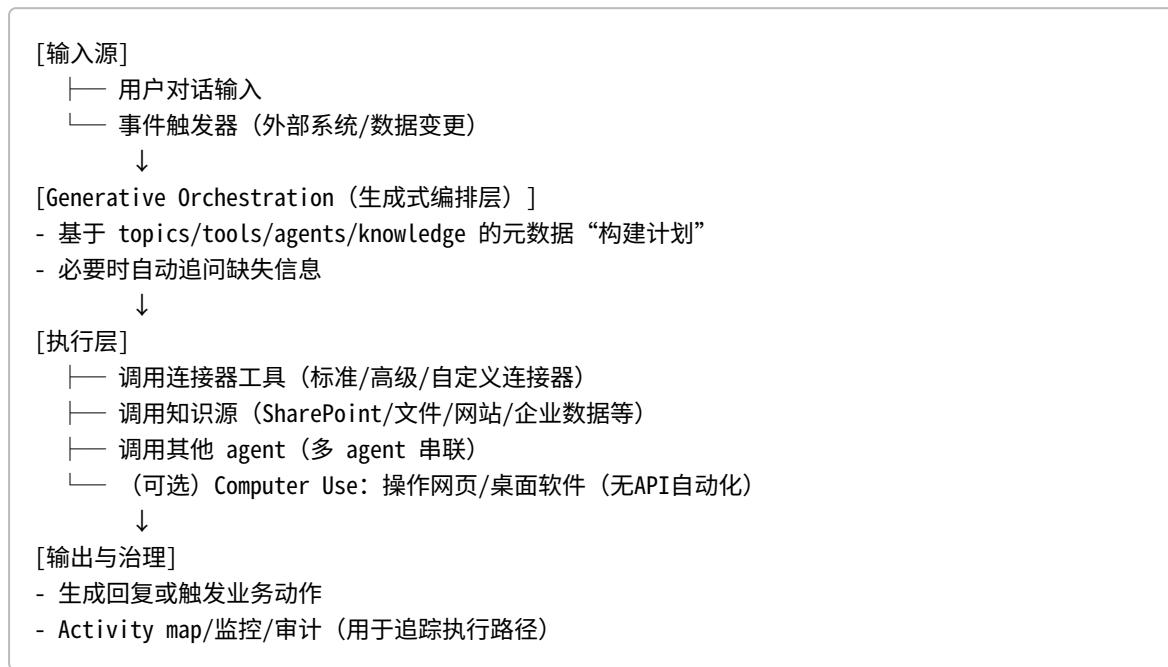


Agent 属性判定：Computer Use 本身更像“Agent 的行动接口与循环范式”，自治程度取决于你把决策权交给模型的程度。官方 reference implementation 已展示完整闭环与安全策略，因此它可支撑 L2-L3：若仅作为工具被外部编排调用更像 agentic workflow；若允许模型自主推进任务闭环则可构成自治 agent。³¹

Microsoft Copilot Studio (Enterprise Agent Platform + Generative Orchestration + Computer Use)

Copilot Studio 的关键在于“企业级可控自治”。Microsoft Learn 明确写到：Copilot Studio 的 autonomous agents 能在不等待用户提示的情况下，通过触发器、指令与护栏在后台持续运行；同时强调每个 agent 在企业场景中应处于“权限范围、决策边界与可审计流程”内。³²

运行流程图（文字化）：



32

在行动面上，Copilot Studio 支持大量连接器（含标准/高级/自定义 API 连接器），并可以把连接器动作作为工具在 agent 或 topic 中调用。³³ 在无 API 自动化上，Microsoft 在 2025-09 宣布 computer use 公共预览：允许 agent “使用自己的电脑”执行数据提取、填表等任务，并提供 allow-list、凭据存储、托管浏览器（Windows 365）等增强；这使其从“对话式助手”拓展到“能操作软件”的行动型 agent。³⁴ Microsoft Learn 进一步解释 computer use 可在“托管浏览器/Cloud PC 池/自带机器”三种模式运行，其中 Cloud PC 池可与 Entra/Intune 结合实现更强治理。³⁵

在治理层，Copilot Studio 的安全与治理概念页列举了数据策略控制（可治理连接器/技能/触发器等）、Purview 审计、Sentinel 监控告警、DLP、CMK 等能力。³⁶

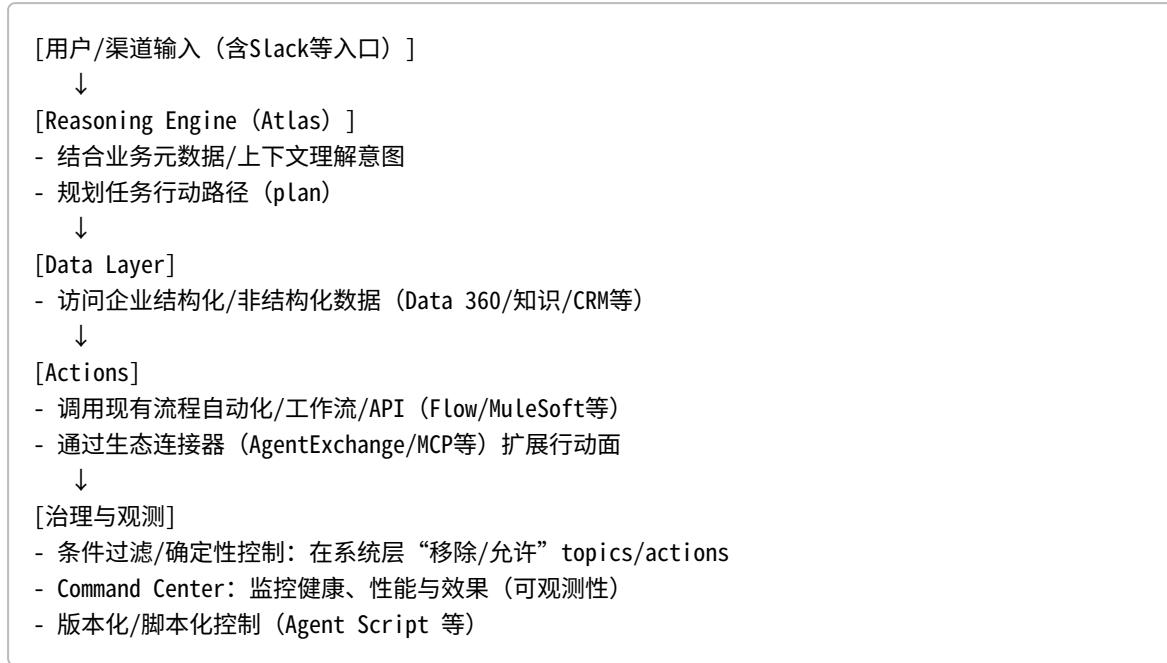
Agent 属性判定：Copilot Studio 在“生成式编排”层面既可表现为 agentic workflow（系统用元数据构建计划并串联执行），当结合事件触发、后台运行与 computer use 行动面后，已满足自治 agent 的核心特征，并通过企业治理把“自主”限制在可审计边界内，因此整体更接近 L3→L4 的企业平台型 Agent 系统。³⁷

Salesforce Agentforce (Enterprise Agent Platform, 强调可观测性与确定性控制)

Salesforce 将 Agentforce 的基本构成归纳为“data + reasoning + actions”，并宣称 agent 可在护栏内自治检索数据、构建行动计划并执行。³⁸ 更具“平台化”特征的是其 2025-06 发布 Agentforce 3：强调解决规模化

的最大阻碍是“看不见 agent 在做什么、难以快速演进”，因此引入 Command Center 作为可观测性层，并宣称原生支持 MCP、扩展 AgentExchange 生态、提升 Atlas 架构性能与信任。³⁹

运行流程图（文字化）：



⁴⁰

在控制策略上，Salesforce 的 Agentforce Guide 明确提出“Conditional Filtering”：它不是用提示词影响 LLM，而是在系统层把不该出现的 topic/action 直接移除，从而降低语义噪声、提高分类准确性与安全性。

⁴¹ 其“Five levels of determinism”文档也明确讨论“自治的流动性与企业确定性约束”的张力，强调要在可控治理下实现可靠 agent 行为。⁴²

落地与 ROI 方面，Agentforce 3 的新闻稿给出若干客户案例口径（例如平均处理时长下降、税务高峰期自治解决大量对话等），并明确将“可观测性、互操作与预置行业动作”作为“快速实现价值”的关键。⁴³ 与此同时，Reuters 报道 Salesforce 在 2025-10 宣布加深与 OpenAI、Anthropic 合作以增强 Agentforce 360 平台，指向企业平台正在走向“多模型、多生态”的集成路线。⁴⁴

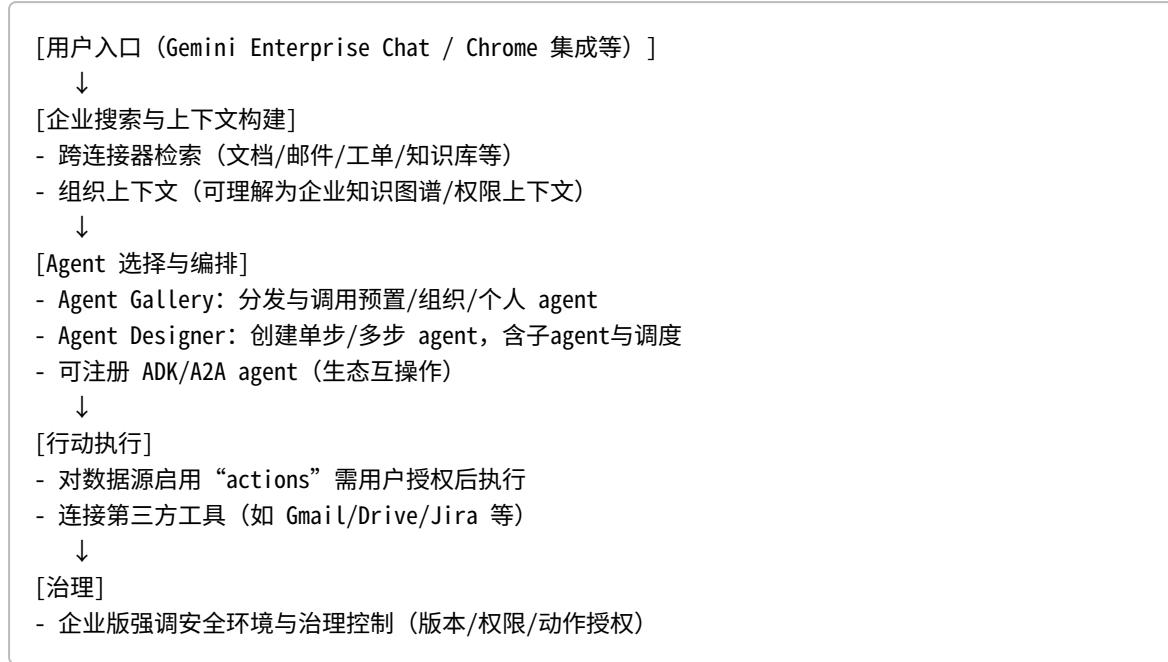
Agent 属性判定：Agentforce 在自治、行动面、平台治理（可观测性/确定性控制/互操作协议）上均具备成熟体系，属于 L4 级企业规模化 Agent 平台。⁴⁵

Google Agentspace → Gemini Enterprise (Enterprise Agentic Platform + Agent Designer/Gallery + A2A)

Google 2024-12 发布 Agentspace，核心卖点是把 Gemini 推理、Google 搜索能力与企业数据连接起来，支持跨 Confluence、Drive、Jira、SharePoint、ServiceNow 等连接器，并提供角色/权限控制（RBAC、VPC Service Controls、IAM 等）。⁴⁶ 2025-04 的更新强调 Agent Gallery（发现与分发 agent）、Agent Designer（无代码创建 agent）、以及推出 Deep Research 等专家 agent，并支持开放的 A2A 协议实现跨生态的多 agent 协作。⁴⁷

2025-10-09 起，Google 官方将“Agentspace”并入并更名为 **Gemini Enterprise**；Google Cloud 文档的 release notes 明确写到：控制台与默认应用中的 Agentspace 体验现称 Gemini Enterprise，且相关资源与文档一并迁移。⁴⁸

运行流程图（文字化）：



49

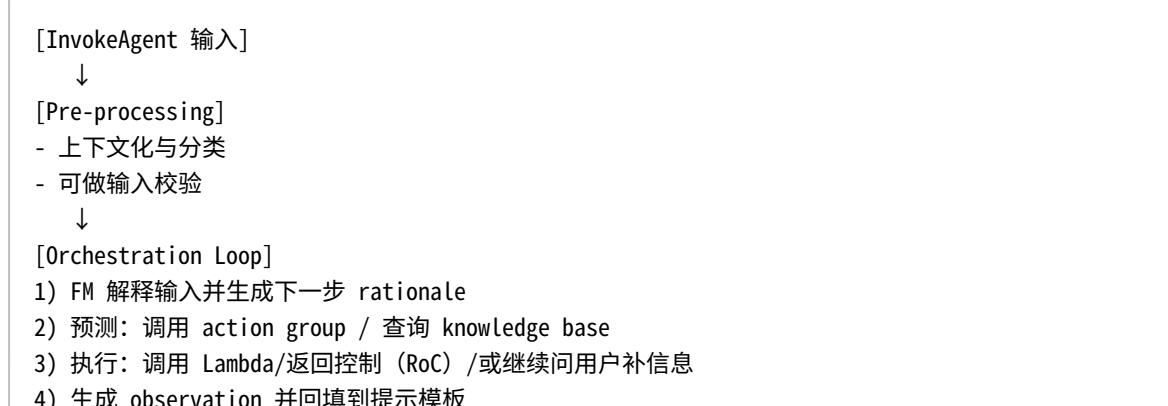
Google 的 Agent Designer 文档对能力边界写得比较“工程化”：它是 no-code/low-code 平台，可创建与管理单步、多步 agent，支持多 agent (subagents)、连接数据源与工具，并可“定时调度执行”。⁵⁰ Release notes 还显示 Gemini Enterprise 支持注册 A2A agent，并对“assistant actions”引入显式授权按钮，体现对行动权限的收口。⁵¹

Agent 属性判定：在“企业搜索+agent 分发+多步编排+动作授权+互操作协议 (A2A)”组合下，Gemini Enterprise 属于 **L4 倾向的平台型 Agent 体系**（尤其在多 agent、治理、分发与动作授权方面）。⁵²

Amazon Bedrock Agents (Cloud Orchestration Agents, 强调可配置编排与工具执行)

Bedrock Agents 的价值在于“托管运行时 + 企业集成”。官方文档将其运行过程拆成 pre-processing 与 orchestration 等环节，并明确 orchestration 是一个循环：基础模型生成 rationale、预测要调用的 action group 或知识库、执行并生成 observation，再决定是否继续循环或向用户追问。⁵³ Bedrock Agents 同时以 action groups 把“可执行动作”抽象成一组 API/函数调用入口。⁵⁴

运行流程图（文字化）：



- 5) 判断是否继续循环
 ⚡ 直到完成或需要用户补充信息
 ↓
 [输出]

55

更关键的是，AWS 在 2024-11 引入 **custom orchestrator feature**：允许用户用 Lambda 实现自定义状态机来控制规划、完成与验证，并对比指出默认 ReAct 的顺序反思可能引入延迟；custom orchestrator 通过“契约交互+状态转换”把编排权交还给企业，以更精确地在速度/准确/韧性之间取舍。¹¹ 这也从侧面解释了“Agent vs agentic workflow”的连续谱：Bedrock 提供默认 agentic loop，但也允许你将其收紧为更确定的工作流。

Agent 属性判定：默认 Bedrock Agents 已满足多步决策、工具行动与反馈循环，可归为 **L3 云端自治 Agent** 运行时；当使用 custom orchestrator 强化步骤控制时，会向 agentic workflow 更靠近，但治理与可复制性更强。¹⁹

Devin (Engineering Agent, 强调工程闭环)

Devin 的核心是把软件工程任务变成“状态目标”（例如“修复 issue 并通过测试”），并在环境内自主推进。官方介绍与文档强调：Devin 是自治 AI 软件工程师，能写、跑、测、重构、修 bug、写单测等。¹⁸ Cognition 的早期介绍还给出 SWE-bench 端到端解决率等基准，并强调 Devin 会自行搭建环境、复现 bug、编码并测试修复。⁵⁶

运行流程图（文字化）：

- [Ticket/Issue 输入（例如 Jira/GitHub）]
 ↓
 [Plan: 生成方案与修改计划 → 人类快速审阅]
 ↓
 [Act: 在代码库中实现修改]
 ↓
 [Test: 运行测试/CI/本地复现]
 ↓
 [Fix Loop: 读失败日志 → 调整策略 → 重跑测试]
 ⚡ 直到通过或达到阈值/人工介入
 ↓
 [PR: 生成可审阅变更并提交]

57

从产业动向看，Reuters 在 2025-07 报道 Cognition 收购 Windsurf (IDE 平台)，并提及该收购将帮助把 Windsurf 的能力整合进 Devin，反映“工程 agent + IDE 基础设施”正在走向更深的产品化与企业化。⁵⁸

Agent 属性判定：Devin 具有明确持续目标（跑通/交付）、强反馈闭环（测试驱动）、环境行动能力与一定自主决策权，因此属于 **L3 工程自治 Agent**；其企业治理能力更多取决于团队的审批/权限/审计接入程度，而非单纯模型能力。⁵⁹

能力成熟度模型：L0-L4 分层与产品映射

结合市场实践与上文的“Agent vs workflow”边界，本报告建议采用一个五级成熟度模型，既能解释当下产品差异，也能解释“为什么很多所谓 agent 落不了地”。

L0 对话式助手

以生成文本/建议为主，不执行外部动作；可有短期上下文，但缺乏可执行闭环。该层不在本报告核心列表中，但它是很多“被 agent 洗牌的产品”的真实底座。¹⁴

L1 工具调用增强

具备函数/工具调用，但通常由外部明确触发或单步调用为主；行动面有限，多用于检索、写入、简单 API 交互。OpenAI computer use 文档将“动作+截图回传”描述为一个循环工具链，可视为 L1-L2 的基础构件。

²⁷

L2 Agentic Workflow

多步执行、可串联工具/检索/写作，但“流程推进权”主要在系统侧（编排器/规则/人类）。Microsoft 的 generative orchestration 明确“基于元数据建计划、串联调用、必要时追问”，典型属于 L2（可升级到 L3 的前台）。¹⁰

L3 有边界的自治 Agent

以状态目标为导向，具备反馈闭环，能在一定边界内自主试错，并提供人工接管/确认/回滚机制。Operator/ChatGPT agent、Bedrock Agents、Devin 大体落在此层。⁶⁰

L4 企业级规模化与多 Agent 生态

不仅能自治，还具备企业级治理（RBAC/IAM、审计、可观测性、数据策略、生态互操作），并支持多 agent 分发与协作。Agentforce 3/360 强调 Command Center 可观测性与 MCP 互操作，Google 以 Agent Designer/Gallery 与 A2A 推进多 agent 生态，Copilot Studio 以数据策略与企业审计强化自治边界，均指向 L4。⁶¹

产品映射表（以“主导形态”为准，允许跨层）：

产品/能力	主要形态	推荐成熟度定位	关键依据（摘要）
OpenAI Operator / ChatGPT agent	Browser/Computer-use Agent	L3	虚拟电脑执行任务、关键动作确认、Watch Mode、可中断接管 ⁶²
Anthropic Computer Use	Computer-use capability	L2-L3（取决于编排权）	提供 agent loop 与注入检测/确认，自治程度依赖宿主实现 ³¹
Microsoft Copilot Studio	Enterprise Agent Platform	L3→L4	事件触发后台自治 + 连接器/知识源 + 数据策略/审计/安全治理 ⁶³
Salesforce Agentforce 3/360	Enterprise Agent Platform	L4	可观测性(Command Center) + 确定性控制 + MCP 互操作 + 生态动作面 ⁴⁵
Google Agentspace/Gemini Enterprise	Enterprise Agentic Platform	L4（平台倾向）	Agent Designer/Gallery + actions 授权 + A2A 多 agent 互操作 ⁶⁴
Amazon Bedrock Agents	Cloud Orchestrated Agents	L3（可向 L2 收紧）	官方运行时循环 + action groups + custom orchestrator 状态机 ⁶⁵

产品/能力	主要形态	推荐成熟度定位	关键依据（摘要）
Devin	Engineering Agent	L3	计划-测试-迭代闭环，目标导向交付与 PR 工作流 ⁶⁶

对比分析框架：三维评分体系、对比表与雷达图

为了可重复对比，本报告采用三维评分框架（每维 1-5 分）。评分不是官方结论，而是“基于公开资料的分析者评估”，用于把不同形态产品放到同一张坐标系中。

- 三维评分定义：
- **自主性与可控性**：是否具备闭环自治（计划/行动/纠错/追问）、是否提供人工接管/确认、是否存在明确护栏与决策边界。⁶⁷
 - **集成与行动表面**：可调用的工具/连接器生态、对企业数据的可达性、能否在无 API 场景用 GUI 自动化扩展动作面。⁶⁸
 - **治理与安全**：权限/审计/可观测性、DLP 与数据策略、提示注入/敏感动作确认、生态互操作带来的供应链风险管理。⁶⁹

产品三维评分表（作者评估，1-5）：

产品	自主性与可控性	集成与行动表面	治理与安全
OpenAI Operator/ChatGPT Agent Mode	4.0	3.2	2.6
Anthropic Computer Use	3.2	3.0	3.6
Microsoft Copilot Studio	4.3	4.6	4.8
Salesforce Agentforce	4.7	4.8	4.9
Google Agentspace/Gemini Enterprise	4.1	4.5	4.6
Amazon Bedrock Agents	4.0	4.1	4.3
Devin	4.2	3.6	3.2

评分解读（与公开证据对齐）：企业平台型产品在“治理与安全、集成行动面”上普遍更高，因为其官方文档/发布强调数据策略、审计、可观测性与生态互操作；例如 Copilot Studio 列出 Purview 审计、Sentinel 监控、DLP 与数据策略等控制项。⁷⁰ Agentforce 3 把“可观测性与控制”作为规模化关键引入 Command Center，并强调 MCP 互操作与生态。³⁹ Google Gemini Enterprise 则通过 Agent Designer/Gallery 与 actions 授权机制体现“平台治理+分发”。⁷¹

雷达图已在文中插入（Analyst Assessment, 1-5），用于直观呈现三维差异；对比结论应以“三维 profile”而非单一分数理解（例如 Browser Agent 自主性高但企业治理相对弱，企业平台则相反）。⁷²

技术架构模式总结：市场主流 Agent 的底层实现路径

市场上的“前沿 agent 产品”底层并非神秘黑箱，大体可归纳为四种可复用的架构模式；它们对应了本报告开头对“workflow vs agent”的讨论：模式越往右，越强调“闭环自治与治理”，也越接近企业可规模化落地。

Computer-use Agent 模式：GUI 作为通用行动接口

核心是“截图感知 + 动作执行 + 反馈循环”。OpenAI 与 Anthropic 都把 computer use 描述为一个连续循环：

模型提出动作（点击/输入/滚动），宿主执行后回传截图，模型再决策下一步。⁷³ 该模式最大价值是“无 API 自动化”，但也带来 UI 脆弱性与提示注入风险，因此两家都强调沙箱、最小权限、allowlist 与关键动作确认。

⁷⁴ Microsoft Copilot Studio 通过 allow-list、凭据管理、Windows 365 托管环境把该模式产品化，进一步证明“GUI agent 正在成为企业自动化新层”。⁷⁵

Enterprise Agent Platform 模式：元数据驱动编排 + 企业治理

这一模式的共同点是把“工具/主题/知识源/连接器”都做成有元数据的可管理资源，让系统能基于元数据构建计划与执行路径，并在权限与审计体系内运行。Copilot Studio 的 generative orchestration 就是典型：系统用可用能力的名称/描述/输入输出元数据构建计划并串联回调。¹⁰ Agentforce 将“数据-推理-行动”作为三要素，并将“条件过滤（系统层 gating）与确定性控制”作为降低语义噪声、提高可靠性与安全性的工程手段。

⁷⁶ Gemini Enterprise 则把“agent 分发（Gallery）+ no/low-code 创建与多步调度（Designer）+ actions 授权”组合为企业级前门。⁷⁷

Cloud Orchestrator 模式：托管编排循环 + action groups/knowledge bases/guardrails

Bedrock Agents 代表了云厂商的“托管自治”：把编排循环标准化（pre-processing + orchestration loop），把外部动作抽象为 action groups（典型落到 Lambda），把企业数据引入 knowledge bases，并允许通过 custom orchestrator 把编排权改为企业自定义状态机。⁶⁵ 这种模式的优势是“快速落地”，风险是“过早追求自治会导致成本与治理复杂度上升”，与 Gartner 对取消项目的警告相呼应。⁷⁸

Engineering Loop Agent 模式：测试驱动的行动闭环

Devin 的“Ticket→Plan→Test→PR”展示了工程 agent 的典型循环：以可验证的测试/构建结果作为反馈，允许多轮试错，最后交付可审阅的变更。⁵⁷ 这种模式之所以更容易成为“真 agent”，根本原因是软件工程环境天然具备可回滚、可测试、可审计的闭环基础设施。

跨模式的共同趋势是：互操作协议与工具生态正在成为“平台竞争的主战场”。MCP 试图标准化“agent ↔ 工具/数据源”的连接，A2A 试图标准化“agent ↔ agent”的协作。⁷⁹ 企业平台在 2025-2026 的关键升级方向也高度一致：可观测性（trace/metrics）、权限治理、生态连接器与互操作标准。⁸⁰

商业化、风险与战略建议：从 ROI 到治理落地的关键路径

商业价值来源可以归纳为三类：其一是“替代重复劳动”，尤其在 IT/知识管理/客服等职能，McKinsey 与其后续解读都指出 agent 的规模化采用更常出现在这些场景。⁸¹ 其二是“打通无 API 的长尾系统”，computer-use / browser agent 让传统 RPA 的高维护成本和脆弱性被重新定义，Microsoft 在 computer use 预览中直接把“无 API 表单填写、数据搬运、市场研究”作为核心用例。⁸² 其三是“平台化的数字劳动力”，以 Agentforce 3/360 为代表，供应商把产品定位从“助手”升级为“可管理的 AI workforce”，并将可观测性/互操作与预制行业动作与 ROI 绑定。⁸³

成本结构需要比传统 AI 应用更“全栈”：除了模型推理成本与工具调用成本，还包括连接器/数据管道建设、权限治理、审计与监控、以及不可忽视的人类监督成本。Gartner 指出大量项目会因为成本上升、价值不清或风险控制不足而取消，并强调很多用例其实不需要 agentic 实现。⁸⁴ 因此一个可行的采用路径通常是：先用 L2 级 agentic workflow 在低风险闭环试点（以可观测性和可回滚为前提），逐步把“动作面”收紧并引入权限/审计，再在特定职能/流程中升级到 L3 自治；最后才谈 L4 多 agent 协作与跨系统自治。⁸⁵

风险与挑战方面，AI Agent 相比传统助手的风险更“操作型”：

提示注入与恶意内容诱导是被反复证实的首要风险。OpenAI 在系统卡与安全文章中明确把 prompt injection 作为关键风险向量，并通过 Watch Mode、关键动作确认、拒绝高风险任务与红队/监控体系来缓解；同时建议在“logged-out mode”等策略下减少敏感数据暴露。⁸⁶ Anthropic 的 computer use 文档同样强调网页/图像中的指令可能覆盖用户意图，因此建议域名 allowlist、最小权限 VM、关键决策人工确认，并引入注入检测分类器触发确认。³¹ 这类风险也在 OWASP 的 LLM Top 10 中被归为首要威胁（Prompt Injection），并与不安全输出处理、供应链漏洞等共同构成 agent 系统的攻击面。⁸⁷

环境脆弱性与责任归属是第二类难题：computer-use agent 依赖 UI 变化与外部网站策略，虽厂商宣称“能适应界面变化”，但本质仍是把复杂性从 API 层迁移到 UI/策略层。⁸⁸ 当 agent 错误执行动作时，责任需要在“模型→系统编排→权限配置→监督者”之间可追溯，这也是企业平台强调审计与可观测性的原因。Copilot Studio 将“可审计流程与数据策略”、Agentforce 3 将“Command Center 可观测性”作为规模化 blocker 的解决方案，体现治理已成为产品核心差异。⁸⁹

一个可落地的治理策略组合（与供应商实践对齐）通常包括：

最小权限与动作分级（高风险动作必须确认/审批）、沙箱与隔离（尤其 computer-use 环境）、域名/应用 allowlist、对话与动作日志（审计与取证）、以及对工具生态的供应链管理。OpenAI 在 MCP 工具文档中也提醒：第三方 MCP server 可能请求敏感数据或包含隐藏指令，并默认要求审批与建议日志记录。⁹⁰ 在更广泛的風險管理框架上，NIST AI RMF 强调为 AI 系统建立跨生命周期的风险识别、评估与缓解体系，可作为企业引入 agent 时的治理参考。⁹¹

机会与建议可以聚焦三条确定性方向：

第一，**平台化治理与可观测性**会持续成为企业购买决策的核心，Agentforce 3 的定位就是“解决规模化的可见性与控制”，并将互操作协议与生态动作面作为平台扩展关键。³⁹ 第二，**无 API 自动化 (computer-use)**会把 agent 的可行动表面显著扩大，但也必须与 allowlist/凭据治理/沙箱结合；Copilot Studio 与 OpenAI/Anthropic 的安全建议均指向“默认不信任环境、用制度把风险关进笼子”。⁹² 第三，**垂直行业的“受控自治”**将是更现实的商业化路径：Gartner 与 McKinsey 都提示大多数组织仍停留在试点，能规模化的用例常集中在 IT、知识管理等特定职能；垂直场景更容易定义动作边界与审计要求，从而让 L3/L4 的自治变得可控。⁹³ 最后，建议把互操作标准 (MCP/A2A) 视为中长期的“生态杠杆”，尤其在多 agent 协作与跨系统编排成为常态后，标准化连接将显著降低集成成本。⁷⁹

附录：检索日志、数据来源表与案例单页模板

本附录提供“可检索、可复现”的报告结构件，便于你把本报告方法迁移为自己的研究流程（例如在作品集或团队调研中复用）。

检索日志模板 (Query Log)

建议在研究中记录每条 query 的时间、目的与筛选规则，确保后续复盘可复现（尤其面对 2025–2026 快速迭代的 agent 市场）。
- 字段建议：检索日期、检索工具（官方站/媒体/学术）、Query、筛选标准（官方优先/近 12 个月优先/是否需二次核验）、关键证据点摘要、来源链接/引用 ID。
- 典型 query 例子可围绕“产品发布+官方文档+运行机制+治理能力”四类展开：例如 OpenAI “computer use loop”、Bedrock “InvokeAgent runtime process”、Copilot Studio “autonomous agents governance” 等。⁹⁴

数据来源分层表 (Source Quality Table)

为避免“agent washing”带来的营销噪声，建议把来源分层管理：

- 一手材料（最高优先）：官方产品文档、系统卡、开发者文档、发布说明（如 Operator 系统卡、Copilot Studio Learn Docs、Bedrock Docs、Gemini Enterprise Docs）。⁹⁵
- 权威媒体（用于市场与商业化）：Reuters 等（例如 Salesforce 与 OpenAI/Anthropic 合作、Cognition 收购 Windsurf）。⁹⁶
- 行业研究（用于趋势与采用率）：Gartner、McKinsey、BCG 等（采用率、取消率、管理建议）。⁹⁷
- 社区/二手解读（最低优先）：仅用于补充背景，不作为关键结论依据（本报告已尽量避免依赖此类来源）。

¹⁴

典型案例单页模板 (Case Sheet)

每个产品建议用 1 页结构化描述，便于横向对比：
- 产品定位与目标用户 (Browser Agent / Enterprise Platform / Cloud Runtime / Engineering Agent)⁹⁸
- 运行架构（文字流程图）与关键模块（感知/规划/行动/反馈/治理）⁹⁹

- 行动表面清单（API、连接器、RAG、GUI computer-use） 100
- 治理机制清单（权限、审计、确认、可观测性、allowlist、沙箱） 101
- Agent 属性判定（对照本报告定义：满足/部分满足/不满足） 102
- 主要风险与缓解策略（Prompt Injection、供应链、环境脆弱性） 103

以上附录结构与本报告方法一致：先定义边界、再分层市场、再用同一框架拆解产品并评分，最后回到治理与价值实现路径。Gartner 对“agent washing”与项目取消的警告，恰好说明这种“先立标准再比较”的方法论在 2025–2026 的 agent 市场中不仅必要，而且是减少试错成本的关键。 84

- 1 5 102 <https://www.gartner.com/en/newsroom/press-releases/2024-10-21-gartner-identifies-the-top-10-strategic-technology-trends-for-2025>
<https://www.gartner.com/en/newsroom/press-releases/2024-10-21-gartner-identifies-the-top-10-strategic-technology-trends-for-2025>
- 2 81 93 https://www.mckinsey.com/~media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai/november%202025/the-state-of-ai-2025-agents-innovation_cmyk-v1.pdf
https://www.mckinsey.com/~media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai/november%202025/the-state-of-ai-2025-agents-innovation_cmyk-v1.pdf
- 3 6 37 63 67 89 <https://learn.microsoft.com/en-us/microsoft-copilot-studio/guidance/autonomous-agents>
<https://learn.microsoft.com/en-us/microsoft-copilot-studio/guidance/autonomous-agents>
- 4 <https://docs.aws.amazon.com/bedrock/latest/userguide/agents.html>
<https://docs.aws.amazon.com/bedrock/latest/userguide/agents.html>
- 7 16 24 26 98 <https://openai.com/index/introducing-operator/>
<https://openai.com/index/introducing-operator/>
- 8 27 73 94 99 <https://platform.openai.com/docs/guides/tools-computer-use>
<https://platform.openai.com/docs/guides/tools-computer-use>
- 9 86 95 <https://openai.com/index/operator-system-card/>
<https://openai.com/index/operator-system-card/>
- 10 <https://learn.microsoft.com/en-us/microsoft-copilot-studio/faqs-generative-orchestration>
<https://learn.microsoft.com/en-us/microsoft-copilot-studio/faqs-generative-orchestration>
- 11 12 85 <https://aws.amazon.com/blogs/machine-learning/getting-started-with-amazon-bedrock-agents-custom-orchestrator/>
<https://aws.amazon.com/blogs/machine-learning/getting-started-with-amazon-bedrock-agents-custom-orchestrator/>
- 13 <https://www.mckinsey.com/featured-insights/week-in-charts/agentic-ai-advances>
<https://www.mckinsey.com/featured-insights/week-in-charts/agentic-ai-advances>
- 14 78 84 97 <https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027>
<https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027>
- 15 <https://www.bcg.com/publications/2025/machines-that-manage-themselves>
<https://www.bcg.com/publications/2025/machines-that-manage-themselves>

17 39 43 45 61 80 83 <https://www.salesforce.com/news/press-releases/2025/06/23/agentforce-3-announcement/>

<https://www.salesforce.com/news/press-releases/2025/06/23/agentforce-3-announcement/>

18 <https://docs.devin.ai/>

<https://docs.devin.ai/>

19 53 55 65 <https://docs.aws.amazon.com/bedrock/latest/userguide/agents-how.html>

<https://docs.aws.amazon.com/bedrock/latest/userguide/agents-how.html>

20 79 <https://www.anthropic.com/news/model-context-protocol>

<https://www.anthropic.com/news/model-context-protocol>

21 <https://www.anthropic.com/news/donating-the-model-context-protocol-and-establishing-of-the-agentic-ai-foundation>

<https://www.anthropic.com/news/donating-the-model-context-protocol-and-establishing-of-the-agentic-ai-foundation>

22 <https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/>

<https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/>

23 90 <https://platform.openai.com/docs/guides/tools-connectors-mcp>

<https://platform.openai.com/docs/guides/tools-connectors-mcp>

25 28 60 62 <https://openai.com/index/introducing-chatgpt-agent/>

<https://openai.com/index/introducing-chatgpt-agent/>

29 <https://www.anthropic.com/news/developing-computer-use>

<https://www.anthropic.com/news/developing-computer-use>

30 31 74 <https://platform.claude.com/docs/en/agents-and-tools/tool-use/computer-use-tool>

<https://platform.claude.com/docs/en/agents-and-tools/tool-use/computer-use-tool>

32 <https://learn.microsoft.com/en-us/microsoft-copilot-studio/advanced-generative-actions>

<https://learn.microsoft.com/en-us/microsoft-copilot-studio/advanced-generative-actions>

33 68 100 <https://learn.microsoft.com/en-us/microsoft-copilot-studio/advanced-connectors>

<https://learn.microsoft.com/en-us/microsoft-copilot-studio/advanced-connectors>

34 75 82 92 <https://www.microsoft.com/en-us/microsoft-copilot/blog/copilot-studio/computer-use-is-now-in-public-preview-in-microsoft-copilot-studio/>

<https://www.microsoft.com/en-us/microsoft-copilot/blog/copilot-studio/computer-use-is-now-in-public-preview-in-microsoft-copilot-studio/>

35 <https://learn.microsoft.com/en-us/microsoft-copilot-studio/configure-where-computer-use-runs>

<https://learn.microsoft.com/en-us/microsoft-copilot-studio/configure-where-computer-use-runs>

36 69 70 101 <https://learn.microsoft.com/en-us/microsoft-copilot-studio/security-and-governance>

<https://learn.microsoft.com/en-us/microsoft-copilot-studio/security-and-governance>

38 40 76 <https://www.salesforce.com/agentforce/how-it-works/>

<https://www.salesforce.com/agentforce/how-it-works/>

41 <https://www.salesforce.com/agentforce/guide/>

<https://www.salesforce.com/agentforce/guide/>

42 <https://www.salesforce.com/agentforce/five-levels-of-determinism/>

<https://www.salesforce.com/agentforce/five-levels-of-determinism/>

- 44 96 <https://www.reuters.com/business/salesforce-deepens-ai-ties-with-openai-anthropic-power-agentforce-platform-2025-10-14/>
https://www.reuters.com/business/salesforce-deepens-ai-ties-with-openai-anthropic-power-agentforce-platform-2025-10-14/
- 46 <https://cloud.google.com/blog/products/ai-machine-learning/bringing-ai-agents-to-enterprises-with-google-agentspace>
https://cloud.google.com/blog/products/ai-machine-learning/bringing-ai-agents-to-enterprises-with-google-agentspace
- 47 <https://cloud.google.com/blog/products/ai-machine-learning/google-agentspace-enables-the-agent-driven-enterprise>
https://cloud.google.com/blog/products/ai-machine-learning/google-agentspace-enables-the-agent-driven-enterprise
- 48 51 <https://docs.cloud.google.com/gemini/enterprise/docs/release-notes>
https://docs.cloud.google.com/gemini/enterprise/docs/release-notes
- 49 <https://cloud.google.com/gemini-enterprise>
https://cloud.google.com/gemini-enterprise
- 50 52 64 71 <https://docs.cloud.google.com/gemini/enterprise/docs/agent-designer>
https://docs.cloud.google.com/gemini/enterprise/docs/agent-designer
- 54 <https://docs.aws.amazon.com/bedrock/latest/userguide/agents-action-create.html>
https://docs.aws.amazon.com/bedrock/latest/userguide/agents-action-create.html
- 56 <https://cognition.ai/blog/introducing-devin>
https://cognition.ai/blog/introducing-devin
- 57 59 66 <https://devin.ai/>
https://devin.ai/
- 58 <https://www.reuters.com/legal/transactional/cognition-ai-buy-windsurf-doubling-down-ai-driven-coding-2025-07-14/>
https://www.reuters.com/legal/transactional/cognition-ai-buy-windsurf-doubling-down-ai-driven-coding-2025-07-14/
- 72 <https://openai.com/index/prompt-injections/>
https://openai.com/index/prompt-injections/
- 77 <https://docs.cloud.google.com/gemini/enterprise/docs/agent-gallery>
https://docs.cloud.google.com/gemini/enterprise/docs/agent-gallery
- 87 <https://owasp.org/www-project-top-10-for-large-language-model-applications/>
https://owasp.org/www-project-top-10-for-large-language-model-applications/
- 88 <https://learn.microsoft.com/en-us/microsoft-copilot-studio/computer-use>
https://learn.microsoft.com/en-us/microsoft-copilot-studio/computer-use
- 91 <https://www.nist.gov/itl/ai-risk-management-framework>
https://www.nist.gov/itl/ai-risk-management-framework
- 103 <https://genai.owasp.org/lmrisk/lm01-prompt-injection/>
https://genai.owasp.org/lmrisk/lm01-prompt-injection/