JIU SI HANG YAN



2026年中国AI智能体商业化落地深度研究报告

九思行研
2026.11

前言

AI 智能体是指能够自主感知环境、规划任务、调用工具并执行操作的智能系统。与聊天式 AI 的单轮问答不同，智能体具备多步骤任务拆解与自主闭环能力，其产业价值从提供信息转向交付结果。这一能力跃迁使智能体成为大模型商业化最受关注的载体，也使其面临比聊天式 AI 更复杂的部署门槛与成本结构。

2026 年是智能体商业化的验证之年。从市场数据看，中国企业级 AI 智能体市场规模 2025 年约 212 亿元，2026 年预计增至 449 亿元，2029 年有望突破 3320 亿元，年复合增速超过 100%。从部署进度看，2026 年第一季度交付或更新的企业应用中已有 80%嵌入智能体，但仅 31%的机构有智能体运行于生产环境，18%的企业将其纳入核心业务流程。从收入模式看，66%的中国企业倾向基于业务成果计费，按结果付费正在从概念走向合同条款。三项数据共同勾勒出一个基本判断：智能体的技术可行性已被验证，商业闭环正在少数场景中跑通，但规模化复制仍受制于治理能力、数据准备度与成本控制。

场景分化的节奏已经清晰。金融与编程处于规模化验证的收尾阶段，商业闭环基本跑通；法律处于渐进渗透的中段，单点效率验证完成、规模化复制进行中；政务在公共价值维度表现突出，商业模式以项目制与采购制为主；医疗受监管与数据成本约束，节奏后置但政策确定性增强；工业处于早期信号阶段，预测性替换单点回报明确，系统级落地尚需数据基础设施先行。

当前行业最值得关注的结构性变化，是商业模式从卖工具向卖结果的代际切换。这一切换重新定义了智能体的价值锚点，从软件预算转向人力成本，从席位数量转向业务产出。供应商承担交付风险的同时获得收入天花板的上移，客户以效果付费换取确定性。模式的成熟度将决定未来三年智能体产业的价值分配格局，也决定了不同场景、不同规模企业之间的分化速度。

目 录

第一章 AI 智能体商业化落地的分析框架与拐点判断 1

1.1 AI 智能体商业闭环验证的判断标准与数据口径 1
1.1.1 部署率、续费率与单位经济模型的衡量作用 1
1.1.2 企业级市场与整体市场的统计口径辨析 2
1.1.3 项目制收入与订阅制收入的质量差异 2
1.1.4 2025-2026 年制度节点与产业节奏的对应关系 3
1.2 AI 智能体部署率与价值交付之间的现实鸿沟 4
1.2.1 企业应用集成率与真实部署率的落差来源 4
1.2.2 试点项目取消率背后的成本结构变化 5
1.2.3 从按席位收费到按结果收费的早期信号 5
1.2.4 不同行业与不同规模企业的部署节奏差异 6
1.3 AI 智能体供给格局中的平台型、垂直型与自研型力量 6
1.3.1 头部科技企业的平台化路径与生态卡位 6
1.3.2 垂直领域专业厂商的行业 Know-how 壁垒 6
1.3.3 需求方自研团队的边界与局限 7
1.3.4 开源生态与闭源生态对供给格局的影响 7
1.4 报告中 AI 智能体场景选择与评估框架 8
1.4.1 落地深度、市场规模、付费意愿与监管约束的四维评估 8
1.4.2 金融、法律、医疗、编程、政务与工业的节奏图谱 8
1.4.3 投资窗口的判断标准与数据来源说明 9

第二章 金融 AI 智能体商业化验证的领先赛道 10

2.1 金融 AI 智能体成为最快落地场景的结构性原因 10
2.1.1 合规与数据安全要求对落地加速的反向作用 10
2.1.2 业务标准化程度与付费能力的结构性优势 10
2.1.3 2026 年金融智能体元年的供需信号与机构动作 11
2.1.4 金融 IT 预算结构与智能体采购模式的匹配关系 11
2.2 金融 AI 智能体平台生态与落地规模 11
2.2.1 腾讯 WorkBuddy 金融版等平台型智能体的公开落地进展 11
2.2.2 头部券商、基金公司与银行的自研及合作模式差异 11
2.2.3 投研、合规、信贷尽调与财富资配等核心场景的渗透顺序 12
2.2.4 保险与银行客服场景的部署节奏比较 13
2.2.5 中小金融机构的采纳障碍与替代路径 13
2.3 金融 AI 智能体从辅助工具到数字员工的业务跃迁 13
2.3.1 华泰、国泰海通、中金与国信等券商的智能体部署路径 13
2.3.2 业务流程重构与组织变革的实证观察 13

2.3.3 增收、降本与风险控制三类商业模型变化 14
2.4 金融 AI 智能体投资窗口研判 14
2.4.1 平台层、垂直应用层与模型层的价值分配 14
2.4.2 未来 2-3 年确定性最高的细分环节 15
2.4.3 竞争拥挤度与差异化机会 15
2.4.4 数据合规与模型可解释性对投资节奏的约束 16

第三章 法律 AI 智能体在专业壁垒下的渐进渗透 17

3.1 法律 AI 智能体在“错不起”行业中的智能化逻辑 17
3.1.1 合同审查成为第一个规模化场景的原因 17
3.1.2 人机回圈作为行业刚需而非过渡方案 17
3.1.3 法律数据高壁垒与低容错对商业化节奏的影响 17
3.1.4 律所与公司法务部门的需求差异 18
3.2 法律 AI 智能体落地数据与真实效率提升 18
3.2.1 幂律智能等厂商公开披露的客户规模与合同审查数据 18
3.2.2 审查时间与风险发现率等指标的口径问题及可比性挑战 18
3.2.3 跨法域与跨行业合同模板差异对规模化的实际影响 19
3.2.4 效率提升向收入增长转化的实证观察 19
3.3 法律 AI 智能体推动律所端组织变革 19
3.3.1 大成与惠诚等律所的 AI 部署路径及战略合作 19
3.3.2 技能与智能体可复用化的组织含义 19
3.3.3 法律服务成本结构与竞争格局的变化 20
3.4 法律 AI 智能体投资窗口研判 20
3.4.1 垂直厂商与通用平台之间的边界 20
3.4.2 领域语料与行业 Know-how 的估值权重 20
3.4.3 2026-2028 年商业化节奏的关键节点 21
3.4.4 法律科技监管变化对投资窗口的影响 21

第四章 医疗 AI 智能体在监管红线下的价值锚点 22

4.1 医疗 AI 智能体的“辅助”定位与落地边界 22
4.1.1 三类医疗器械与药监局注册证的约束机制 22
4.1.2 诊断权归属与责任边界对产品设计的影响 22
4.1.3 医院采购与科室预算逻辑的现实约束 22
4.1.4 医生使用习惯与工作流嵌入的摩擦成本 23
4.2 医疗 AI 智能体已落地案例的结构性分析 23
4.2.1 协和、地坛与河北医大一院等综合智能体案例 23
4.2.2 仁济医院泌尿专科等专科智能体的规模化服务数据 24
4.2.3 临床决策、病历生成与预后评估的场景分化 24
4.2.4 区域医疗中心与基层医院部署模式的差异 24
4.2.5 案例数据公开程度与可复制性评估 25

4.3 医疗 AI 智能体数据成本与商业化瓶颈 25
4.3.1 脱敏、授权与专家标注成本居高不下的原因 25
4.3.2 医保支付改革对新付费场景的创造作用 25
4.3.3 规模化速度与投资回报周期的现实匹配 26
4.4 医疗 AI 智能体投资窗口研判 26
4.4.1 长周期背景下政策确定性增强的逻辑 26
4.4.2 影像、病历与辅助诊断等细分方向的优先级 26
4.4.3 平台型与专科型厂商的竞争格局 26
4.4.4 医院回款周期与商业化可持续性 27

第五章 编程 AI 智能体与其他高价值场景的商业成熟度 28

5.1 编程 AI 智能体渗透率最高的技术适配性 28
5.1.1 代码结构化、可验证与反馈闭环短的技术优势 28
5.1.2 开发者付费意愿与价值量化的商业优势 28
5.1.3 从补全工具到自主智能体的能力跃迁 29
5.1.4 编程智能体对软件工程流程的重塑 29
5.2 编程 AI 智能体企业级部署的规模化信号 29
5.2.1 字节 TRAE 与腾讯 CodeBuddy 等内部渗透数据 29
5.2.2 OpenAI Codex 用户增长与 tokens 处理量的公开信号 30
5.2.3 软件开发组织形态与外包市场的影响 30
5.2.4 企业级安全与代码资产保护的实际要求 30
5.3 公共价值驱动下的政务 AI 智能体特殊赛道 31
5.3.1 北京 12345“京京”、深圳福田与台州检察院等案例 31
5.3.2 政府直采、国企平台运营与科技企业共建的采购模式 31
5.3.3 商业可持续性与效率指标的衡量方式 31
5.4 工业与制造 AI 智能体的早期信号 32
5.4.1 设备故障响应与预测性维护的落地数据 32
5.4.2 工业智能体目标下的部署节奏与行业差异 33
5.4.3 与金融、编程场景在商业化逻辑上的根本区别 34
5.4.4 工业数据孤岛与 OT 系统集成难题 34
5.5 跨场景 AI 智能体投资窗口研判 35
5.5.1 编程赛道商业化最确定的方向 35
5.5.2 政务与工业场景的公共价值及长周期回报 35
5.5.3 跨场景可复用的平台能力与工具链机会 35
5.5.4 垂直场景之间的投资组合与风险分散 35

第六章 AI 智能体商业模式、投资窗口与风险研判 37

6.1 AI 智能体从卖工具到卖结果的商业模式代际切换 37
6.1.1 按结果付费的全球实践与 Salesforce、Zendesk、Sierra、OpenAI 案例 37
6.1.2 中国市场的智能体即服务与价值分成模式 38

6.1.3 对单位经济模型与收入质量的实际影响 38
6.1.4 按结果付费的计量争议与合同设计难点 39
6.2 AI 智能体资本流向与投资优先级 39
6.2.1 代理式 AI 融资规模与增速的权威数据 39
6.2.2 平台层、垂直应用层与基础设施层的资本分布 41
6.2.3 金融、法律、医疗、编程、政务与工业的投资吸引力排序 41
6.3 AI 智能体核心风险与项目取消率的结构性成因 42
6.3.1 单 Agent 任务 LLM 调用量远高于聊天交互的成本失控问题 42
6.3.2 业务价值不清与风险管控不足的结构性原因 42
6.3.3 技术路线、监管走向与地缘政治的不确定性 43
6.3.4 项目取消后的组织学习与二次部署机会 43
6.4 AI 智能体未来 12-18 个月关键观察指标 44
6.4.1 企业级规模化部署率与续费率 44
6.4.2 按结果付费渗透率与单位经济模型 44
6.4.3 各垂直场景 ROI 验证数据 44
6.4.4 监管框架演进方向与合规成本变化 45
6.4.5 智能体安全事件与信任修复机制 45

图表 1 中国企业智能体采纳阶段分布 1
图表 2 2025-2029 年中国企业级 AI 智能体市场规模及预测 2
图表 3 “人工智能+”行动中智能体应用普及率目标路线 4
图表 4 2026 年智能体企业级渗透的落差结构 4
图表 5 六大场景商业化成熟度四维评估 8
图表 6 六大场景规模化验证窗口期 9
图表 7 2026 年证券行业算法与模型备案进展 10
图表 8 金融业智能体核心场景渗透顺序（相对指数） 12
图表 9 2026-2029 年金融智能体产业链价值分配及演变预测 15
图表 10 合同审查 AI 辅助前后的核心指标对比（幂律智能口径） 18
图表 11 医疗 AI 监管准入进展 23
图表 12 代表性医疗智能体服务规模对比 24
图表 13 2025-2026 年中国 AI 编程市场规模 28
图表 14 2026 年中国 AI 编程工具市场份额（使用率口径） 30
图表 15 政务智能体核心效率指标 32
图表 16 工业预测性维护的核心改善幅度 33
图表 17 工业智能体的成熟度与数据利用缺口 34
图表 18 六大场景投资吸引力与确定性综合评分 36
图表 19 主要厂商按结果付费定价对比 37
图表 20 AI 与人工单次解决成本对比 39
图表 21 2025-2026 年中国代理式 AI 融资规模 40
图表 22 2026 年全球智能体垂直场景融资分布 40
图表 23 资本配置优先级：六大场景投资吸引力排序 41
图表 24 智能体任务的 Token 消耗量级对比（指数） 42

图表 25　代理式 AI 项目取消率与已部署项目回报并存 43
图表 26　未来 12-18 个月关键观察指标的基线与阈值 45

第一章　AI 智能体商业化落地的分析框架与拐点判断

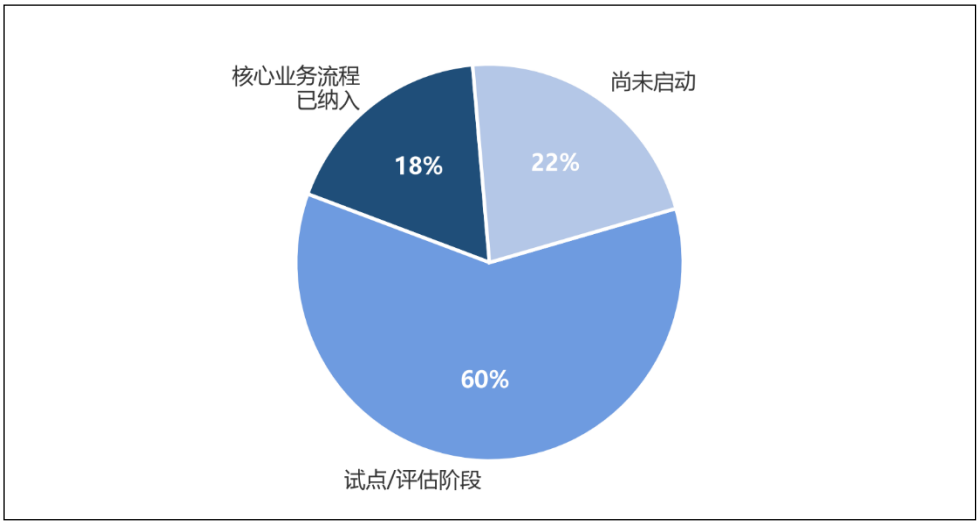
1.1　AI 智能体商业闭环验证的判断标准与数据口径

1.1.1　部署率、续费率与单位经济模型的衡量作用

判断一个行业的智能化是否真正跨越商业化拐点，不能只看概念热度与融资新闻，而要回到三个可核验的底层指标上。其一，部署率反映技术从采购到真实使用的转化深度，其核心问题在于企业购买与实际部署之间的落差；其二，续费率反映用户在首个合同周期结束后是否愿意继续付费，是检验产品真实价值最诚实的指标，比任何厂商自述的效率提升都更具说服力；其三，单位经济模型衡量单次服务或单个智能体的收入与成本之比，决定商业模式能否规模化复制。只有三个指标同时向好，才能认定商业闭环已经跑通。

从 2026 年 10 月的最新数据看，三个指标呈现明显的分化格局。部署维度，IDC 调研显示约 60%的中国企业仍处于了解评估与试点智能体阶段，仅 18%的企业将智能体纳入核心业务流程；Gartner Q1 2026 调查显示，2026 年第一季度交付或更新的企业应用中有 80%至少嵌入了一个 AI 智能体，但 S&P Global Market Intelligence 的数据表明仅 31%的机构有智能体实际运行于生产环境。续费维度，头部场景（金融、编程）的复购信号已经出现——中工互联董事长公开表示，2026 年公司目标转向“从 1 到 1010”的规模化复制阶段，并称“99%的精力都砸在技术上，泡在客户现场推动场景与技术结合”；而长尾场景的续费仍高度依赖项目验收而非主动续约。单位经济维度，成功部署的代理式 AI 项目全球平均投资回报率达 171%；另据 BCG 和 Forrester 2026 年调查，智能体部署的中位回本周期为 5.1 个月，但这一回报集中于少数跑通闭环的项目，多数项目止步于治理、范围界定或数据质量问题。

图表 1　中国企业智能体采纳阶段分布

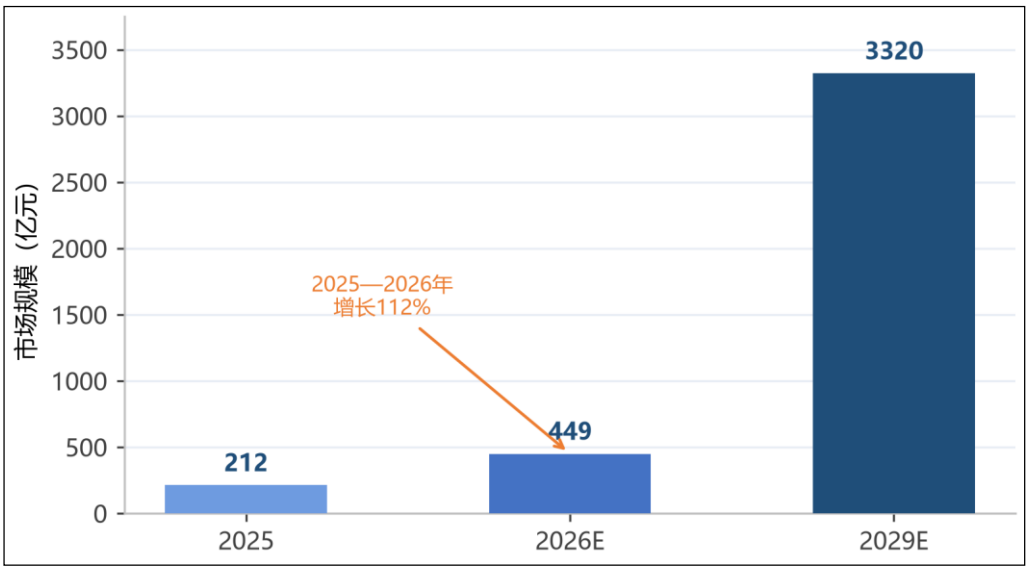


数据来源：IDC，九思行研整理

1.1.2 企业级市场与整体市场的统计口径辨析

市场上关于智能体市场规模的数字差异极大，根源在于统计口径不同。企业级市场口径只统计企业客户为智能体产品、平台与相关服务支付的金额，科智咨询数据显示，2025 年中国企业级 AI 智能体市场规模约 212 亿元，预计 2026 年增至 449 亿元、2029 年有望突破 3320 亿元，年复合增长率高达 107%；赛迪顾问测算的智能体整体市场 2026 年约 135.3 亿元，增速超 70%，政务、制造、能源、金融四大领域合计占比超 70%。两个数字并不矛盾——科智咨询口径包含智能体运行所依赖的平台订阅、开发工具与实施服务，赛迪口径更聚焦智能体本体销售与集成。本报告统一采用科智咨询企业级口径作为市场规模基准，以避免全文数据冲突。

图表 2　2025-2029 年中国企业级 AI 智能体市场规模及预测



数据来源：科智咨询，九思行研整理

另一组需要辨析的口径是活跃智能体数量。IDC 预测（《中国智能体规模预测，2026-2031》），2026 年和 2027 年将是中国企业场景中活跃智能体数量增速最快的两年，单年同比增长超过 200%，到 2031 年将达到 3.5 亿个活跃智能体；其中基于低代码、无代码平台构建的智能体将从 2026 年的约 300 万个增长至 2031 年的近 2 亿个，始终占据全部活跃智能体的半数以上；独立智能体的活跃数量占比将从 2026 年的 7.5%升至 2031 年的 20.1%。数量口径反映生态繁荣度，但不能直接折算收入，企业在引用时应注明口径属性。

1.1.3 项目制收入与订阅制收入的质量差异

智能体供应商的收入结构直接决定其估值逻辑与抗周期能力。项目制收入以定制

开发、一次性实施为主，收入规模大但不可持续，毛利受人力成本挤压，且客户复购依赖新一轮立项；订阅制收入以平台年费、按席位或按调用量计费为主，具备可预测性与高毛利特征，是软件行业长期验证的优质收入形态。2025 年中国智能体开发平台私有化市场约 17.5 亿元、公有云约 2.5 亿元，私有化份额接近公有云的七倍——这一结构说明中国企业当前更愿意为“可控”付费，但也意味着大量收入仍以私有化项目形式存在，订阅化转型尚在早期。

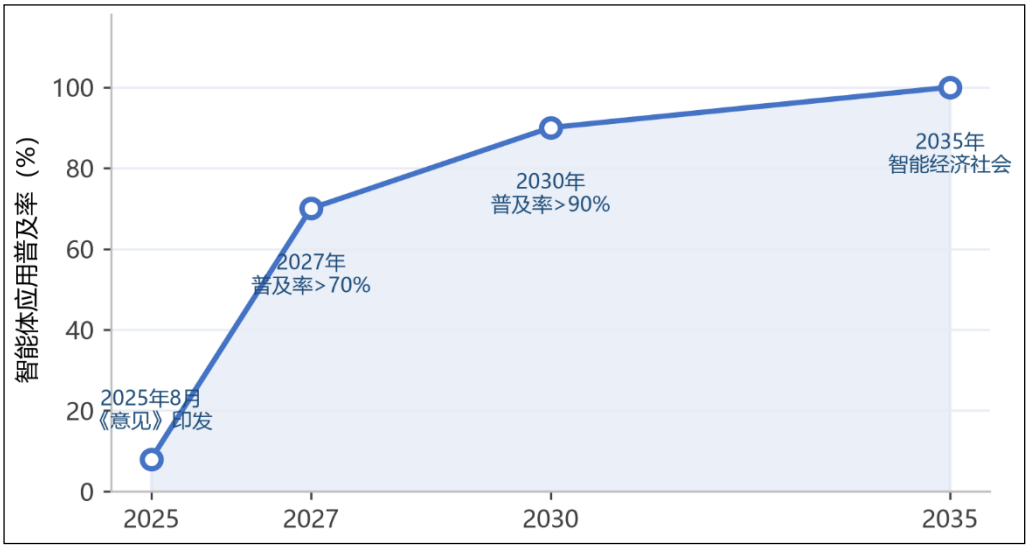
从收入质量演进的方向看，行业正在出现第三种形态——按结果付费，其收入兼具订阅制的可持续性与项目制的价值锚定，但计量与合同设计的复杂度最高。三种形态的并存与迁移，构成了判断智能体厂商商业成熟度的关键标尺。

1.1.4 2025-2026 年制度节点与产业节奏的对应关系

制度供给与产业节奏的共振，是 2025-2026 年智能体商业化加速的深层原因。2025 年 8 月，国务院印发《关于深入实施“人工智能+”行动的意见》（国发〔2025〕11 号），明确提出到 2027 年新一代智能终端、智能体等应用普及率超 70%，到 2030 年超 90%，并要求在软件、信息、金融、商务、法律等领域推动智能体广泛应用——这是智能体首次以专门术语进入国务院层级政策文件。随后各领域实施细则密集落地：2025 年 10 月，中央网信办、国家发展改革委联合印发《政务领域人工智能大模型部署应用指引》；2025 年 11 月，国家卫生健康委等五部门印发《关于促进和规范“人工智能+医疗卫生”应用发展的实施意见》；2026 年 3 月，国家医保局将 12 项 AI 辅助诊断项目纳入医保乙类目录。政策从顶层设计到行业细则的传导周期不到一年，显著压缩了需求方的决策观望期。

资本节奏与制度节奏形成呼应。毕马威数据显示，2025 年中国 AI 领域 PE/VC 投资总额达 406.4 亿美元，同比增长 58.5%，交易 2336 笔，同比增长 54%；其中代理式 AI 融资规模 2025 年达 294.7 亿美元，同比增长 39%，2026 年第一季度进一步增至 94.5 亿美元，同比增长 129.4%。资本从“大模型参数竞赛”转向“场景落地验证”，与政策对应用牵引的强调一致，共同将 2026 年推至智能体商业化的验证之年。

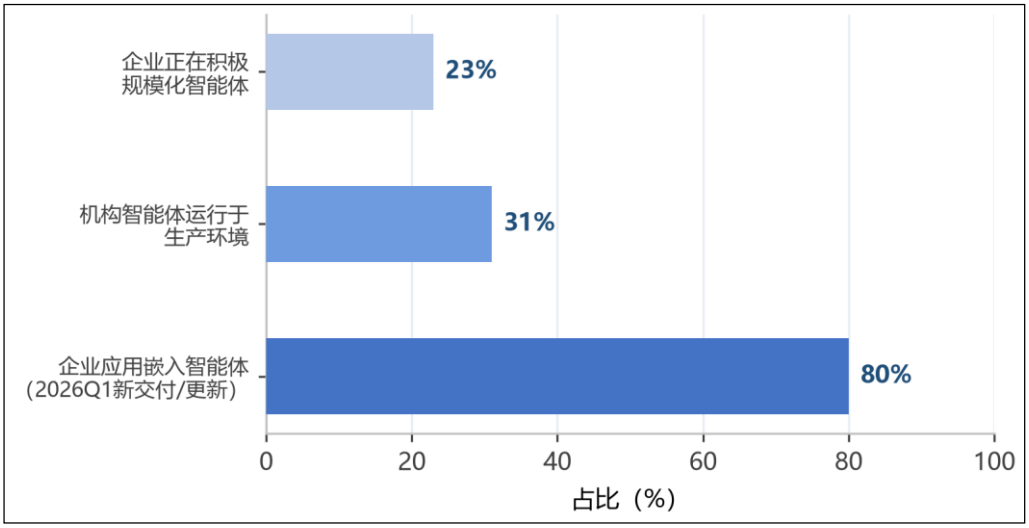
图表 3 “人工智能+”行动中智能体应用普及率目标路线



资料来源：国务院《关于深入实施“人工智能+”行动的意见》（国发〔2025〕11 号），九思行研整理

1.2 AI 智能体部署率与价值交付之间的现实鸿沟

图表 4 2026 年智能体企业级渗透的落差结构



数据来源：Gartner、S&P Global Market Intelligence、麦肯锡，九思行研整理

1.2.1 企业应用集成率与真实部署率的落差来源

80%的企业应用嵌入智能体与 31%的机构有智能体运行于生产环境之间，存在约 49 个百分点的落差。这一落差的来源可以从供给与需求两端理解。供给端，2025 年以来主流软件厂商将智能体能力作为标配嵌入产品更新，属于“供给侧默认搭载”，企业

采购新版本时被动获得智能体功能，但未必启用；需求端，企业真正将智能体接入核心业务流程需要完成数据准备、权限治理、流程再造与人员培训四项前置工程，任何一项缺位都会使智能体停留在演示层面。IDC 数据显示 38%的企业将数据准备度不足列为 AI 落地的首要障碍，这是落差最直接的技术成因。

从治理视角看，落差还反映了责任机制的缺位。智能体一旦接入核心系统，其错误输出的责任归属、调用边界与审计要求都需要制度化安排，多数企业在试点阶段尚未建立这套机制，因而主动将智能体隔离在生产环境之外。深圳福田区的实践提供了正面参照：其出台全国首个政务辅助智能机器人管理办法，建立“标准流程+负面清单”双重管控，为每一名数智员工配备专属“监护人”，用制度供给换部署深度——2026 年该区已有 43 类智能体全面上岗。

1.2.2 试点项目取消率背后的成本结构变化

Gartner 预测，到 2027 年底超过 40%的代理式 AI 项目将被取消，成本攀升、业务价值不清与风险管控不足是三大原因。其中成本结构的变化最值得警惕：与聊天式 AI 相比，单智能体任务的 LLM 调用量呈量级上升——企业月度 LLM 支出中位数在 2026 年第一季度同比增长 7.2 倍；一个编码智能体任务每日若执行 1000 轮、每轮重读 20 万 token 上下文，其算力成本将迅速吞噬项目预算。试点阶段以小规模验证掩盖了规模化后的成本曲线，是项目取消率居高不下的技术根源。

价值不清则是更普遍的管理成因。德勤 2026 年调研显示，55%的受访高管表示难以向利益相关方清晰证明 AI 项目价值；麦肯锡 State of AI 研究显示 62%的企业至少在做智能体试验，但仅 23%正在积极规模化。试点与规模化之间的断层，本质上是从“技术验证”到“业务验证”的门槛未跨过——试点证明的是模型能力，规模化需要证明的是单位经济。

1.2.3 从按席位收费到按结果收费的早期信号

部署率与价值交付之间的鸿沟，正在倒逼商业模式变革。IDC 数据显示，66%的中国企业倾向基于业务成果计费，远超全球平均水平；IDC 同时预测到 2028 年传统按席位收费模式将被淘汰，70%的软件供应商将转向按业务成果、交易量或自动化结果计费。国际市场上，Sierra、Salesforce、Zendesk 已完成按结果付费的产品化；国内市场上，蚂蚁数科、百融云创、零犀科技等厂商已跑通价值分成模式。按结果付费将供应商收益与客户业务产出绑定，直接回应了“部署率高但价值交付低”的核心矛盾，是跨越鸿沟最有力的商业机制。

1.2.4 不同行业与不同规模企业的部署节奏差异

部署节奏呈现清晰的行业梯度。赛迪顾问数据显示，政务、制造、能源、金融四大领域的智能体市场份额合计超过 70%——政策驱动型行业（政务）与数据密集型行业（金融）走在最前，制造与能源依托设备数据与流程标准化紧随其后。企业规模维度上，大型机构凭借数据基础与预算优势率先完成从试点到规模化的跨越，头部券商、国有大行与三甲医院是 2026 年的典型代表；中小企业则受制于数据准备度与实施成本，更多通过低代码平台与 SaaS 化产品低门槛接入。

这种节奏差异对供给方意味着市场策略必须分层：面向头部机构的深度定制与共建模式、面向中长尾的标准化产品与生态分发模式，二者不可混用。对政策制定者而言，节奏差异提示扶持政策应避免“一刀切”，对中小企业宜以算力补贴与共性平台降低接入门槛。

1.3 AI 智能体供给格局中的平台型、垂直型与自研型力量

1.3.1 头部科技企业的平台化路径与生态卡位

平台型厂商以头部科技企业为代表，通过提供智能体开发、编排、托管与治理的全栈平台占据生态入口。国内市场已形成清晰格局：字节跳动扣子、百度文心 AgentBuilder、阿里云百炼位列行业前三，腾讯元器、华为云 Versatile 等进入前十；平台竞争焦点正从“开发与编排”转向“全生命周期运营”——测试评估、运行托管、权限治理、日志审计与成本管理成为头部平台补齐能力的方向。平台型路径的核心优势在于生态卡位：智能体数量越多、互操作协议（如 MCP）越普及，平台的网络效应越强。IDC 预测基于低代码、无代码平台构建的智能体将长期占据活跃总量半数以上，印证了平台模式在数量维度的统治力。

平台化的价值分配优势同样显著。2026 年一季度交付或更新的企业应用中 80% 嵌入智能体，这些智能体大多运行在少数几个平台上；平台层截取了产业链中最稳定的“过路费”，这也是资本持续流向编排、监控与治理层的原因——2025 年纯智能体初创企业捕获 29 亿美元融资、集中在多智能体编排平台与工作流执行工具，投资者已将基础模型层视为既定格局，转而押注模型之上的部署价值。

1.3.2 垂直领域专业厂商的行业 Know-how 壁垒

垂直型厂商以行业 Know-how 与领域数据构筑壁垒，在专业壁垒高、容错率低的场景中反而具备更强议价权。法律领域的幂律智能是典型样本：其 MeCheck 合同审查智能体覆盖 600 余个合同模板类别、横跨 60 余个行业，2023 年 6 月完成 8000 万元融资（蓝驰领投），已服务超过 2000 家企业客户。垂直壁垒的本质不是模型能力，而是

三样东西的沉淀——标注过的领域语料、经过验证的行业审查规则库、以及与客户业务流程的深度耦合。这三样东西无法靠通用模型能力一蹴而就，构成垂直厂商的护城河。

垂直路线的风险在于场景天花板与平台挤压。当平台型厂商将行业模板、预置规则与领域微调做成开箱能力时，垂直厂商的浅层 Know-how 会被快速商品化；只有深至工作流重构、强至监管级可解释性的 Know-how 才具备持续溢价。这也是本报告在场景评估中将“落地深度”置于“市场规模”之前的原因。

1.3.3 需求方自研团队的边界与局限

需求方自研是 2025 年以来不可忽视的第三股力量。金融机构、大型制造企业与头部医院纷纷组建自研团队，将智能体能力内化。自研的动机有三：数据安全与合规要求（金融、医疗尤为突出）、业务 Know-how 的外部不可得性（券商研究方法论、医院临床路径）、以及成本长期化考量。国信证券非银团队观察到，券商 AI 路径已呈现清晰分化——少数头部机构自研大模型并开放生态（如国泰海通“君弘灵犀”），更多券商务实发力应用层。

但自研的边界同样清晰。其一，基础模型能力迭代速度远超单一企业的研发资源，自研团队最终仍需绑定外部基座模型，形成“伪自研、真集成”的结构；其二，自研的机会成本高，智能体工程涉及数据工程、评测体系、安全护栏等完整技术栈，非核心业务的机构难以长期投入；其三，人才竞争激烈，头部科技公司对 AI 工程人才的虹吸效应使传统机构自研团队面临流失风险。理性的自研边界是：把数据、评测与业务流程的掌控权留在内部，把基础模型与开发框架交给外部生态。

1.3.4 开源生态与闭源生态对供给格局的影响

开源与闭源生态的博弈正在重塑供给格局的成本曲线。开源侧，DeepSeek V4 Flash 将调用成本压至同类闭源模型的约九十五分之一，月之暗面 Kimi K3 在前端代码生成评测中登顶、成为首个登顶的开源模型，通义千问衍生系统占 Hugging Face 新发布语言模型 40%以上——开源模型在成本与可控性上的优势，使私有化部署的技术门槛与费用同步下移，这对政务、金融、医疗等强合规行业尤为关键。闭源侧，头部模型厂商持续以能力代差维持溢价，闭源模型在复杂推理、多模态与长程任务上仍保持领先。

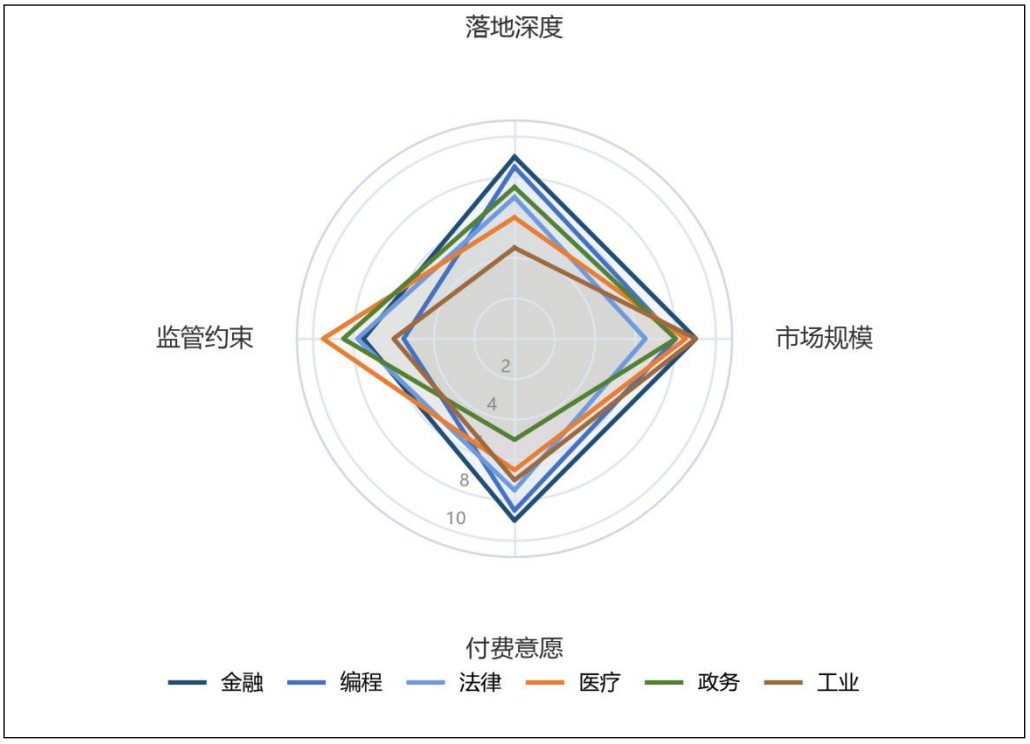
对供给格局的综合影响是：开源生态降低了智能体创业与自研的门槛，使供给主体更加多元；闭源生态则通过能力代差维持平台与模型层的话语权。两者叠加的结果是“模型层利润变薄、应用层与治理层价值上移”，九思行研判断这一趋势在 2026-2030 年将持续强化。

1.4 报告中 AI 智能体场景选择与评估框架

1.4.1 落地深度、市场规模、付费意愿与监管约束的四维评估

本报告以四个维度评估各场景的商业化成熟度。落地深度衡量智能体是否嵌入核心业务流程并产生可量化产出，权重最高；市场规模衡量场景的可寻址收入空间；付费意愿衡量需求方为结果付费的预算基础与决策机制；监管约束衡量合规要求对落地节奏的正反两方面作用——强监管既抬高门槛、也保护先发者。四维评分采用 1-10 分制，评分依据为公开数据与案例证据，属证据化推断而非精确测算。

图表 5 六大场景商业化成熟度四维评估

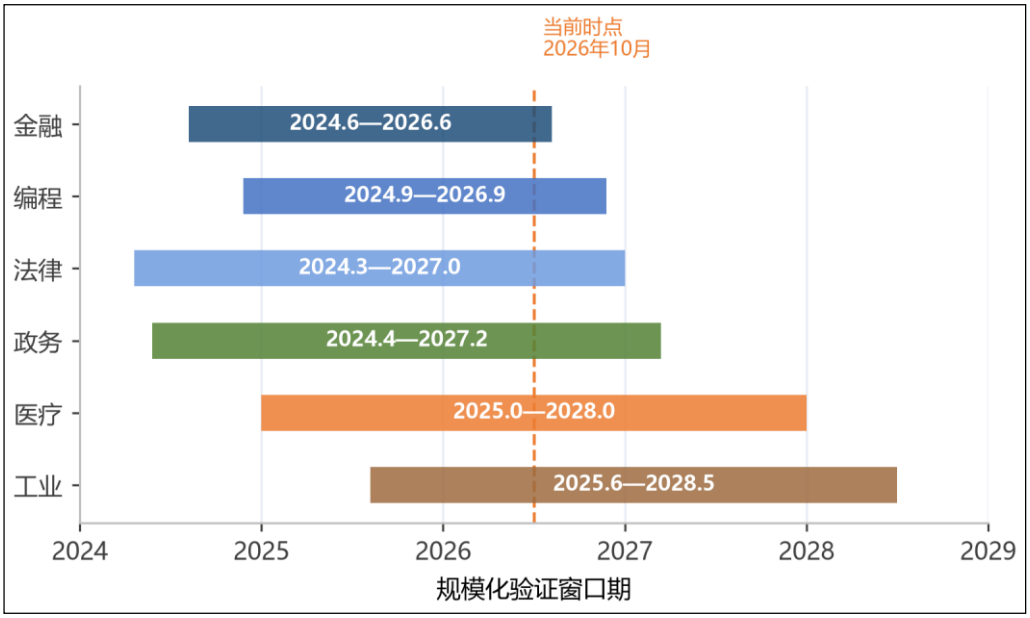


资料来源：九思行研基于公开案例与机构调研数据综合评估

1.4.2 金融、法律、医疗、编程、政务与工业的节奏图谱

六大场景呈现清晰的节奏梯度。金融与编程处于规模化验证的收尾阶段，商业闭环已基本跑通；法律处于渐进渗透的中段，单点效率验证完成、规模化复制进行中；政务在公共价值维度表现突出但商业模式以项目制与采购制为主；医疗受监管与数据成本约束、节奏后置但政策确定性增强；工业处于早期信号阶段，预测性维护单点 ROI 明确但系统级落地尚远。

图表 6　六大场景规模化验证窗口期



资料来源：九思行研综合评估

1.4.3　投资窗口的判断标准与数据来源说明

本报告的投资窗口判断基于三项标准：一是商业闭环验证进度，以部署率、续费率与单位经济模型衡量；二是价值分配位置，判断投资标所处的产业链环节是否具备持续议价权；三是确定性与弹性的平衡，确定性来自已验证的付费行为，弹性来自渗透率提升空间。数据来源方面，市场规模与企业采纳数据主要引用 IDC、Gartner、赛迪顾问、麦肯锡、毕马威、德勤等机构的公开研究；企业级市场规模基准数据（212 亿元/449 亿元/3320 亿元）引用科智咨询《AI 智能体赋能行业决策：趋势与实践白皮书（2026）》；企业落地数据引用公司年报、公开披露与权威媒体报道；政策数据引用国务院及各部委原文。所有关键数据均标注来源与统计时点，推断性结论均标明依据与成立条件。

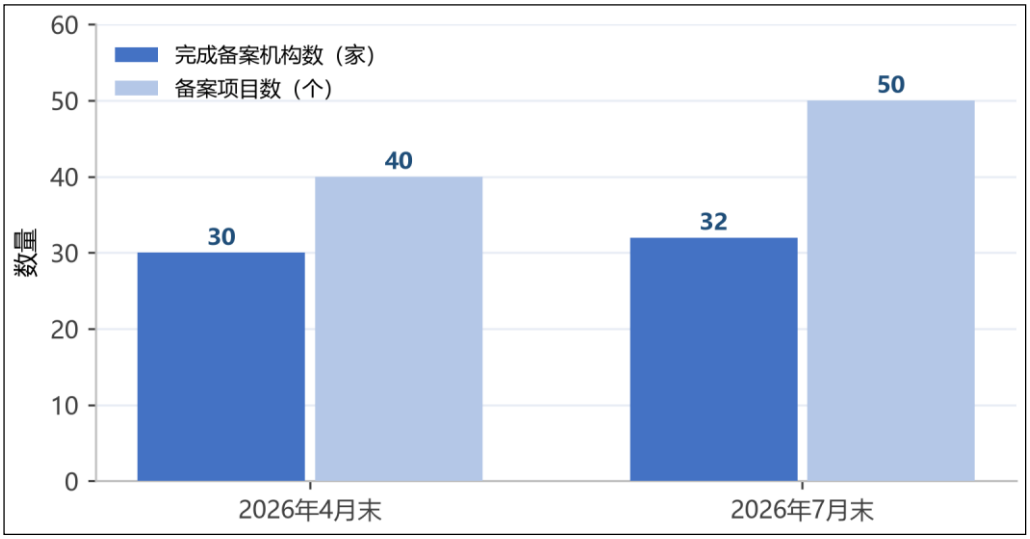
第二章　金融 AI 智能体商业化验证的领先赛道

2.1　金融 AI 智能体成为最快落地场景的结构性原因

2.1.1　合规与数据安全要求对落地加速的反向作用

直觉上，强监管行业的技术采纳应该更慢，但金融业的实践给出了相反的答案。合规与数据安全要求非但没有拖慢智能体落地，反而通过三个机制加速了进程。其一，合规要求筛选掉了不可靠的供给，留下愿意投入私有化部署、审计追踪与权限治理的厂商，使需求方的选型风险下降；其二，监管明确性降低了机构内部的决策争议——当算法备案、模型备案成为行业通行动作，业务部门引入智能体的合规路径变得清晰；其三，合规本身成为智能体的应用场景，合规审查、反洗钱监测、投研合规检查等高价值环节率先智能化。证券业的备案数据直接印证了这一进程：据易观千帆统计，截至 2026 年 4 月，证券行业已有近 30 家机构完成算法或模型备案、备案项目总数超 40 个；到 2026 年 7 月末，这一数字增长至 32 家机构、备案项目超 50 个，其中 10 余家头部券商完成多模型备案。

图表 7　2026 年证券行业算法与模型备案进展



数据来源：易观千帆，九思行研整理

2.1.2　业务标准化程度与付费能力的结构性优势

金融业具备智能体商业化最理想的结构条件。业务流程高度标准化——投研报告生成、信贷审批、财富配置建议等环节均有成熟的业务规则与输出规范，智能体的任务边界易于界定，结果的“对错”可核验，这直接降低了按结果付费的计量难度。付费能力上，金融业是中国 IT 投入强度最高的行业之一，头部金融机构的信息科技预算以十亿元计，且采购决策机制成熟，从试点到规模化的资金路径通畅。数据基础上，

金融业的数据治理起步早、质量高，结构化数据占比高，为智能体的可靠运行提供了供给基础。三重优势叠加，使金融成为智能体单位经济模型最先跑正的行业。

2.1.3 2026 年金融智能体元年的供需信号与机构动作

2026 年被业内普遍视为金融智能体的元年，供需两侧的信号均已明确。需求侧，2026 年上半年券商 AI 布局从“投入期”迈向“产出期”，重心由技术建设转向业务价值兑换，AI 智能体已快速渗透投研、投顾、运营、审核等核心业务场景，从“试验场”走向“主战场”。供给侧，头部券商密集发布产品：2026 年 4 月国泰海通率先推出灵犀 Skills，覆盖研报查询、行情数据、标的筛选等六大基础功能；2026 年 5 月广发证券、国信证券先后发布 Skill 矩阵；2026 年 6 月中金公司、华泰证券相继推出重磅产品。截至 2026 年 4 月，国内前十券商中已有半数落地 Skill 相关产品。供给的密集投放与需求的快速消化同时发生，是市场进入元年的典型特征。

2.1.4 金融 IT 预算结构与智能体采购模式的匹配关系

智能体采购与金融业既有 IT 预算结构高度匹配。金融机构的 IT 预算通常分为基础设施建设、应用系统开发与运维服务三大块，智能体产品恰好横跨三者：平台订阅计入基础设施与应用预算，定制开发计入项目预算，按结果付费部分则可从业务部门的费用预算列支——这使智能体采购摆脱了纯 IT 预算的约束，业务部门首次以“效果采购”的主体身份参与技术决策。IDC 数据显示 66%的中国企业倾向基于业务成果计费，金融业是这一倾向最强烈的行业；蚂蚁数科的实践显示，按效果付费模式反而促成客户签署 3 到 5 年长约，预算结构的匹配转化为商业关系的长期化。

2.2 金融 AI 智能体平台生态与落地规模

2.2.1 腾讯 WorkBuddy 金融版等平台型智能体的公开落地进展

平台型智能体在金融业的落地已进入实质阶段。以腾讯 WorkBuddy 为代表的办公与业务智能体平台，通过与金融机构的联合运营进入核心工作流；WorkBuddy、千问办公等办公智能体不再停留在功能演示层面，而是真正进入工作流、承担具体交付任务，国内市场对 AI 办公的接受度正从“尝鲜”转向“常态使用”，AI 办公进入实质性付费阶段。平台模式的金融落地路径呈现“外围到核心”的规律：先从会议纪要、文档处理、代码辅助等通用办公场景切入建立信任，再向投研辅助、合规审查等业务场景渗透。

2.2.2 头部券商、基金公司与银行的自研及合作模式差异

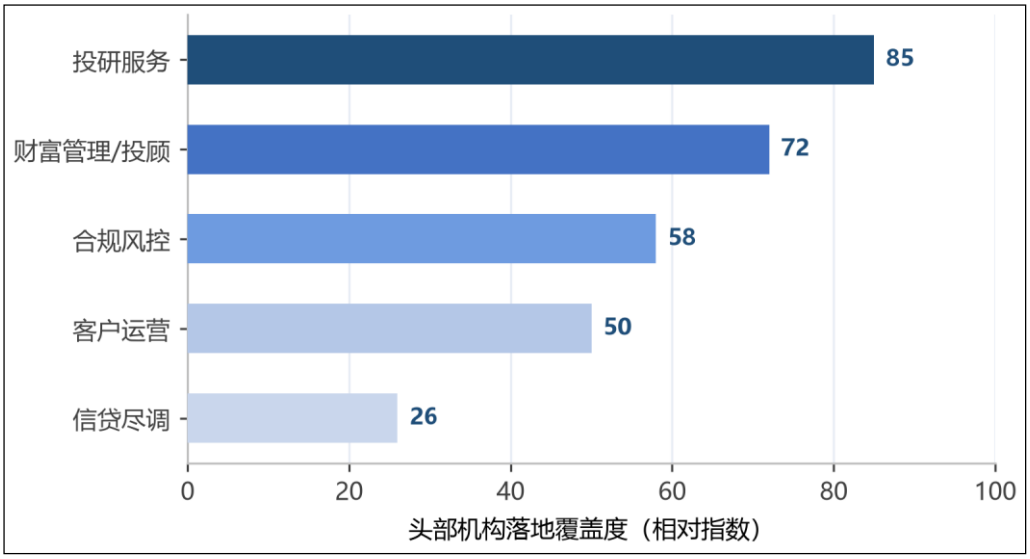
金融机构的智能体获取路径已分化为三种模式。自研模式以国泰海通为代表，其

自研大模型支撑“君弘灵犀”体系并向生态开放；华泰证券走“技术基建派”路线，整套体系深度依托自有投研数据底座与算力基建，降低对通用大模型的依赖，核心目标是搭建自主可控的 AI 技术底层架构。合作模式以中小机构为主，普遍与恒生聚源、同花顺等第三方技术服务商合作，快速上线标准化 Skill 产品，其中财富管理、投顾辅助类场景的落地需求最为迫切。混合模式以中金公司为代表——基座能力借助外部生态，但将自身最有价值的分析师研究方法论、估值体系与行业认知进行 AI“蒸馏”封装，形成“分析师.Skill”这样的差异化产品。三种模式没有绝对优劣，选择取决于机构的数据资产独特性、技术储备与战略定位。

2.2.3 投研、合规、信贷尽调与财富资配等核心场景的渗透顺序

金融智能体的场景渗透呈现清晰的顺序规律：先投研、后合规、再信贷、终财富。投研场景渗透最快，因其数据密度高、产出标准化、时效敏感——智能体的信息聚合与初步分析能力直接命中痛点；合规场景紧随其后，审计追踪需求与监管备案要求为智能体创造了刚需入口；信贷尽调场景对智能体的多源数据核验能力依赖度高、落地稍晚但单点价值大——金智维金融智能体在重庆银行等标杆案例中实现信贷审批时效缩短 87%、单笔录入时间从 210 分钟降至 15 分钟；财富资产配置场景涉及投资者适当性与信义义务，落地最为审慎，但国信证券等机构已将智能选股、ETF 筛选、基金对比、宏观数据查询等普惠投顾场景产品化。

图表 8 金融业智能体核心场景渗透顺序（相对指数）



注：上图中的“头部机构落地覆盖度（相对指数）”为九思行研基于公开案例的综合评估结果，构建方法为：对头部机构在各场景中是否有公开披露的智能体落地案例进行赋值（有明确案例=10 分，有试点信号=5 分，无公开信息=0 分），以各场景得分均值归一化至 0-100 区间，反映头部机构在不同场景中的相对落地密度而非绝对渗透率。

资料来源：九思行研基于公开案例综合评估

2.2.4 保险与银行客服场景的部署节奏比较

客服是金融业最早接触 AI 的场景，但保险与银行的智能化节奏存在明显差异。银行客服以咨询与业务办理为主，任务边界清晰、知识库结构化程度高，智能体自主解决率提升快；保险客服涉及条款解释、理赔指引等高专业性与高情感敏感度的内容，智能体多定位为辅助坐席而非自主应答。零犀科技的实践给出了保险场景智能体的另一种可能——不卖工具、直接交付销售结果：其 RaaS 模式帮助头部险企新增保费超 20 亿元，部分场景销售转化率高出行业平均 35%，达到 Top 销售员水平的 75%，证明保险智能体的价值锚点在销售转化而非客服降本。

2.2.5 中小金融机构的采纳障碍与替代路径

中小金融机构面临三重采纳障碍：自研能力不足、数据基础薄弱、采购预算有限。其替代路径已经清晰——加入行业级平台与生态分发。国泰海通将灵犀 Skills 上架华为小艺、腾讯 SkillHub 等第三方智能体平台，本质上是为行业提供了“能力批发”的基础设施；中小券商通过接入此类平台，以极低边际成本获得投研、投顾等高频技能。行业级基础设施的完善程度，将决定中小机构智能体化的速度，这也是监管推动行业共享算力与语料基础设施的政策逻辑所在。

2.3 金融 AI 智能体从辅助工具到数字员工的业务跃迁

2.3.1 华泰、国泰海通、中金与国信等券商的智能体部署路径

头部券商的部署路径提供了组织层面智能体化的完整样本。国泰海通的路径是“平台+生态”：自研灵犀体系，建设行业首家 AI 技能仓库，累计智能化服务截至 2026 年 6 月末达 8300 万人次，并通过第三方平台扩大触达。华泰证券的路径是“底座+双线”：依托自有投研数据底座，机构端“智研 Skill&MCP”与零售端“AI 涨乐 Skills”双线并行，覆盖机构投资者与零售客户。中金公司的路径是“方法论产品化”：将首席分析师团队的完整研究方法论、估值体系与行业认知封装为专属 AI Skills，机构客户可与分析师“数字分身”深度对话——这是对卖方研究服务模式的解构与重构。国信证券的路径是“普惠投顾”：六大 AI Skills 覆盖智能选股、ETF 筛选、基金对比、宏观数据查询等场景，将智能体能力下沉至长尾客户。

2.3.2 业务流程重构与组织变革的实证观察

智能体从辅助工具跃迁为数字员工，组织层面的变化比技术层面更深刻。实证观察显示三个组织变化正在发生：一是岗位结构变化，券商研究所的助理分析师岗位职能被智能体部分承接，人力向客户服务与方法论沉淀迁移；二是能力封装变化，资深从业者的隐性经验通过 Skill 封装变为可复用的组织资产，个人能力组织化的速度加

快；三是协作界面变化，业务人员与智能体的协作成为日常工作方式，AI 素养从加分项变为基本要求。金融业因知识密集与流程标准化兼备，成为观察这一组织跃迁的最佳样本。

2.3.3 增收、降本与风险控制三类商业模型变化

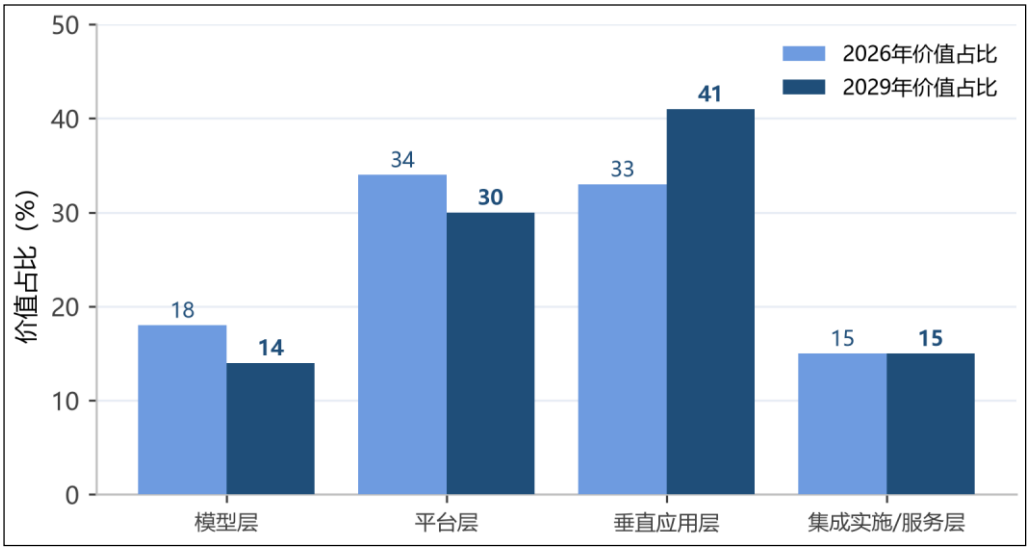
金融智能体的商业价值沿三条线展开。增收线：财富管理与投顾场景的智能体通过服务下沉扩大客户覆盖，国信证券的普惠投顾 Skill 即属此类；零犀科技的 RaaS 模式直接以保费增量计价，将增收逻辑推至极致。降本线：信贷、运营、客服场景的智能体替代重复人力，重庆银行案例中信贷审批时效缩短 87%是量化代表。风险控制线：合规监测、反欺诈、舆情预警等场景的智能体以降低风险损失计价，其价值虽难以直接观测，但在监管处罚趋严的背景下，风险控制类智能体的预算刚性反而最强。三类商业模型中，增收类对供应商的能力要求最高，也是按结果付费模式最先跑通的领域。

2.4 金融 AI 智能体投资窗口研判

2.4.1 平台层、垂直应用层与模型层的价值分配

金融智能体产业链的价值分配正在向应用层倾斜。模型层因开源冲击与同质化竞争，价值占比趋于下降；平台层凭借生态卡位保持稳定份额；垂直应用层则因行业 Know-how 的不可替代性，价值占比持续上升。九思行研测算，2026 年金融智能体价值分配中模型层、平台层、垂直应用层、集成服务层约为 18%、34%、33%、15%；到 2029 年，垂直应用层占比将升至 41%左右，模型层降至 14%。价值上移的直接含义是：具备场景纵深与客户关系的应用层厂商，其议价能力与利润空间将持续优于纯技术供应商。

图表 9　2026-2029 年金融智能体产业链价值分配及演变预测



注：上图中数据为九思行研基于产业链结构的测算模型输出。测算依据为：模型层参考开源模型调用成本占比及闭源模型授权收入规模；平台层参考头部智能体平台公开营收结构中平台订阅与工具服务收入占比；垂直应用层参考金融科技公司行业解决方案业务毛利率及项目均价；集成服务层参考 IT 服务商金融行业实施收入占比。上述测算未包含硬件与算力基础设施收入。

资料来源：九思行研测算

2.4.2　未来 2-3 年确定性最高的细分环节

综合部署数据、备案进度与付费行为，未来 2-3 年确定性最高的细分环节有三。第一是投研与投顾的智能体化——备案机构数与 Skill 产品数持续增长、头部机构已形成可复制的实施路径，且采购预算来自业务部门、决策链条短。第二是合规科技智能体——算法备案已成为行业通行动作，合规审查的智能化是监管鼓励方向，需求刚性最强。第三是金融数据智能体基础设施——数据治理、语料建设、评测体系是所有场景的前置工程，“卖水人”逻辑明确，IDC 数据显示 38%的企业将数据准备度不足列为落地首要障碍，该环节的供需缺口最大。

2.4.3　竞争拥挤度与差异化机会

金融智能体赛道的竞争拥挤度呈结构性分布。平台层已形成寡头格局，新进入者机会有限；投研、投顾等热门场景同质化竞争加剧，前十券商半数落地 Skill 产品即是信号；差异化机会存在于三处：一是垂直深度，如债券承销、并购重组等低频高价值场景的智能体化仍属空白；二是中小机构市场，行业级共享平台的运营与增值服务；三是按结果付费的计量与风控能力——谁能建立可信的结果核验机制，谁就能在商业模式代际切换中占据先机。

2.4.4 数据合规与模型可解释性对投资节奏的约束

金融智能体的投资节奏受两大约束。数据合规方面，金融机构数据高度敏感，智能体的数据调用边界、跨境流动与留存要求均受严格监管，私有化部署成为主流选择，这抬高了供应商的实施成本但也保护了本土厂商。模型可解释性方面，金融决策涉及投资者保护与系统性风险，监管对模型输出的可解释、可审计要求将持续提高，无法提供依据链的智能体产品将被主流市场排除。两项约束的净效应是：利好具备合规工程能力与行业积累的供应商，利好私有化部署路线，约束纯云端 SaaS 模式的扩张速度。

第三章　法律 AI 智能体在专业壁垒下的渐进渗透

3.1　法律 AI 智能体在“错不起”行业中的智能化逻辑

3.1.1　合同审查成为第一个规模化场景的原因

法律行业被认为是“错不起”的行业，其智能化路径与金融、编程截然不同——它不会出现爆发式替代，只能渐进渗透。合同审查之所以成为第一个规模化的场景，有三个结构性原因。第一，合同文本介于结构化与半结构化之间，付款条款、违约责任、保密义务、争议管辖等风险点相对可枚举，可以抽象成规则库与审查清单，任务边界清晰。第二，合同审查是高频、重复、劳动密集的工作，人工审查一份复杂合同平均需要约 4 小时，成本痛点足够大，AI 替代的收益容易量化。第三，审查结果由律师复核把关，human-in-the-loop 机制天然存在，为 AI 的错误输出提供了安全垫，降低了试错成本。幂律智能披露的数据显示，其 MeCheck 智能体预置 300 余个审查点，覆盖货物买卖、商铺租赁、营销活动等合同类型，综合准确率超 95%，2026 年第一季度累计审查合同超 100 万份、发现约 350 万个风险点——规模化的证据已经充分。

3.1.2　人机回圈作为行业刚需而非过渡方案

在多数行业，人机回圈（human-in-the-loop）被视为 AI 成熟前的过渡方案；在法律行业，它是 permanence 的行业刚需。其原因在于责任结构：律师对法律意见承担执业责任，AI 无法承担责任主体资格，因此最终判断权必须保留在人。2026 年司法部启动 AI 法律服务合规试点，要求 AI 法律工具必须标注“仅供参考，不构成法律意见”，重大案件必须由人类律师复核——监管以制度形式确认了人机边界。这一定位对商业模式的含义深远：法律智能体的定价逻辑不是替代人力成本，而是放大单位人力的产出，其市场空间按“律师生产力工具”而非“律师替代”计算，天花板取决于渗透率而非存量人数。

3.1.3　法律数据高壁垒与低容错对商业化节奏的影响

法律数据的获取与加工壁垒显著高于通用领域。高质量的法律语料分散在律所卷宗、企业法务系统与司法数据库中，涉密属性强、脱敏成本高；法律知识的更新频率高——法规修订、司法解释与裁判口径变化要求规则库持续维护，这构成了持续的知识工程投入。低容错则意味着产品必须提供依据链：律师要的不是一句“建议修改”，而是“为什么这条标红、依据哪条规则”，可解释性是法律 AI 的生存底线。高壁垒与低容错共同决定了法律智能体的商业化节奏是渐进的：单点验证快（合同审查）、全流程替代慢（诉讼代理），市场将长期呈“点状突破、面上渐进”格局。

3.1.4 律所与公司法务部门的需求差异

律所与公司法务是法律智能体的两类核心买方，需求差异显著。律所的核心诉求是提升律师单位产出与标准化交付质量——批量审查与条款比对、法律意见书初稿生成等场景需求突出，iCourt 的 AlphaGPT 系统自动检索判例并生成法律意见书初稿，文书起草效率提升约 40 倍（传统需 2 小时压缩至 3 分钟）。公司法务的核心诉求是合规刚性落地与法务职能转型——将内部合规红线接入智能审查系统后，合规要求从“依赖个人经验”变为“系统刚性执行”，法务部门从成本中心走向价值中心。两类需求的付费逻辑不同：律所为效率付费、对成本敏感，公司法务为风控付费、对确定性敏感；后者天然更适配按结果付费模式。

3.2 法律 AI 智能体落地数据与真实效率提升

3.2.1 幂律智能等厂商公开披露的客户规模与合同审查数据

法律智能体头部厂商的公开数据提供了行业最扎实的落地证据。幂律智能 2023 年完成近 8000 万元 Pre-B 轮融资（蓝驰创投领投），为其迄今公开的最大单笔融资，也是中国法律科技赛道迄今最大的单笔融资，已服务超过 2000 家企业客户，包括字节跳动、美团等头部企业；其 MeCheck 合同审查系统与 MeFlow 合同全生命周期管理系统构成产品双线。同业对比方面，法大大披露其 iTerms 产品 AI 合同审查准确率超过 95%，其电子签名加 AI 合同审查组合在房地产、金融、人力资源行业渗透率超 30%；百度“AI 律师”在标准租赁合同审查中可识别 90%的常见风险点，速度是人工律师的 50 倍，其 C 端产品“法行宝”截至 2025 年 9 月已累计为超过 7000 万用户提供法律咨询服务，日均处理量超过 8 万次。厂商数据的密集披露本身即是市场升温的信号，但引用时须注意口径差异。

3.2.2 审查时间与风险发现率等指标的口径问题及可比性挑战

法律 AI 的效率指标存在显著的口径差异，跨厂商比较需极为谨慎。以审查准确率为例：法大大声称其 AI 合同审查准确率超 95%，而幂律智能 MeCheck 综合准确率同样超 95%——两个数字的统计对象不同（识别准确率与审查准确率）、风险点定义不同、测试集不同，数字本身“更像一把需要校准的尺子，而非放之四海皆准的标尺”。风险漏检率的口径差异更大：幂律智能披露其将风险漏检率从人工审查的约 8%降至 0.5%，但漏检率的分母取决于风险点清单的完备性，不同厂商的清单差异可能高达数倍。

图表 10 合同审查 AI 辅助前后的核心指标对比（幂律智能口径）

	人工/传统模式	AI 辅助模式
审查耗时	4 小时	15 分钟

风险漏检率	8.0%	0.5%

数据来源：幂律智能公开披露，九思行研整理

3.2.3 跨法域与跨行业合同模板差异对规模化的实际影响

合同模板的多样性是法律智能体规模化的最大技术摩擦。不同法域、不同行业的合同结构差异巨大——幂律智能处理的模板类别超过 600 种、横跨 60 余个行业，每进入一个新行业都需要重建或适配审查规则库。这意味着法律智能体的规模扩张不是纯软件复制，而是“软件复制+知识工程”的复合过程，边际成本高于通用 SaaS。对厂商而言，可行的策略是聚焦少数高价值行业做深规则库（如幂律智能对建设工程、知识产权、劳动用工、投融资的行业积累），而非追求全行业覆盖；对买方而言，选型时应重点考察厂商在本行业的规则库深度而非通用能力宣传。

3.2.4 效率提升向收入增长转化的实证观察

效率提升能否转化为商业收入，是检验法律智能体商业闭环的最终标准。实证证据正在累积：幂律智能披露的标杆案例显示，某城商行引入智能审查后合同审核时间缩短约 60%，预计年节约人力成本 30%、减少法律纠纷损失 40%；某央企制造板块将内部合规红线接入系统后，合同审查耗时从小时级压缩至分钟级，法务职能从成本中心走向价值中心。收入侧，2026 年第一季度全球法律科技融资总额达 23.4 亿美元（103 笔交易），同比增长 25.4%；其中 Harvey 和 Legora 两家头部公司在半个月内合计融资 7.5 亿美元，资本集中度显著提升。效率到收入的转化链条已经打通，但转化效率因客户规模与行业而异，大型机构转化快、中小机构仍需教育。

3.3 法律 AI 智能体推动律所端组织变革

3.3.1 大成与惠诚等律所的 AI 部署路径及战略合作

律所端的 AI 部署呈现“头部自建、腰部合作、长尾接入”的梯度。头部律所依托案件数据与品牌优势，与法律科技厂商共建定制化系统，将事务所知识沉淀为私有规则库；腰部律所主要通过采购标准化产品快速获得 AI 能力；长尾律师则通过通用平台与 C 端工具接入。部署路径的选择本质上取决于数据资产规模与技术预算，行业分层因此加深——头部律所的 AI 能力差距正在转化为服务报价与客户获取的竞争优势，法律服务的马太效应因 AI 而放大。

3.3.2 技能与智能体可复用化的组织含义

法律智能体的深层组织含义在于律师能力的可复用化。传统律所的知识管理依赖师徒制与案例库，隐性经验难以规模化传承；智能体架构下，资深律师的审查规则、

文书范式与谈判策略可封装为可复用的 Skill，新律师调用 Skill 即继承组织经验。这一变化对律所的竞争格局影响深远：能力封装能力强的律所，其新人成长周期缩短、服务质量方差收窄，可以承接更大规模的标准化业务；反之，未能完成能力组织化的律所将在价格与效率双重压力下边缘化。律所的核心资产正从“合伙人个人声誉”向“组织级智能体资产池”扩展。

3.3.3 法律服务成本结构与竞争格局的变化

AI 渗透正在重塑法律服务的成本结构与竞争格局。成本端，合同审查、尽调初筛、文书起草等标准化环节的边际成本大幅下降，法律服务定价从“按小时”向“按事项”“按结果”迁移的压力增强；竞争端，标准化服务的价格竞争加剧，倒逼律所向高价值非标业务（诉讼策略、交易设计、合规架构）上移。对法律科技厂商而言，格局变化创造了两类机会：向律所提供能力封装工具的“赋能型”机会，与直接面向企业法务提供审查服务的“替代型”机会。后者增长更快——企业法务市场的付费主体明确、决策链条短，是法律 AI 商业化最顺畅的通道。

3.4 法律 AI 智能体投资窗口研判

3.4.1 垂直厂商与通用平台之间的边界

法律智能体领域，垂直厂商与通用平台的边界正在动态演变。通用大模型厂商向法律场景的渗透集中在信息检索、文书初稿等浅层场景；垂直厂商的护城河在审查规则库、依据链可解释性与私有化交付能力。边界的稳定条件是：只要法律场景的容错成本足够高、监管对可解释性要求足够强，垂直厂商的专业壁垒就能维持；一旦通用模型的可解释性与领域知识突破临界点，浅层垂直产品将被平台吸收。九思行研判断，合同审查、合规管理两个深水场景的垂直壁垒在未来 2-3 年仍然稳固，法律检索、普法咨询等浅层场景将快速被通用平台商品化。

3.4.2 领域语料与行业 Know-how 的估值权重

资本对法律 AI 的估值逻辑已从“模型能力”转向“数据资产”。幂律智能近 8000 万元 Pre-B 轮融资、2026 年第一季度法律科技全行业融资 23.4 亿美元（同比增长 25.4%）背后，投资机构普遍将领域语料规模、规则库完备度与客户工作流耦合深度作为核心估值参数。海外对照更加明显：法律 AI 公司 Harvey 拥有超过 3000 家客户，全美百大律所超 80%已使用其产品，2026 年 9 月完成 5.5 亿美元融资后估值达 155 亿美元，年经常性收入已超过 4 亿美元——多年期合同背后是客户工作流的深度嵌入与数据资产的持续沉淀。领域语料与 Know-how 已取代算法先进性，成为法律 AI 估值的第一权重。

3.4.3 2026-2028 年商业化节奏的关键节点

法律智能体未来三年的商业化节点可按三个层次预判。2026 年：合同审查场景完成头部企业渗透，按结果付费模式在企业法务市场出现规模化合同；司法部合规试点的细则落地，为 AI 法律工具的责任边界提供更明确的制度框架。2027 年：审查场景向尽调、合规监测等相邻场景扩展，律所端能力封装平台成为新的竞争焦点；预计行业出现首轮整合，缺乏规则库深度的厂商被收购或出清。2028 年：企业法务智能体渗透率进入双位数区间，法律服务定价模式的迁移从讨论变为实践，按事项与按结果计费在标准化服务中占比过半。上述节点为证据化推断，其成立条件是监管框架保持“鼓励应用、明确边界”的基调不发生逆转。

3.4.4 法律科技监管变化对投资窗口的影响

监管是法律 AI 投资窗口的最大变量。当前监管基调呈现“双轨”特征：一手推动应用——司法部启动 AI 法律服务合规试点，鼓励 AI 辅助司法与法律服务；一手划定边界——要求 AI 法律工具标注“仅供参考，不构成法律意见”、重大案件必须人类律师复核。这一基调对投资窗口的影响是结构性的：合规试点的扩大直接扩大可商业化场景清单，责任边界的明确降低买方的采购顾虑；但若未来监管收紧责任认定（如要求厂商对 AI 输出承担连带责任），将显著抬高厂商合规成本、压缩利润空间。投资者应将监管试点进展作为法律 AI 赛道的第一优先观察指标。

第四章　医疗 AI 智能体在监管红线下的价值锚点

4.1　医疗 AI 智能体的“辅助”定位与落地边界

4.1.1　三类医疗器械与药监局注册证的约束机制

医疗是所有场景中监管约束最强的一个，其约束的核心机制是第三类医疗器械注册证制度。基于深度学习的辅助诊断软件按第三类医疗器械管理，须通过国家药监局（NMPA）的严格审评——截至 2025 年 10 月，NMPA 已批准超过 110 款基于深度学习的第三类医疗器械独立软件产品，数据类型包括 CT 图像、X 射线图像、眼底图像等 9 种类型。三类证的意义是双向的：它既是高门槛——一款三类证的取证周期以年计、临床验证成本以千万计，将大量玩家挡在门外；也是信任锚——持证产品获得了医院采购与医保支付的准入资格。博睿康 NEO-ONE SCI 植入式脑机接口系统于 2026 年 3 月获 NMPA 批准上市，成为全球首个获批的植入式脑机接口三类医疗器械，华山医院作为临床试验牵头单位完成全球首例植入手术，显示监管审批正在向大模型与脑机接口等新形态延伸，监管红线的位置决定着医疗智能体的产品形态与商业化边界。

4.1.2　诊断权归属与责任边界对产品设计的影响

诊断权归属是医疗智能体产品设计的首要约束。现行制度下，诊断结论必须由执业医师作出并署名，AI 的定位是“辅助”——国家卫生健康委等五部门 2025 年 11 月印发的《关于促进和规范“人工智能+医疗卫生”应用发展的实施意见》明确坚持“人工智能赋能而不替代”的定位。这一定位直接影响产品架构：智能体的输出必须附带依据与置信度、必须嵌入医师复核环节、必须保留完整审计轨迹。2026 年 3 月，国内 40 余家机构联合发布《医疗机构人工智能应用与治理专家共识（2026 版）》（以下简称“《共识》“），提出“人机协同、以人为主”的核心原则、“数据不出院”原则与 AI 系统熔断机制——行业治理框架的成型，意味着产品设计自此有了明确的合规基准，“以人为主”不是过渡状态而是长期制度安排。

4.1.3　医院采购与科室预算逻辑的现实约束

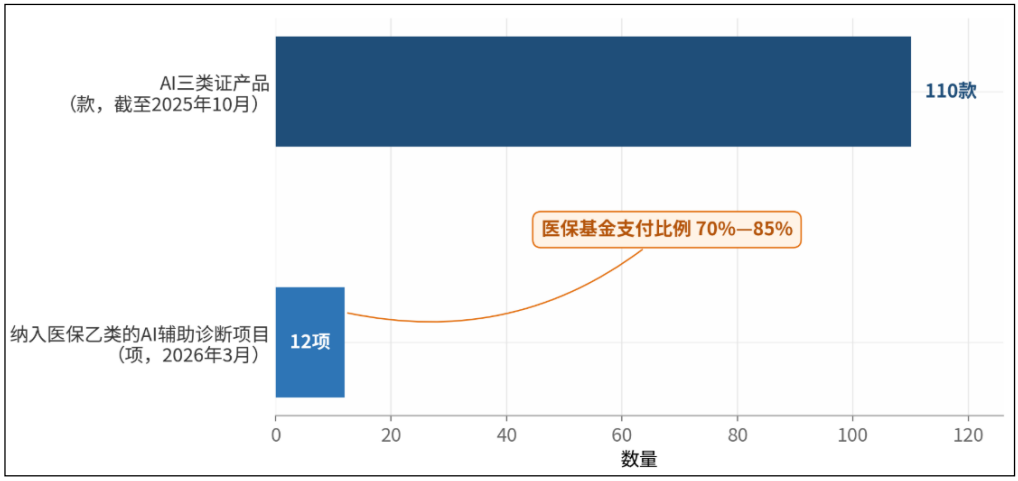
医院采购逻辑对智能体商业化的约束常被低估。公立医院实行预算制管理，设备与软件采购须纳入年度预算并履行招标程序，采购周期以年计；科室层面，信息科预算与临床科室预算的分割使智能体采购常陷入“临床要用、信息科付钱”的错位。破局的关键在于将智能体纳入既有付费通道——2026 年 3 月国家医保局将 12 项 AI 辅助诊断项目纳入医保乙类目录，医保基金支付比例 70%-85%，覆盖肺结节 CT 辅助筛查、眼底病变识别、心电图自动分析等高频场景，这一破冰使医院采购 AI 辅助诊断从“纯成本项”变为“有收入对应的服务项”，采购逻辑被根本性理顺。医保支付的示

范效应预计将逐步扩展至更多 AI 医疗场景，是未来三年医疗智能体商业化最重要的制度变量。

4.1.4 医生使用习惯与工作流嵌入的摩擦成本

医生的处方式工作流对智能体的嵌入精度要求极高。临床场景中，医生的时间以分钟计，任何增加操作步骤的产品都会被弃用；电子病历系统的操作授权、院内网络的安全隔离、与 HIS/PACS 系统的对接，都构成嵌入摩擦。成功案例的共同经验是“无感嵌入”——上海中试基地的全院级电子病历生成智能体、胸部影像一扫多查智能体均以“原有流程内自动触发”为设计原则，后者累计处理病例超 250 万例、覆盖全国 20 余家医院，证明嵌入深度决定使用深度。摩擦成本的另一个侧面是责任心理：医生对 AI 错误的容忍度远低于其他行业用户，产品必须以“初筛+提示”而非“结论+确认”的交互姿态建立信任。

图表 11 医疗 AI 监管准入进展



数据来源：国家药监局、国家医保局，九思行研整理

4.2 医疗 AI 智能体已落地案例的结构性分析

4.2.1 协和、地坛与河北医大一院等综合智能体案例

综合型医院智能体的代表性案例揭示了两种建设路径。一种是国家项目中试基地模式：国家人工智能应用中试基地（医疗领域）依托复旦大学附属中山医院建设，从 200 余款医学 AI 应用中遴选 51 项优质成果开展中试推广，2026 年 3 月发布五大核心创新成果与九款医疗智能应用，包括“观心”心血管病辅助诊疗智能体——其依托超 180 万例病例及 300TB 多模态语料，自 2025 年 6 月上线以来已累计处理 2 万余例病例，大幅压缩病历书写、术前谈话、多学科会诊筹备时间。另一种是医院自建模式：北京协和、北京地坛医院、河北医科大学第一医院等综合医院依托自身临床数据与专

家资源建设院内智能体，覆盖病历生成、预后评估、患者服务等全院场景。两种路径的对比显示：中试基地模式以标准化推广见长，医院自建模式以场景贴合度取胜，未来将长期并存。

4.2.2 仁济医院泌尿专科等专科智能体的规模化服务数据

专科智能体的服务数据提供了医疗 AI 规模化最直接的证据。上海交通大学医学院附属仁济医院泌尿外科智能体累计服务超 80 万人次，其中 65%的咨询来自三线及以下城市，是专科智能体服务量的标杆；上海中试基地的肝胆肿瘤智能体已推广至 122 家医疗机构、服务超百万人次，胸部影像一扫多查智能体累计处理病例超 250 万例；浙江大学医学院附属邵逸夫医院的肝癌多模态 AI 大模型智能体于 2026 年 6 月上线，首创“开放算力、共享模型、按需结算”的 AI Token 跨院全流程结算服务，上线以来 AI 服务调用总量已突破 25 亿 Token。专科智能体的规模化证据链已经完整，其共性规律是：选择病例量大、影像依赖度高、诊断标准相对明确的专科（影像、肝胆、泌尿），规模化速度最快。

图表 12 代表性医疗智能体服务规模对比

智能体类型	服务规模
胸部影像一扫多查智能体	累计处理病例超 250 万例
肝胆肿瘤智能体	累计服务超 100 万人次
仁济泌尿外科智能体	累计服务超 80 万人次
中山“观心”心血管病辅助诊疗智能体	累计处理 2 万余例病例

数据来源：国家人工智能应用中试基地（医疗领域）、各医院公开披露，九思行研整理

4.2.3 临床决策、病历生成与预后评估的场景分化

医疗智能体的内部场景呈现明显的分化。病历生成场景落地最快——上海中试基地的全院级电子病历生成智能体已在多家医院应用，该场景不涉及诊断决策、监管风险低、医生书写负担重的痛点明确，是“低风险高粘性”的入口场景。临床决策辅助场景价值最高但节奏最慢——受三类证与诊断权约束，落地形态以“影像初筛+医师复核”为主。预后评估与健康管理场景介于两者之间——上海基地多款智能体向患者服务延伸，国家政策层面“探索推广人人可享的高水平居民健康助手”的表述为该场景打开了政策空间。三者的商业含义不同：病历生成按效率付费、临床决策按准确性付费、健康管理按服务订阅付费，厂商须按场景选择对应的商业模式。

4.2.4 区域医疗中心与基层医院部署模式的差异

医疗智能体的部署呈“中心辐射”结构。区域医疗中心作为能力中枢，承担模型训练、语料建设与专家规则沉淀；基层医院通过区域平台以极低成本接入智能体服务。

国家政策明确“突出基层”导向——《关于促进和规范“人工智能+医疗卫生”应用发展的实施意见》把更多篇幅放在基层和预防，以人工智能支撑预防、诊疗、康复、健康管理全链条集成服务，大幅提高基层医疗健康服务能力和效率。区域化部署的经济学逻辑清晰：单个基层医院无力承担智能体建设成本，但区域中心的一次性投入可通过服务分摊覆盖数十家基层机构。上海中试基地与 28 家医院、20 家领军企业的生态签约，正是这一模式的组织化呈现。

4.2.5 案例数据公开程度与可复制性评估

医疗智能体案例数据的公开程度总体高于其他行业，但可复制性评估须审慎。公开数据集中在服务量（人次、Token 调用量、覆盖机构数）层面，而单位经济数据（单次服务成本、采购价格、回款周期）公开极少。可复制性方面，中试基地模式因有标准化推广机制，复制性最强——肝胆肿瘤智能体推广至 122 家机构即为例证；医院自建案例的技术与流程高度依赖本院资源禀赋，直接复制难度大。评估一个医疗智能体案例能否复制，应看三项条件：是否有标准化产品形态、是否有区域平台承接能力、是否有医保或财政支付通道。

4.3 医疗 AI 智能体数据成本与商业化瓶颈

4.3.1 脱敏、授权与专家标注成本居高不下的原因

医疗数据成本是智能体商业化的最大瓶颈。其成本高企有三重原因：一是脱敏成本，医疗数据涉及个人敏感信息，《共识》确立“数据不出院”原则后，脱敏与本地化处理成为强制要求，技术与合规成本叠加；二是授权成本，临床数据用于模型训练须经伦理审查与患者知情授权，流程周期长、法律风险高；三是标注成本，医学标注必须由副高以上职称专家参与，单位标注成本远高于通用数据。三重成本使医疗智能体的前期投入显著高于其他场景，也解释了为何“开放算力、共享模型、按需结算”的共享模式在医疗领域具有特殊价值——分摊是消化数据成本的最现实路径。

4.3.2 医保支付改革对新付费场景的创造作用

医保支付改革正在为医疗智能体创造全新的付费场景。2026 年 3 月 31 日，国家医保局将 12 项 AI 辅助诊断项目纳入全国统一医保乙类目录，覆盖癌症早筛、慢性病管理和急诊急救三大方向，这是 AI 医疗服务首次获得全国性医保支付地位。医保纳入的连锁反应是结构性的：医院端，AI 辅助诊断从成本项变为收入项，采购意愿被激活；厂商端，支付标准的确立使按次计费有了权威锚点，商业模式的确定性大幅提升；产业端，医保目录的动态调整机制为后续 AI 医疗项目进入打开了通道。预计未来三年医保支付范围将沿“筛查－诊断－康复－健康管理”链条逐步扩展，每一个新增支

付项目都会创造一个规模化的付费场景。

4.3.3 规模化速度与投资回报周期的现实匹配

医疗智能体的投资回报周期显著长于金融与编程场景。其原因是三重周期的叠加：产品周期（三类证取证以年计）、采购周期（医院预算与招标以年计）、回款周期（公立医院结算延迟）。三者叠加使厂商的收入确认严重后置，现金流压力远高于 SaaS 行业常态。现实的匹配策略有三：一是以病历生成、患者服务等低门槛场景的现金流养高门槛的诊断类产品；二是绑定区域医疗中心与中试基地，以“一次建设、多院分摊”摊薄成本；三是优先进入已获医保支付的场景，以支付确定性对冲回款周期。资本进入该赛道须以长周期资金匹配，短期回报预期不适用。

4.4 医疗 AI 智能体投资窗口研判

4.4.1 长周期背景下政策确定性增强的逻辑

医疗智能体的投资逻辑正在从“高风险”向“高确定性长周期”转变。政策确定性来自三个层面：顶层设计——国务院“人工智能+”行动意见明确要求有序推动人工智能在辅助诊疗、健康管理、医保服务等场景的应用；行业细则——《关于促进和规范“人工智能+医疗卫生”应用发展的实施意见》确立 2027 年、2030 年两阶段目标，覆盖 8 个方向 24 项重点应用；支付保障——医保乙类目录的破冰打通了最重要的支付闭环。三层政策以法律与预算形式固化，医疗智能体由此获得了其他场景不具备的“制度确定性”。长周期属性依然存在，但确定性增强意味着风险溢价的下降，适合以产业资本与耐心资本为主的投资结构。

4.4.2 影像、病历与辅助诊断等细分方向的优先级

医疗智能体细分方向的优先级排序，应综合支付保障、监管成熟度与复制性三要素。第一优先级是影像辅助诊断：12 项医保支付项目中多数为影像类（肺结节 CT、眼底、心电图），支付通道最畅通，且胸部影像智能体 250 万例的处理量证明复制性已验证。第二优先级是病历与文书智能化：监管风险最低、医生痛点最明确、部署门槛最轻，是现金流最好的入口场景。第三优先级是专科辅助诊断：单点价值高但取证与推广周期长，适合产业资本长线布局。健康管理与服务机器人方向政策空间大，但支付闭环尚未形成，暂列观察级。

4.4.3 平台型与专科型厂商的竞争格局

医疗智能体的竞争格局呈“平台型守入口、专科型占纵深”的态势。平台型厂商（医疗信息化企业、云厂商）以 HIS/PACS 入口与全院集成能力守卫入口层，其优势

是客户关系与渠道纵深；专科型厂商以单一病种或单一影像模态的深度积累获取溢价，其优势是专家资源与注册证壁垒。两种类型的竞合关系大于对抗——平台型需要专科产品丰富生态，专科型需要平台渠道触达医院。上海中试基地的生态签约模式（28 家医院、20 家企业）预示着“平台组织、专科供能”的产业协作结构将成主流，单纯的全栈型厂商面临两头挤压。

4.4.4 医院回款周期与商业化可持续性

医院回款周期是医疗智能体商业化可持续性的最大考验。公立医院对软件与服务的结算普遍存在延迟，叠加医保支付的结算周期，厂商应收账款周转天数显著高于其他行业。缓解路径有三：优先选择医保支付覆盖的项目，医保结算的确定性高于医院自付；以订阅制替代项目制，平滑收入曲线；与医疗信息化厂商合作，借助其既有结算通道。对投资者的建议是：评估医疗智能体标的时，应收账款质量与回款条款应作为与收入增速同等重要的尽调项目。

第五章　编程 AI 智能体与其他高价值场景的商业成熟度

5.1　编程 AI 智能体渗透率最高的技术适配性

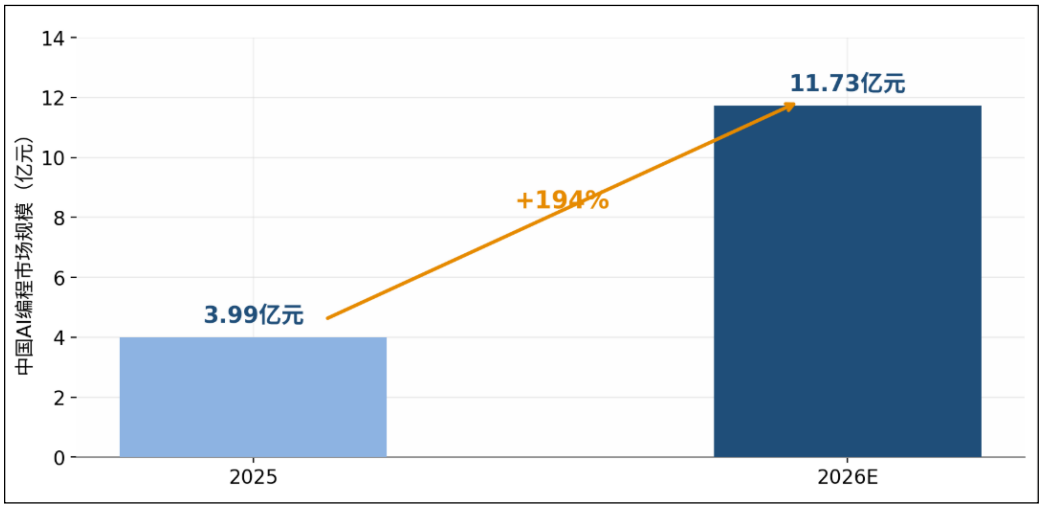
5.1.1　代码结构化、可验证与反馈闭环短的技术优势

编程是智能体渗透率最高的场景，其根本原因在于技术适配性完美。第一，代码是结构化语言，语法严格、语义明确，大模型的理解与生成误差远低于自然语言场景。第二，代码可验证——单元测试、编译检查与持续集成构成机器可执行的验证体系，智能体的输出正确性可以自动判定，无需人工逐项核对，这使“自主闭环”在技术上成为可能。第三，反馈闭环短——开发者从提交任务到获得可运行代码的周期以分钟计，快速迭代天然适配智能体的强化学习逻辑。三项技术优势使编程成为智能体能力演进的最前沿阵地，其他场景的智能体形态（金融的 Skill 封装、法律的审查规则库）在方法上都借用了编程场景的验证范式。

5.1.2　开发者付费意愿与价值量化的商业优势

编程场景的商业优势在于价值量化极其容易。开发者的产出以代码交付量与缺陷率衡量，智能体的贡献可以直接度量——麦肯锡 2026 年初报告显示，采用 AI 编程的团队开发效率平均提升 35%-55%，超过一半的新增代码已有 AI 参与生成。付费意愿因此强劲：Gartner 预测到 2027 年底超过 70%的企业开发团队将把 AI 编程助手列为标准工具；IDC 数据显示 2025 年中国 AI 编程市场规模为 3.99 亿元，预计 2026 年底达 11.73 亿元，同比增长约 194%。企业的采购逻辑已经从“要不要用”转变为“用哪家”，这是商业成熟度最高的表现形式。

图表 13　2025-2026 年中国 AI 编程市场规模



数据来源：IDC，九思行研整理

5.1.3 从补全工具到自主智能体的能力跃迁

编程智能体正在经历从补全工具到自主智能体的能力跃迁。补全工具阶段（2022-2024 年），产品形态是 IDE 内的行级补全，人是任务的主导者；对话协作阶段（2024-2025 年），产品以对话方式理解需求、生成函数级代码，人机协作成型；自主智能体阶段（2025 年至今），产品可以接受模糊需求描述、自主拆解任务、调用工具、执行多文件修改并交付完整功能模块。阿里 Qoder 从 AI IDE 升级为智能体自主开发工作台、字节 Trae 的 Agent 化演进、OpenAI Codex 的重写升级，都标志着行业已进入自主智能体阶段。能力跃迁直接反映在使用深度上：据阿里 Qoder 1.0 发布时的实测数据，用户代码保留率提升 11%、输入 Token 消耗降低 40%、对话轮次减少 33%。

5.1.4 编程智能体对软件工程流程的重塑

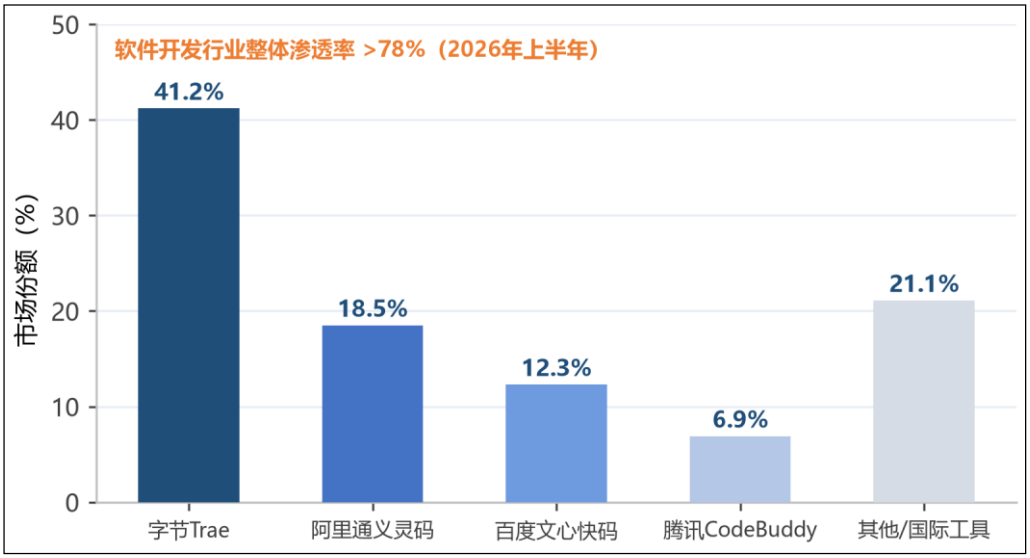
编程智能体对软件工程的影响已从工具层面深入流程层面。需求环节，自然语言描述可以直接转化为原型代码，需求评审与开发实现的边界模糊化；开发环节，代码生成、测试用例编写、代码审查等子任务大规模智能化，工程师的角色向需求定义与架构设计上移；运维环节，智能体开始承担故障定位、日志分析与修复建议生成。流程重塑的组织后果是软件工程的人力结构变化：初级编码岗位需求收缩、具备 AI 协作能力的工程师溢价上升。麦肯锡数据显示超过一半的新增代码已有 AI 参与生成，软件开发正在从“人力密集”转向“智能体密集+人力杠杆”的新范式。

5.2 编程 AI 智能体企业级部署的规模化信号

5.2.1 字节 TRAE 与腾讯 CodeBuddy 等内部渗透数据

国产编程智能体的规模化数据在 2026 年密集披露。渗透率维度，截至 2026 年 4 月，AI 编程工具在中国软件开发行业的渗透率已超过 78%。市场份额呈现“使用率与收入份额双口径”的特征：IDE 使用率口径下，字节 Trae 以 41.2%居首、阿里通义灵码 18.5%居次、百度文心快码 12.3%居三；收入份额口径下（IDC 2026 年 7 月报告），阿里 Qoder 以 47.6%居首、智谱 CodeGeeX 11.5%、商汤 Raccoon 10.5%、腾讯 CodeBuddy 6.9%、百度 Comate 6.0%。IDC 报告同时确认，阿里 Qoder 的收入份额超过第二名至第五名的总和。两个口径的背离揭示了中国市场的关键特征：编程智能体对大厂而言不是直接收入来源，而是云计算的获客入口——字节以免费策略换用户规模，阿里以企业订阅与云生态绑定获取收入，战略选择不同导致口径分化。

图表 14　2026 年中国 AI 编程工具市场份额（使用率口径）



数据来源：九思行研整理

5.2.2　OpenAI Codex 用户增长与 tokens 处理量的公开信号

海外编程智能体的公开信号提供了能力演进的参照系。Cursor 付费用户超过 100 万、年经常性收入约 20 亿美元；Claude Code 周活跃用户约 420 万，以 46%的开发者喜爱度居首；OpenAI Codex 已追至 Cursor 约 60%的使用量，其模型家族迭代至多档上下文共享架构。增长信号背后的结构性变化是：编程智能体成为基础模型厂商最重要的企业级入口——模型能力通过编程场景验证、企业习惯通过编程场景培养、算力消耗通过编程场景变现，编程是 AI 商业化的“第一性场景”。

5.2.3　软件开发组织形态与外包市场的影响

编程智能体对软件外包市场的影响正在显现。外包行业的定价基础是人力工时，智能体使单位工时的代码产出大幅提升，按人天计价的传统模式面临重估——具备 AI 能力的外包商可以更低报价交付同等质量，行业将出现“效率红利再分配”，智能化领先的外包商扩大份额、落后者退出。组织形态上，软件团队的最小作业单元从“若干工程师”缩小为“工程师+智能体编队”，小微团队可以承接过去需要中型团队的项目。对中国软件产业而言，编程智能体既是外包行业的效率革命，也是软件出海的竞争力重构机会——以智能体杠杆对冲人力成本优势的流失。

5.2.4　企业级安全与代码资产保护的实际要求

企业级部署对编程智能体提出了安全与代码资产保护的硬要求。代码是企业的核心知识产权，金融、政务、运营商等行业明确要求代码不出境，私有化部署与专有云

方案成为采购必备条件；智能体的上下文读取范围、生成代码的许可证合规、训练数据的中性化，均须纳入企业安全评估。百度文心快码主攻金融政企与央国企客户、以合规和私有化部署为核心卖点，正是对这一要求的响应；卓易信息 EazyDevelop 等厂商亦以私有化部署与全链路审计合规为差异化定位。安全要求提高了企业级市场的进入门槛，但同时也保护了具备合规能力的本土厂商。

5.3 公共价值驱动下的政务 AI 智能体特殊赛道

5.3.1 北京 12345“京京”、深圳福田与台州检察院等案例

政务智能体的标杆案例已形成完整的证据链。北京 12345 热线在“京通”App 上线智能体“京京”，支持语音双向交互，实现诉求信息自动采集、工单智能生成与精准派发；据《北京 12345 市民服务热线 2025 年度数据分析报告》，热线 2025 年全年受理诉求 2580.2 万件，诉求解决率和满意率分别达到 97.3%和 97.7%；自 2019 年接诉即办改革以来累计受理诉求已超 1.8 亿件。深圳福田区在全国率先推出 AI 数智员工，据深圳市政府及广东省发改委公开信息，43 类智能体已全面上岗：“AI 小福”实现重复填报减少 50%，民生诉求助手累计处置社情民意事件超 202 万件，仲裁数智员工将裁决书撰写从半天压缩至几分钟。检察机关方面，台州检察院等地的智能辅助办案系统将证据审查、文书生成等环节智能化，台州检察已上线“公益诉讼线索采集研判助手”等智能体，构建“智能体初筛－人工复核－系统迭代”的工作闭环，形成“人主机辅”的办案新格局，检察办案效率显著提升。

5.3.2 政府直采、国企平台运营与科技企业共建的采购模式

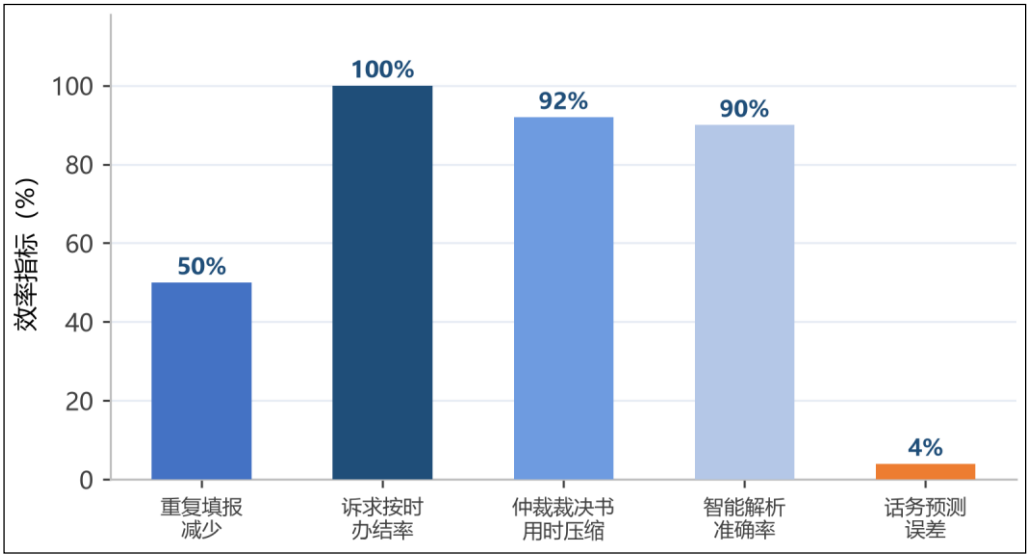
政务智能体的采购模式呈三种形态并进。政府直采模式适用于技术成熟的标准品，通过政府采购程序直接采购智能体产品与服务，效率高但定制性有限。国企平台运营模式以地方国资平台为建设运营主体，云天励飞于 2026 年 3 月中标湛江市 AI 渗透支撑新质生产力基础设施建设项目，中标金额 4.2 亿元，基于自研国产 AI 推理加速卡建设国产 AI 推理千卡集群，搭载 DeepSeek 等国产大模型，构建“国模国芯”生态。科技企业共建模式适用于前沿场景，政府开放场景与数据、企业投入技术与运营，深圳福田“场景超市”累计开放场景需求清单超 1000 个，以真实需求吸引优质 AI 企业，是共建模式的组织化样本。三种模式的共同约束是安全红线：深圳福田区实现 100%使用国产大模型、100%搭载国产算力、全部部署在专属安全政务网络内，自主可控是政务智能体的先决条件。

5.3.3 商业可持续性与效率指标的衡量方式

政务智能体的价值衡量必须兼顾效率指标与公共价值。效率维度，可量化的指标

体系已经形成——工单处理时长、重复填报压缩率、办结率、智能应答准确率、预测误差率等，福田区“重复填报减少 50%、按时办结率 100%、满意度提升”的量化披露为行业建立了衡量基准。公共价值维度，政务智能体的深层价值在于治理能力提升：北京 12345 的“每月一题”机制从海量工单中筛选高频共性民生痛点，累计完成 2000 余项任务、出台 500 余项政策，选取 70 余个高频共性难题开展专项治理，实现从处置一件事到治理一类事。商业可持续性方面，政务智能体以财政预算与国资平台投入为主，市场规模稳定但毛利率受制于采购定价机制，厂商的可持续策略是“以政务场景打磨产品、以行业化复制实现盈利”。

图表 15　政务智能体核心效率指标



数据来源：北京市政务服务和数据管理局、广东省人民政府网、台州市人民检察院公开信息，九思行研整理

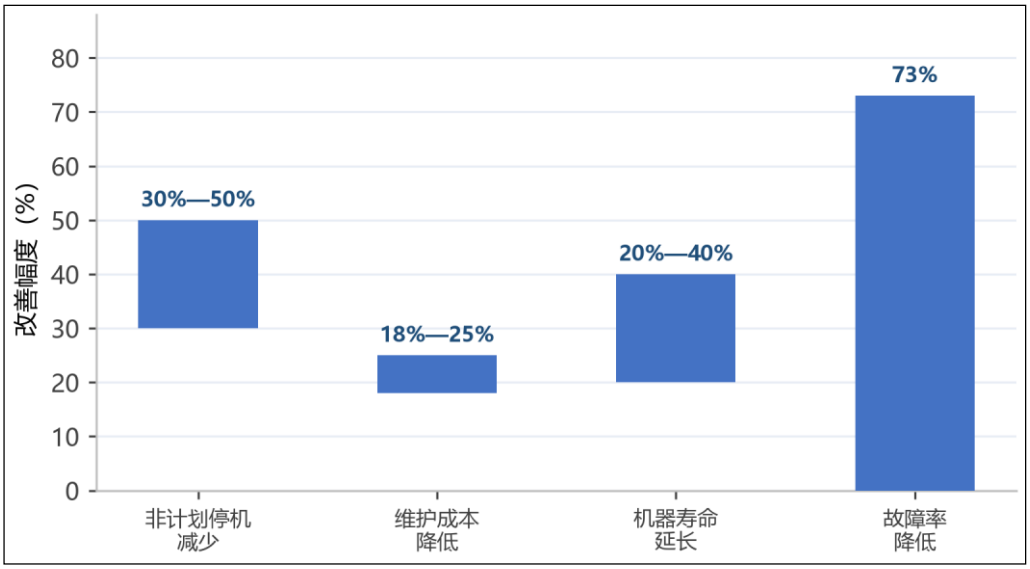
5.4　工业与制造 AI 智能体的早期信号

5.4.1　设备故障响应与预测性维护的落地数据

工业智能体的落地信号集中在预测性维护场景，其投资回报已被充分验证。技术层面，预测性维护模型可在故障发生前 30-90 天预警，多数来源显示准确率区间为 80%-97%。经济层面，麦肯锡分析显示预测性维护可实现非计划停机减少 30%-50%、维护成本降低 18%-25%、机器寿命延长 20%-40%，全球制造业年价值潜力达 6300 亿美元；领先采用者 12-18 个月内实现 10：1 至 30：1 的投资回报，单厂节约 150 万-750 万美元。停机成本的刚性支撑了付费意愿——离散制造业非计划停机平均损失约 26 万美元/小时，汽车装配线停工损失超 230 万美元/小时。国内落地方面，卡奥斯 COSMOPlat“AI 设备故障诊断与智能维保”场景 2026 年 8 月入选山东省首批“人工智能+制造”重点建设应用场景，实现设备数据在线采集、健康工况实时研判、隐患超前预判与维

保工单自动派发。

图表 16　工业预测性维护的核心改善幅度

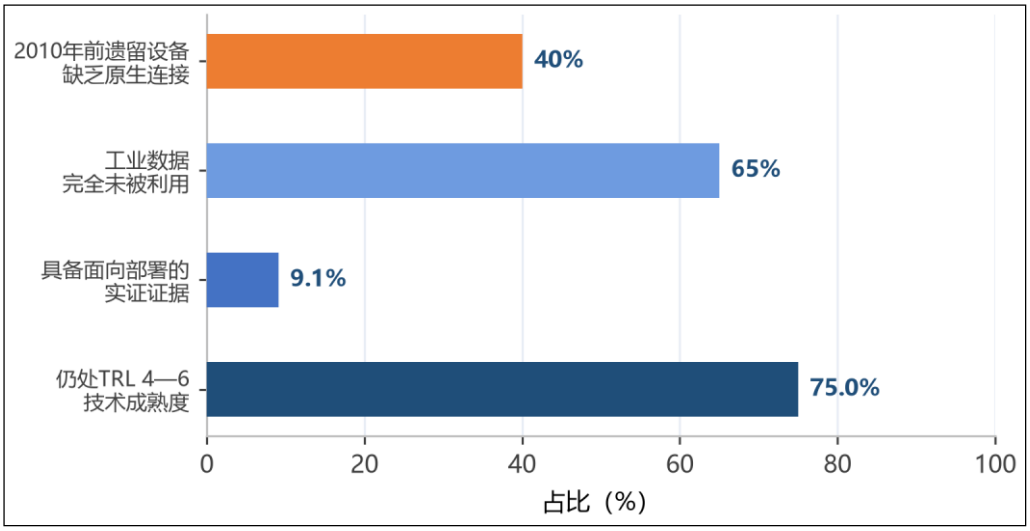


数据来源：麦肯锡、f7i.ai，九思行研整理

5.4.2　工业智能体目标下的部署节奏与行业差异

工业智能体的部署节奏呈现显著的行业与成熟度分层。行业整体仍处早期——一项 2026 年发表的系统性综述筛查 2341 篇文献发现，75.0%的已报道基础模型工业智能体系统仍处 TRL 4-6（技术验证至环境验证）阶段，仅 9.1%的案例具备面向部署的实证证据。行业差异明显：流程工业（化工、冶金）因设备价值高、停机损失大，预测性维护渗透最快；离散制造（汽车、电子）在质量检测与装配辅助环节加速导入；半导体制造因良率敏感成为最前沿——东京电子于 2026 年 7 月宣布与英伟达扩大合作，在 Epsira 数字化转型解决方案中导入 NVIDIA NeMo、NIM 等智能体工具及 Isaac 机器人平台，推动 AI 故障排除、数字孪生与智能维护。企业采纳率方面，德勤预测到 2026 年制造业中 Agentic AI 的采用率将从 6%增长到 24%，仍处爬坡初期。

图表 17　工业智能体的成熟度与数据利用缺口



数据来源：Value Add VC，九思行研整理

5.4.3　与金融、编程场景在商业化逻辑上的根本区别

工业智能体与金融、编程场景在商业化逻辑上存在根本区别。价值形态上，金融与编程智能体的产出是数字资产（报告、代码），交付即完成；工业智能体的产出是物理世界的状态改变（设备启停、工单执行），必须通过控制系统的安全验证，责任链条远长于纯数字场景。集成复杂度上，工业现场存在大量遗留设备，超过 60%的工业设备仍处于“数据孤岛”状态、缺乏数字化接口，超过 65%的制造企业设备联网率不到 30%，改造费用是采纳的实际瓶颈；相比之下，金融与编程场景的数据天然数字化。安全约束上，NSA 与 CISA 于 2025 年 12 月发布的 AI-in-OT 指南明确要求仅在收益大于风险时集成 AI、关键决策保留人类、实施失效安全机制——工业智能体的天花板由安全边界而非模型能力决定。三项区别意味着工业智能体的商业化是长周期、重服务、强集成的“持久战”，不适合追求快速回报的资本。

5.4.4　工业数据孤岛与 OT 系统集成难题

工业数据孤岛是比模型能力更根本的制约。工厂内部数据横跨物联网传感器、PLC、ERP、MES、HMI、OPC-UA 消息与检修报告等多种来源，语义标准不一，数据治理论文已警告语义错配可能传导为服务编排错误与知识推理失败。NIST 在 2025 年 11 月半导体制造研讨会报告中指出，将制造数据视为二等资产将系统性削弱 AI、机器学习与数字孪生系统的效能。破局路径是“先治理、后智能”：领先企业的实践表明，数据基础设施（设备连接、数据管道、质量保证）必须先行于智能体部署，德勤调研亦显示数据基础设施最成熟的资产类别恰是智能体落地最快的类别。对产业方而言，OT/IT 融合咨询与数据治理服务本身就是工业智能体产业链中最先受益的环节。

5.5 跨场景 AI 智能体投资窗口研判

5.5.1 编程赛道商业化最确定的方向

编程赛道未来 2-3 年最确定的方向有三。其一是企业级智能体平台：企业采购从单点工具转向平台化部署，具备代码资产保护、私有化与 DevOps 集成能力的平台将获取最大份额，阿里 Qoder 的收入份额优势（47.6%，据 IDC 2026 年 7 月报告）已验证该逻辑。其二是垂直行业编程智能体：金融、政务等强合规行业的专用编程智能体，以行业代码库、合规规则与私有模型为壁垒，百度文心快码在金融政企市场的卡位是先例。其三是软件工程流程再造服务：智能体嵌入研发流程后的流程咨询、效能度量与工具链整合，是“卖铲子”型机会，麦肯锡数据显示 AI 编程带来的效率提升 35%-55%需要组织流程变革才能兑现，变革服务的市场将随渗透率同步扩张。

5.5.2 政务与工业场景的公共价值及长周期回报

政务与工业智能体的投资逻辑不同于商业场景，其回报形态是“公共价值+长周期商业回报”的复合。政务智能体的公共价值已在效率指标中量化体现（北京 12345、深圳福田数据），其商业回报路径依赖行业化复制——头部厂商以标杆案例为模板向地市级、区县级市场复制，政府采购的预算刚性提供了收入下限。工业智能体的回报锚定在可测算的停机损失削减（单厂 150 万-750 万美元/年），其长周期属性来自数据基础设施先行 requirement，投资窗口在数据治理与 OT 集成层先于应用层打开。两类场景均适合以产业资本、国资背景基金与耐心资本主导的投资结构。

5.5.3 跨场景可复用的平台能力与工具链机会

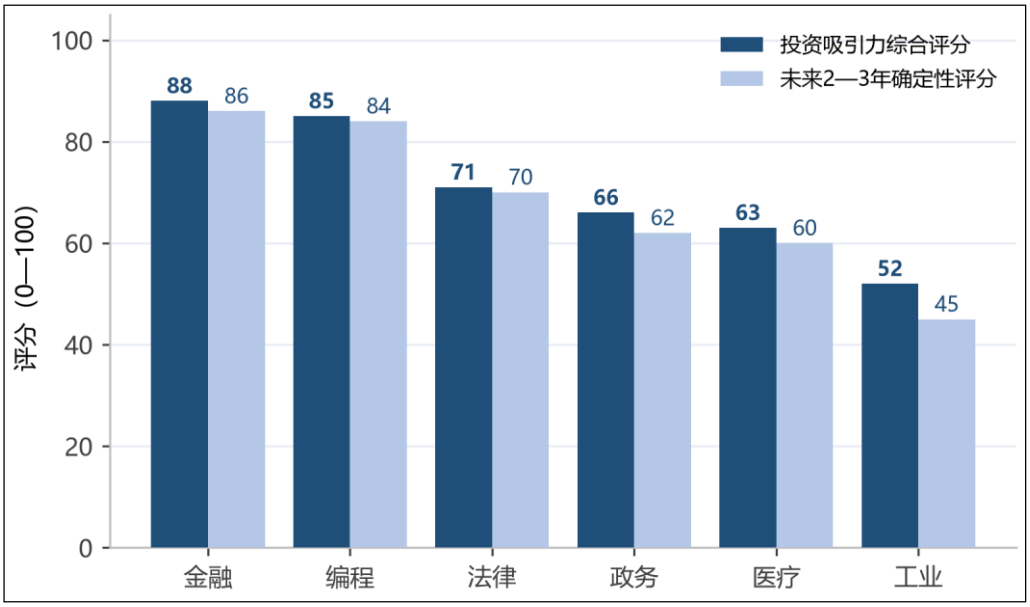
跨场景观察揭示了一个被低估的投资主线：可复用的平台能力与工具链。无论金融 Skill 封装、法律规则库还是工业知识图谱，各垂直场景的智能体工程都依赖四类共性能力——领域数据治理、评测与验证体系、安全护栏、智能体编排。这四类能力的供应商天然具备跨场景扩张性，其市场空间是各垂直场景之和。海外资本流向印证了这一判断：2025 年智能体领域投资正加速流向编排、监控与治理层工具，投资者将基础模型层视为既定格局、转而押注模型之上的部署价值。国内对应环节（智能体评测、安全护栏、MCP 工具生态）尚处早期，是未来 2-3 年最具弹性的平台型机会。

5.5.4 垂直场景之间的投资组合与风险分散

基于六大场景的成熟度与风险特征，本报告建议的投资组合结构为“核心+卫星”。核心仓位配置金融与编程两个已验证场景——部署率、续费率与单位经济三项指标均向好，确定性与流动性最佳；卫星仓位配置法律与政务——法律的单点经济模型已验证、渗透率提升空间大，政务的预算刚性与政策确定性提供防御属性；医疗与工业作

为期权仓位——医疗以医保支付扩展节奏为触发条件分步加仓，工业以数据基础设施成熟度为先行指标。风险分散的关键在于避免同一时点的同质化暴露：六大场景的风险来源各异（金融看合规、编程看技术迭代、法律看监管、医疗看支付、政务看预算、工业看集成），组合的自然分散度较高，投资者应避免在单一环节（如纯模型层）过度集中。

图表 18　六大场景投资吸引力与确定性综合评分



资料来源：九思行研基于公开数据综合评估

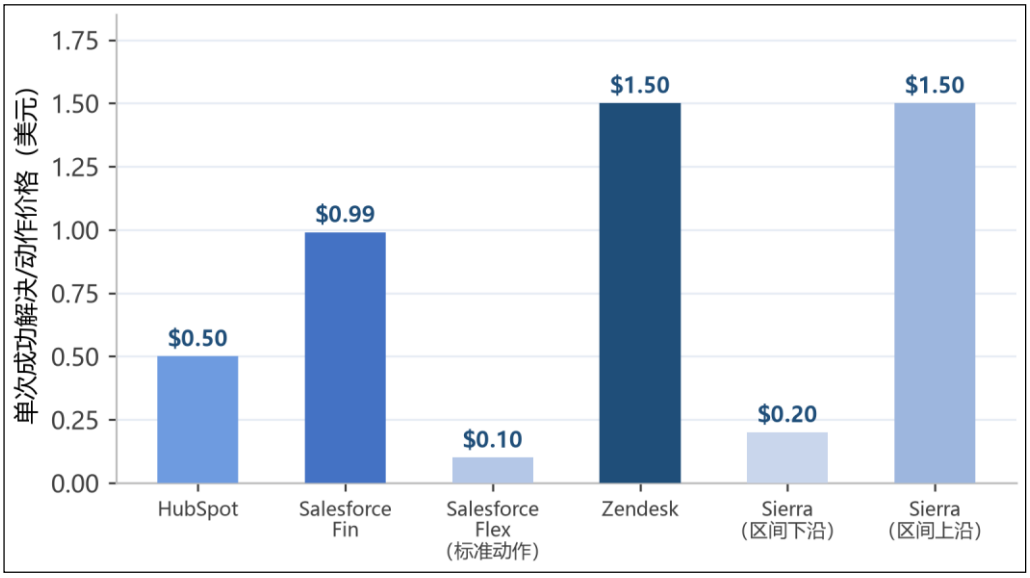
第六章　AI 智能体商业模式、投资窗口与风险研判

6.1　AI 智能体从卖工具到卖结果的商业模式代际切换

6.1.1　按结果付费的全球实践与 Salesforce、Zendesk、Sierra、OpenAI 案例

按结果付费（outcome-based pricing）正在从概念变为真实的合同条款，全球先行者的实践已形成完整样本库。Sierra 是激进派代表：这家由 OpenAI 董事长 Bret Taylor 创办的 AI 客服企业，于 2026 年 5 月完成 9.5 亿美元融资，投后估值超过 150 亿美元（由 Tiger Global 与 GV 联合领投），ARR 于 2026 年 2 月升至 1.5 亿美元，服务超过 40%的财富 50 强企业；Sierra 的 ARR 在 2025 年 11 月为 1 亿美元，至 2026 年 2 月增长至 1.5 亿美元，三个月增长 50%；其定价采用“基础订阅+效果佣金”混合模式，10 万美元年费保障基础接入，按自主解决会话 0.2-1.5 美元计费，转人工不计费——风险从买方挪到了卖方，供应商必须先证明自己能交付。Salesforce 跟进迅速：2026 年 6 月宣布以约 36 亿美元收购 Fin（前身为 Intercom），Fin 按 0.99 美元/成功解决单独计费；2026 年 6 月发布 Agentforce Help Agent，仅在 AI 代理从头到尾自主解决问题时收费，每单 2 美元，转人工或负面反馈不计费，产品于 2026 年 7 月上线；Zendesk 推出经独立评估模型核验结果真实有效才计费的“AI 员工”体系；连麦肯锡都在逐步抛弃按工时计费，转向收益与客户经营利润挂钩。甲骨文在 2026 财年第四季度（截至 5 月 31 日）财报电话会披露，启动 token bundles 与成果计费的“有限推广（limited rollout）”，33 家客户在该季度预购了 token bundles（包括 Aon、Liberty Energy），与此同时 Oracle 也在向 outcome-based pricing 扩展。

图表 19　主要厂商按结果付费定价对比



数据来源：各公司公开披露，九思行研整理

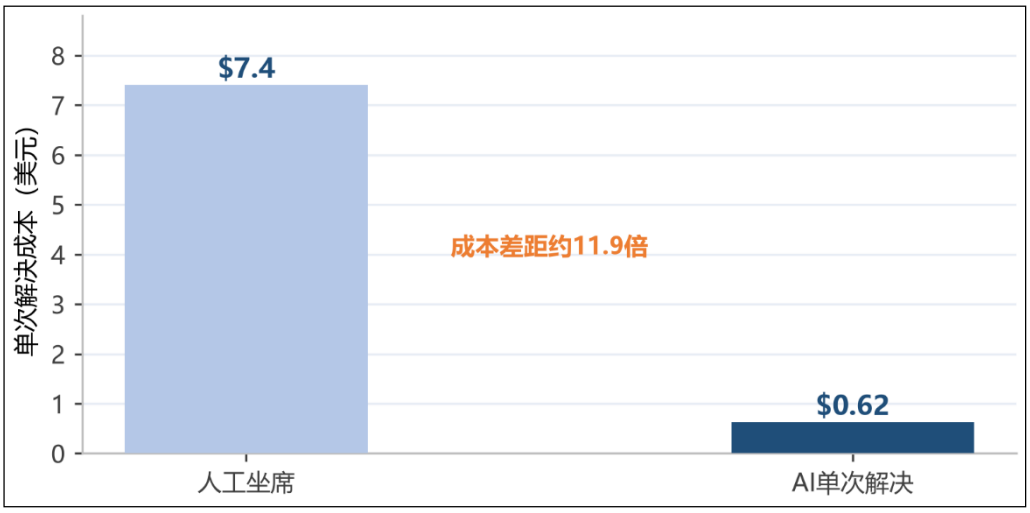
6.1.2 中国市场的智能体即服务与价值分成模式

中国市场的按结果付费实践（业内称 RaaS，Results-as-a-Service）已形成本土化形态。零犀科技早在 2019 年即坚定押注 RaaS 模式，不售卖软件订阅、只交付销售业务成果，2025 年实现规模化盈利与正现金流——其帮助某头部保险机构新增保费超 20 亿元，“这 20 亿保费增收，依靠传统人机结合模式，需要 800-1000 人的销售团队才能实现”；其转化率达专家水平的 75%。蚂蚁数科 2025 年 9 月推出按效果付费：上海银行 AI 手机银行转化率提升 10%，理财代销按千分之二到千分之四分成，部分客户首年收益全归蚂蚁，客户反而愿意签署 3 到 5 年长约——收益分成的深度绑定转化为商业关系的长期化。百融云创发布 ResultsCloud，提供按任务、按岗位薪酬、按价值分成三种模式；天润云 AI 数字劳动力收入一季度同比增长 105%，只为成功办结工单付费；用友网络 2026 年 8 月明确提出“软件及标准服务+业务成果分成”的新付费方式。中国 RaaS 的演进路径与国际同向，但在金融销售、数字劳动力等场景的渗透深度已具领先性。

6.1.3 对单位经济模型与收入质量的实际影响

按结果付费对供应商的单位经济模型与收入质量产生双重影响。正面影响是收入天花板的上移：传统订阅模式按席位或按调用量收费，收入上限由客户人数决定；按结果付费直接分享客户业务产出，当智能体承担“数字员工”职能时，其收入对标的是人力成本而非软件预算——这是一个数量级的差异，垂直 AI 的市场空间按“知识型工作的劳动成本”而非“软件用户数”计算，Bessemer 的研究认为垂直 AI 的市场容量可达同领域传统 SaaS 的十倍。负面影响是收入波动性上升：结果付费使供应商承担了业务不确定性，收入与客户业务量同步波动；同时 AI 单次解决成本 0.62 美元与人工 7.40 美元的成本差距意味着供应商的毛利空间取决于自身推理成本的持续优化，成本控制能力弱化的厂商将被模式反噬。

图表 20　AI 与人工单次解决成本对比



数据来源：麦肯锡，九思行研整理

6.1.4　按结果付费的计量争议与合同设计难点

按结果付费的落地障碍集中在计量与合同设计两端。计量端的核心争议是“什么算结果”：客服场景以“会话自主解决”为计费事件，但解决与否的判定可由供应商自评（存在道德风险）或独立模型核验（Zendesk 的 verified resolution 机制）或客户确认（增加摩擦）；更隐蔽的问题是委托－代理扭曲——供应商有动机最大化计费事件数量而非结果质量，Forbes 的分析将其称为“企业 AI 最昂贵的神话”。合同端的难点是风险定价：供应商承担交付风险后，必然要求风险溢价，长期合同的价格谈判复杂度远超固定订阅；多要素业务（如理财代销的转化率受市场行情影响）的结果归因难以在合同中事先穷尽。可行的制度解法包括第三方计量审计、基线对赌（以历史数据为基线约定改善幅度）与混合定价（低固定费+高分成），国内蚂蚁数科的多年长约实践证明混合定价在金融场景的可行性。

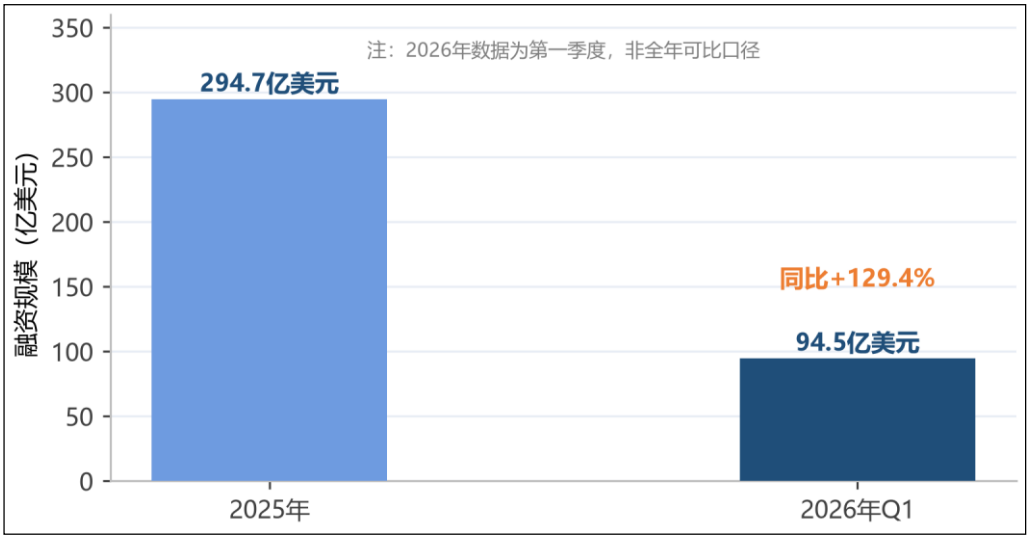
6.2　AI 智能体资本流向与投资优先级

6.2.1　代理式 AI 融资规模与增速的权威数据

代理式 AI 已成为全球风险投资最集中的方向之一。毕马威数据显示，2025 年中国 AI 领域 PE/VC 投资总额达 406.4 亿美元，同比增长 58.5%，交易数量 2336 笔、同比增长 54%；其中代理式 AI 融资规模 2025 年达 294.7 亿美元、同比增长 39.0%，2026 年第一季度进一步增至 94.5 亿美元、同比增长 129.4%，融资热度持续升温。PitchBook 口径下，2025 年 VC-backed agentic AI 公司融资 242 亿美元、交易 1311 笔，占 2015-2024 年累计 VC 交易价值的近 73%。细分赛道增速分化显著：法律 AI 领域，Harvey

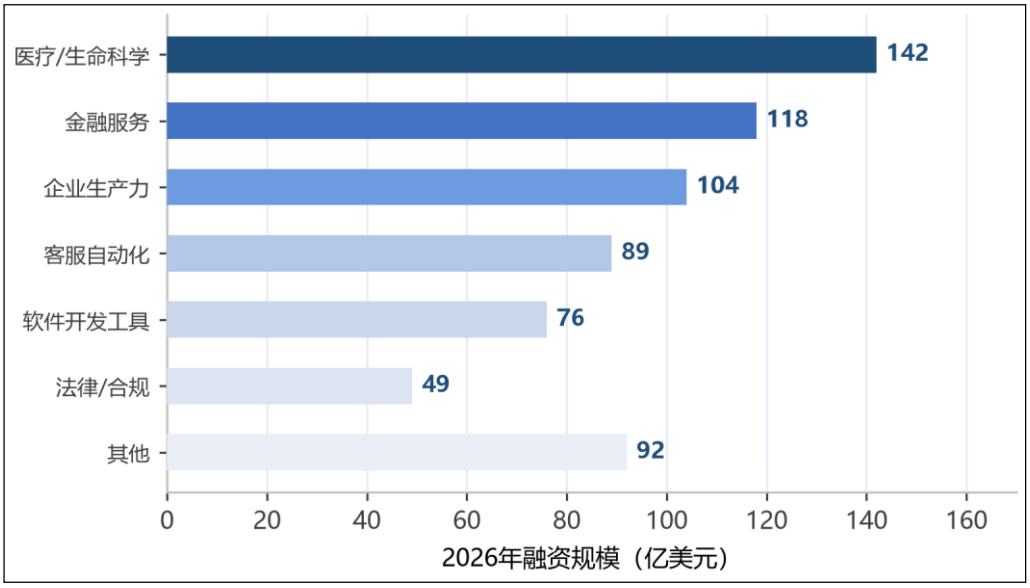
与 Legora 两家公司在 2026 年 3 月合计融资 7.5 亿美元（Legora 于 2026 年 3 月 10 日完成 5.5 亿美元 D 轮，Harvey 于 2026 年 3 月 25 日完成 2 亿美元 G 轮）；法律 AI 年度融资从 2023 年的 6.08 亿美元增长至 2025 年的 43 亿美元；数字健康 AI 增长 33%，金融科技智能体增长 50%，而通用横向平台融资下降 23%——资本明确从“通用平台”转向“垂直深耕”。

图表 21　2025-2026 年中国代理式 AI 融资规模



数据来源：毕马威，九思行研整理

图表 22　2026 年全球智能体垂直场景融资分布



数据来源：Axis Intelligence，九思行研整理

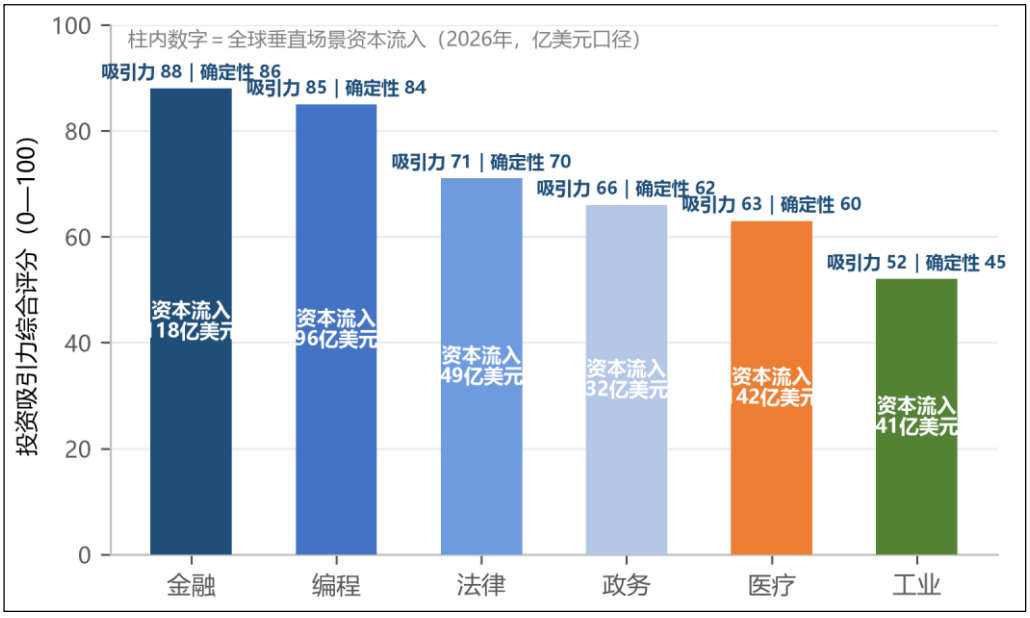
6.2.2 平台层、垂直应用层与基础设施层的资本分布

资本在产业链三层之间的分布正在经历结构性迁移。平台层仍是大额资金的聚集地——基础模型与头部平台以 200 亿美元以上估值持续完成巨额融资，但增长型资本（growth capital）的边际投入在减少。垂直应用层成为增量资本的主战场——垂直行业解决方案 2026 年占据 34%的市场份额（2025 年为 12%），行业 specific 的智能体厂商正获得在位企业的战略投资；应用层初创已展现产品市场匹配、从种子轮与 A 轮毕业进入成长期融资。基础设施层的价值被重新发现——编排、监控、治理、评测工具与智能体可观测性平台获得密集投资，2025 年纯智能体初创捕获 29 亿美元、高度集中于多智能体编排平台与工作流执行工具；2025 年第四季度 AI 初创整合加速，并购与关闭数量较此前显著上升，资本在向头部集中。

6.2.3 金融、法律、医疗、编程、政务与工业的投资吸引力排序

综合部署进度、付费验证、资本热度与政策确定性，本报告对六大场景的投资吸引力排序为：金融（88 分）＞编程（85 分）＞法律（71 分）＞政务（66 分）＞医疗（63 分）＞工业（52 分）。金融居首的依据是备案机构数持续增长（32 家、超 50 个项目）、灵犀体系 8300 万人次的服务规模与按效果付费模式的率先跑通；编程次之，78%的行业渗透率与近两倍的市场规模增速证明其商业化最成熟，但巨头垄断格局压缩了独立投资标的的可用性；法律位居第三，融资同比增长 7-8 倍与垂直 Know-how 壁垒提供弹性，但单客户价值与监管变量仍需观察；政务的预算刚性提供防御性，工业的长周期属性决定其更适合产业资本分阶段布局。该排序为证据化推断，成立条件是各场景的监管框架与支付制度不发生方向性逆转。

图表 23 资本配置优先级：六大场景投资吸引力排序



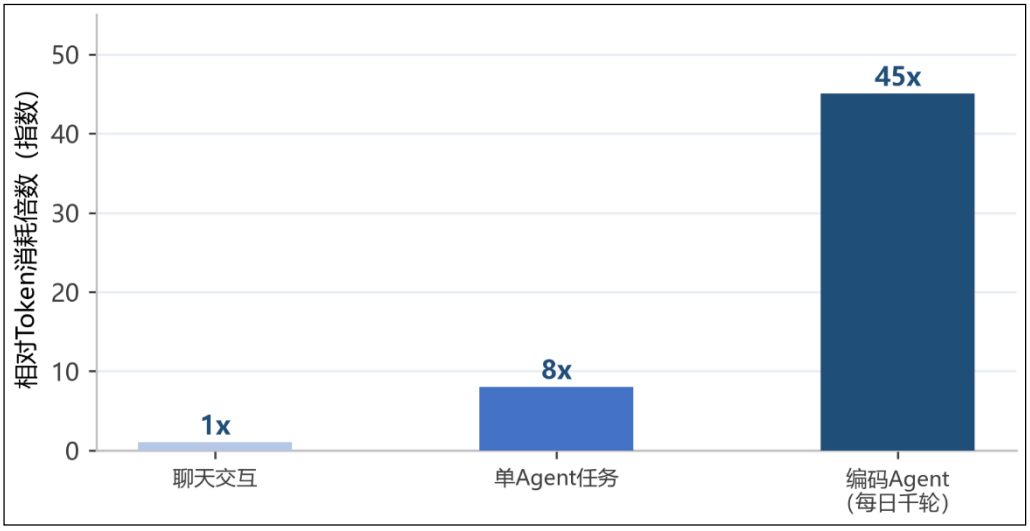
资料来源：九思行研基于公开数据综合评估

6.3 AI 智能体核心风险与项目取消率的结构性成因

6.3.1 单 Agent 任务 LLM 调用量远高于聊天交互的成本失控问题

成本失控是智能体项目取消率的首要技术成因。与聊天式 AI 的单轮交互相比，单智能体任务的 LLM 调用量呈量级上升：一次自主任务需要规划、检索、工具调用、自我校验的多轮循环，据 Menlo Ventures 2026 年数据，企业月度 LLM 支出中位数同比增长 7.2 倍；编码智能体任务每日若执行 1000 轮、每轮重读 20 万 token 上下文，算力成本将迅速吞噬项目预算。成本结构的特殊性在于其非线性——试点阶段的小规模验证无法暴露规模化后的成本曲线，项目从试点转入生产后成本超预期成为普遍现象。应对路径包括：推理成本工程（上下文缓存、模型分级调用、开源模型替代）、计费模式对冲（按结果付费将成本风险部分转嫁）与预算护栏（调用限额与成本熔断机制）。

图表 24　智能体任务的 Token 消耗量级对比（指数）



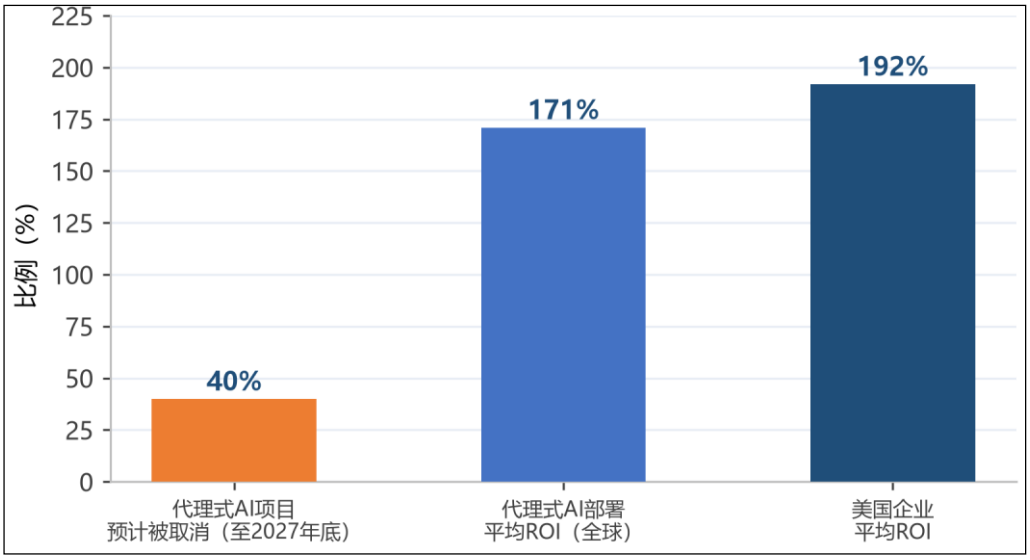
资料来源：九思行研基于行业公开测算整理

6.3.2 业务价值不清与风险管控不足的结构性原因

项目取消的第二、第三大成因是价值不清与管控缺位，两者同源于组织能力缺口。价值不清的管理根源是 AI 采购决策权从 IT 部门向业务部门转移过程中的错位——业务部门为价值买单但缺乏技术评估能力，IT 部门掌握技术评估但不背业务指标，智能体项目常陷入“双重否决”的僵局。风险管控的缺口更为普遍：德勤《2026 年人工智能现状报告》显示 73%的企业将数据隐私与安全列为最关注的 AI 风险，但仅约 21%建立了面向自主智能体的成熟治理模型；近四分之三（74%）的企业计划在 2027 年前

较大程度使用 AI Agent。Gartner 于 2025 年 6 月预测的“到 2027 年底超过 40%的代理式 AI 项目将被取消”与“成功部署平均 ROI 达 171%”并不矛盾——这是同一个市场分化现象的两面：到达生产成熟度的少数项目回报极佳，多数项目止步于治理与数据质量。

图表 25　代理式 AI 项目取消率与已部署项目回报并存



数据来源：Gartner，九思行研整理

6.3.3　技术路线、监管走向与地缘政治的不确定性

三类外部不确定性构成智能体产业的系统性风险。技术路线层面，模型能力的代际跃迁可能使现有应用层产品快速贬值——2026 年开源模型（DeepSeek V4 Flash 成本降至闭源约九十五分之一、Kimi K3 登顶代码评测）已显著压缩闭源模型的价格空间，应用层厂商的护城河若仅建立在模型调用之上，则随时可能被平台化能力覆盖。监管走向层面，人工智能治理框架仍在快速演进，模型备案、算法审计、责任认定的具体规则尚未完全定型，监管收紧将抬高合规成本、监管放松则可能引发低质量供给的泛滥。地缘政治层面，高端算力供给、先进制程与基础软件的国际竞争格局持续演变，算力成本与供给稳定性是全产业链的底层变量。三类不确定性的共性是难以对冲，产业方与投资方应通过“多技术路线组合、场景多元化布局”来分散而非预测。

6.3.4　项目取消后的组织学习与二次部署机会

项目取消并非终点，而是行业成熟过程中的必要学习成本。取消项目的组织学习价值体现在三方面：一是需求认知的校准——取消迫使用户重新审视真实业务需求，二次部署的项目范围显著收窄、目标更聚焦；二是治理能力的沉淀——取消多因治理缺位，取消后的制度建设（权限、审计、熔断）为二次部署铺平道路；三是供应商市

场的出清——取消潮淘汰差异化不足的厂商，幸存者获得更清晰的竞争环境。行业证据支持“二次部署成功率更高”的推断：2025 年第四季度 AI 初创整合加速后，2026 年投资质量改善、垂直解决方案份额提升至 34%。对产业方而言，取消后的复盘应制度化——将失败原因归类为需求、数据、治理、成本四类并建立检查清单，是降低二次部署风险的最有效动作。

6.4 AI 智能体未来 12-18 个月关键观察指标

6.4.1 企业级规模化部署率与续费率

部署率与续费率是未来 12-18 个月检验行业拐点的第一组指标。部署率维度，当前基线为：18%的中国企业将智能体纳入核心业务流程、31%的机构有智能体运行于生产环境；观察阈值应设为 12-18 个月内企业级规模化部署率（核心业务流程纳入口径）突破 45%、生产环境运行率突破 60%——若达标则行业进入规模复制阶段，若停滞则说明治理与成本瓶颈未解。续费率维度，当前公开数据稀缺，恰是需要重点跟踪的缺口指标；建议关注头部厂商财报或披露中的净收入留存率（NRR）数据，NRR 持续高于 110%将是商业闭环最终确认的信号。

6.4.2 按结果付费渗透率与单位经济模型

按结果付费的渗透进程决定收入质量演化的速度。当前全球主要厂商（Salesforce、Zendesk、Sierra、HubSpot）已全面布局结果计费产品，国内 RaaS 厂商在金融销售、数字劳动力场景实现盈利；观察指标应包括：头部 SaaS 厂商按结果付费收入占总收入的比例（预计 2027 年底超过 15%即确认代际切换）、按结果付费合同的平均期限（多年期合同占比提升说明模式成熟）、以及厂商单位经济模型中推理成本占比的下降斜率。IDC 预测到 2028 年 70%的软件供应商将放弃纯按席位收费，转向以消费、成果或能力为核心重构定价，该预测的中间节点验证是本组观察的核心。

6.4.3 各垂直场景 ROI 验证数据

六大场景的 ROI 验证数据将决定资本配置的再平衡。当前已具备的基线证据：金融——重庆银行个人按揭贷款审批时效提升 87%（单案例：资料处理从 210 分钟缩短至 15 分钟，整体审批流程压缩至 4 小时内，数据准确率达 95%）；国泰海通灵犀智能化服务 8300 万人次。法律——合同审查耗时从 4 小时至 15 分钟（该量级在多个来源中出现，AI 审查通常为 15-30 分钟），审查准确率 85%-92%。医疗——影像智能体处理 250 万例、12 项 AI 辅助诊断获医保支付；编程——开发效率提升 35%-55%、渗透率 78%；政务——重复填报减少 50%、办结率 100%；工业——停机减少 30%-50%、12-18 个月实现 10：1 至 30：1 回报。未来 12-18 个月应重点跟踪：各场景从单点案例

到行业平均水平的 ROI 迁移证据、第三方评测机构（而非厂商自述）的独立验证数据、以及 ROI 数据从头部机构向中小机构扩散的梯度曲线。

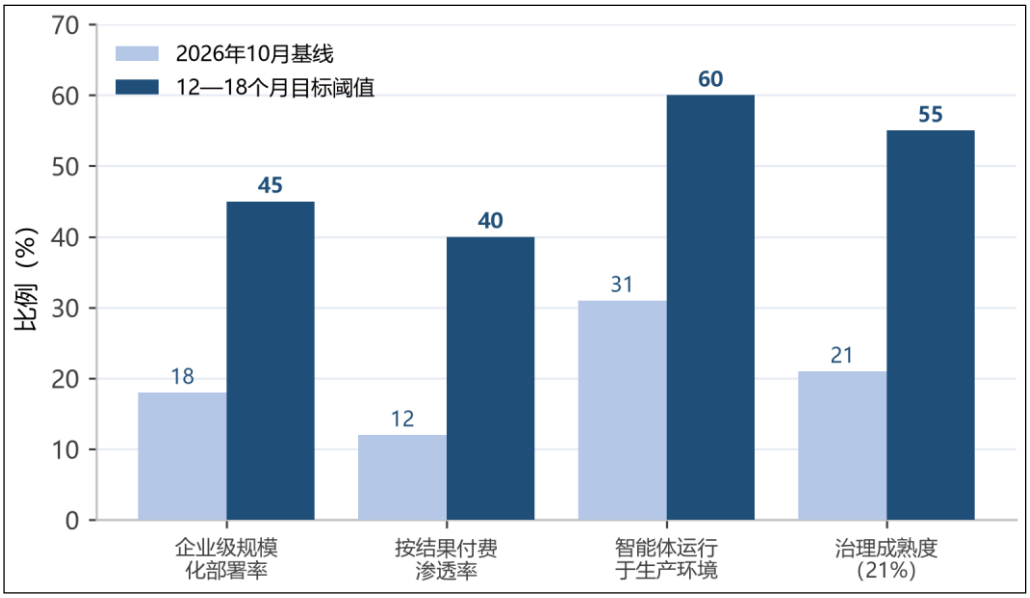
6.4.4 监管框架演进方向与合规成本变化

监管演进是所有场景共用的前置变量。未来 12-18 个月的关键观察点包括：算法与模型备案的覆盖范围扩展（当前证券业 32 家机构、超 50 个项目的备案节奏是否复制到银行、保险）、医疗领域 AI 三类证审批周期的变化与新形态（大模型、智能体化产品）的准入规则、政务领域数据不出院与国产化要求的统一程度、以及司法部 AI 法律服务合规试点扩大后的责任认定细则。合规成本方向上，备案、审计与私有化部署构成的合规成本曲线若趋平，则行业进入“监管友好期”；若显著陡峭，则中小供应商将被挤出、集中度加速提升。

6.4.5 智能体安全事件与信任修复机制

安全事件是智能体信任体系的最大威胁变量。需要建立的观察机制包括：智能体越权操作、数据泄露、幻觉导致业务损失等安全事件的行业通报频率；监管对重大事件的处置案例与追责口径；以及行业层面信任修复机制（保险、审计、熔断标准）的建设进度。当前治理基线是：仅 21%的机构建立成熟治理模型，而《医疗机构人工智能应用与治理专家共识（2026 版）》的熔断机制、福田区的“监护人”制度提供了行业样板。未来 12-18 个月若爆发标志性安全事件，短期将抑制部署节奏、长期将加速治理标准统一；投资者应将治理能力强的供应商视为安全事件冲击下的避险标的。

图表 26 未来 12-18 个月关键观察指标的基线与阈值



资料来源：九思行研基于 IDC、S&P Global、Deloitte 数据综合设定

综合全文分析，本报告的核心判断可以凝练为三句话。第一，AI 智能体的商业化已经越过概念验证阶段，2026 年是金融与编程场景商业闭环的确认之年，行业正处在“少数项目回报极佳、多数项目止步治理”的结构性分化期。第二，商业模式的代际切换——从按席位到按结果——是未来三年最重要的产业变量，它重新定义了智能体的价值锚点（从软件预算到人力成本）、重构了供应商与客户的激励结构，66%的中国企业倾向成果计费的采购偏好预示中国可能成为该模式渗透最快的市场。第三，投资窗口沿“金融－编程－法律－政务－医疗－工业”的顺序梯度展开，确定性与弹性在不同场景间的分布为不同期限与风险偏好的资本提供了各自的配置逻辑；而贯穿所有场景的共同前置变量是治理能力的建设——谁能率先建立可信的智能体治理与结果核验机制，谁就能在代际切换中占据产业与资本的先手。