

企业 AI 落地 白皮书

AI 落进业务，账上见数。

20 个已公开可查证的行业实证 · 95% 项目死因解剖 · 可签字的落地方法

一页读完全书

四篇内容，一条主线：AI 的价值不在演示里，在账本里。

第一篇 · 实证

别人已经省下的人和钱

20 个已公开、可查证的真实部署：顶 700 名坐席、质检省一半人、一座矿一年多赚 ¥2,000 万。从世界巨头到 200 人的中型公司，从中国到中东。

第二篇 · 认知

为什么 95% 的项目死在试点与投产之间

死因从来不是模型——没嵌进核心系统、没人算得清账、数据出不了域、没有验收体系。四个死因，每个都有解法。

第三篇 · 方法

怎么把 AI 落到账上

先定义「好」再动工；阈值不达不上线；每个动作可审计、每分钱可归因；数据不出域，模型进得来。

第四篇 · 行动

怎么开始

第一步是一场 2–4 周的 Bootcamp：驻场工程师在你的数据上跑通第一个真实用例，费用按成本价收取、签约后全额抵扣首期。90 天后，你桌上多出一个运行中的系统和一本算得清的账。

阅读方式：只有 5 分钟——读本页与第 3 页的数字墙；有 15 分钟——每章只读「带走的一句话」；准备立项——把书末的《落地自查清单》复印给团队。

带走的一句话

AI 创造的价值从来不在发布会上，在账本里。

01

第一篇 · 实证

别人已经省下的人和钱

20 个数字，全部来自公开发表的真实部署。

20 个已经发生的数字

每一格，都是一次公开发表的真实部署。

= 700 名坐席 AI 客服顶岗 Klarna	36 万小时 → 0 合同审阅 摩根大通	排查 -60% 反洗钱监控 汇丰银行	1,000 工程师 +20% AI 结对编程 Emirates NBD · 迪拜
一年 \$5 亿 全价值链 AI ADNOC · 阿布扎比	+¥2,000 万/年 洗选优化 山东能源	电耗 -20% 水厂调优 Jacobs	顶 140 名质检员 AI 质检 TCL 华星
缺陷率 -67% 产线视觉 AI 和硕	质检成本 -70% AI 验布 中科视语 · 50 厂	交付 +33% 总装产线 空客	采购 -3% · 11 天 AI 谈 2,000 家供应商 沃尔玛
拣货 +30-180% 仓储协作机器人 DHL	40 人干 3 倍量 自动分拣 中通	1,200 人转岗 AI 安防巡检 万物云	省 ¥750 万/年 酒店运营 AI 亚朵集团
预测准 10-15 个点 1.6 万家门店 百胜中国	运营成本 -35% 政务 AI 办结 迪拜政府	客服量 ×7 不加入 400 人公司 Synthesia	省 \$440 万 信用社 AI 接线 Neighborhood CU

金融 能源工业 制造质检 供应链物流 服务连锁 中小企业

来源见附录。你的数字，由你自己的审计账本验收。

带走的一句话

这些数字都已经发生。问题只剩一个：什么时候轮到你的业务。

风控更严，人却更少

客服 · KLARNA × OPENAI

1 个 AI = 700 名坐席

上线首月接管三分之二客服对话，单次解决 11 分钟 → 2 分钟，一年多赚 4,000 万美元

Klarna 官方 · 2024 ①

合同审阅 · 摩根大通

36 万小时/年 → 秒级

商贷合同审阅全自动，相当于 180 人的全年工时

Bloomberg · 2017 ②

反洗钱 · 汇丰 × GOOGLE CLOUD

排查 -60%，多抓 2-4 倍

机器学习替代规则引擎成为主监控系统，误报大降、漏报同步减少

Google Cloud · 2023 ③

柜面 · 工商银行 × 盘古

柜面 5 步 → 1 步

大模型自动生成办理路径，每笔业务节省 5 分钟以上

证券时报 ④

工程 · EMIRATES NBD × GITHUB

1,000 名工程师 +20%

≈ 白得 200 个工程师——迪拜头部银行的工程组织

GitHub 官方案例 ⑤

套上你的数字：200 席客服中心 → ≈120 席的工作量；10 人审单组 → 省 3 人。你的准确数字，Bootcamp 第一周用你自己的数据算清。

带走的一句话

监管最严的行业都跑通了，「我们行业特殊」不再是理由。

越重的资产，AI 省得越多

集团级 · ADNOC

一年 \$5 亿

30+ 个 AI 工具覆盖勘探到销售全价值链，价值可计量、对外披露

ADNOC 官方 · 2024 ⑥

矿山 · 山东能源 × 盘古

一座矿多赚 ¥2,000 万/年

精煤产率 +0.2%，新场景上线周期 6 个月 → 1 个月

证券时报 ⑦

公用事业 · JACOBS × PALANTIR

全厂电耗 -20%

在一座已经过传统优化的水厂上继续省出 20%，并消除环保罚款

Palantir Impact ⑧

设备维护 · 跨行业实证区间

停机 -30-50%

预测性维护：维护成本 -10-40%，设备寿命 +20-40%（保守区间）

麦肯锡跨行业研究 ⑨

套上你的数字：一条产线每停机 1 小时的损失 × 全年停机时数 × 30%——这就是预测性维护一项的起步账。

带走的一句话

重资产的每一个百分点，都是千万级的钱。

质检线，是 AI 顶人最狠的地方

面板质检 · TCL 华星

顶掉 140 名质检员

质检人力 -50%，单图判定 2 秒 → 0.5 秒，省 ¥1,000 万+（深圳三厂）

雷锋网 · 2025 ⑩

产线视觉 · 和硕 × NVIDIA

人力成本 -7% · 缺陷率 -67%

视觉 AI 上装配线，每条产线直接压人力与不良率

NVIDIA 官方案例 ⑪

纺织验布 · 中科视语

验布成本 -70%

30 人 → 15 人，速度快 3-5 倍；漏检率 20-30% → 3% 以下（50 家纺织厂）

纺织服装周刊 ⑫

总装 · 空客 × PALANTIR

A350 交付 +33%

产能就是收入——生产瓶颈用数据打通

Palantir / Airbus ⑬

套上你的数字：10 人的质检班组，按 -50% 实证 ≈ 省出 5 个人；三班倒就是 15 个。

带走的一句话

人力最密的工位，就是账最好算的工位。

一个 AI，同时跟 2,000 家供应商谈判

采购谈判 · 沃尔玛 × PACTUM

采购成本 -3%

AI 同时与 2,000 家尾部供应商谈判，周期数周 → 11 天，68% 成交

哈佛商业评论 ⑭

仓储拣货 · DHL × LOCUS

每人每小时 +30—180%

协作机器人 + AI 调度，40+ 仓库、10 亿次拣选验证

DHL 官方 ⑮

快递分拣 · 中通

40 人干出 3 倍的量

自动分拣线 750,000 件/日——同样的人，三倍的口径

澎湃新闻 ⑯

需求预测 · 跨行业实证区间

预测误差 -20—50%

库存 -20—30%，物流成本 -5—20%

麦肯锡 ⑰

套上你的数字：年采购额 ¥5 亿的公司，尾部品类按 -3% 实证 ≈ 一年省 ¥400 万上下（按尾部占比折算）。

带走的一句话

供应链的钱是一分一分省出来的——AI 一次省 2,000 家。

工资单越厚的行业，AI 的账越好算

物业安保 · 万物云

1,200 名员工转岗

AI 远程巡检替代现场轮值，人均效率 2.3 倍（479 个项目）

读特 / 36 氪 ⑮

酒店 · 亚朵集团

一年省 ¥750 万

30 万+ 工时的运营与服务自动化

澎湃新闻 ⑰

餐饮连锁 · 百胜中国

比店长预测准 10—15 个点

1.6 万家门店的销量预测与订货，95% 店长每天在用

公司披露 ⑳

政务 · 迪拜政府

运营成本 -35%

60%+ 例行咨询 AI 直接办结，办事快 25%（15 个用例）

迪拜数字局 ㉑

套上你的数字：1,000 人的物业/连锁企业，按万物云口径的转岗比例 $\approx 100+$ 人力重新配置。

带走的一句话

人力成本的大头在哪，AI 的利润表就在哪。

这不是巨头的游戏

400 人公司 · SYNTHESIA

客户量 ×7，客服没加人

87% 自助解决，处理时长 5 天 → 4.6 小时

Intercom 官方案例 ②②

区域信用社 · NEIGHBORHOOD CU

48 小时接管 90% 来电

AI 接线上线两天见效，客服运营省 \$440 万

interface.ai 案例 ②③

中型咨询公司 · ELIASSEN

一年 6,000 小时 = 3 个全职

财务月结自动化，覆盖 80+ 分支机构

Eliassen 官方 ②④

专精特新 · 金石机器人

售后工单 半个月 → 1-3 天

钉钉 AI 工单助理接入 1,000 家服务网点

量子位 · 2025 ②⑤

Forrester 对 200 人、营收 \$3,500 万规模复合企业的独立测算：运营成本 -20%，净收入 +6%。

带走的一句话

200 人的公司和 20 万人的公司，用的是同一批模型。差的只是落地。

客户量涨了七倍，客服一个人没加

SYNTHESIA · AI 视频软件

约 400 人 · 中型科技公司

场景：客户支持

困境

两年内客户从 4 万家涨到 31.6 万家，客服咨询量跟着暴涨近七倍。按老办法，客服团队也要扩编七倍才接得住。

AI 做了什么

AI 客服顶在第一线：常规问题直接解决，复杂问题升级人工；没答好的对话回流知识库，越用越准。

87%

咨询自助解决

5 天 → 4.6 小时

单件解决时长

1,300+ 小时

半年省下的人工

+0 人

客服编制增加

来源：Intercom 官方客户案例（厂商发布，客户具名）

带走的一句话

业务涨七倍，编制加零人——增长被 AI 吸收了。

一条热线，一年省下 \$440 万

NEIGHBORHOOD CU · 美国达拉斯

资产约 \$10 亿 · 区域信用社

场景：会员热线

困境

热线长期只接得住一半来电：高峰排队，会员抱怨，扩编的人力成本又扛不住。

AI 做了什么

语音 AI 第一线接听全部来电：查询、挂失、转账等常规业务直接办结，复杂事项转人工坐席。

48 小时

接管率 50% → 90%

100%

来电由 AI 首接

\$440 万

客服运营成本节省

43,000+ 小时

会员等待时间减少

来源：interface.ai 官方客户案例（厂商发布，客户具名）

带走的一句话 两天上线，当年省出 \$440 万。

20 个案例，三个共同点

① 都落在具体工作流上

没有一个案例是「全公司上 AI」。全部是一条一条工作流：接线、审单、验布、分拣、订货。

② 都算得出账

顶多少人、省多少钱、快多少——每个案例都能用一行数字说完。算不出账的项目，不在这 20 个里。

③ 都不是一次性项目

上线只是计量的开始：数字持续累计、持续验收、持续扩展到相邻工作流。

带走的一句话 AI 创造的价值从来不在发布会上，在账本里。

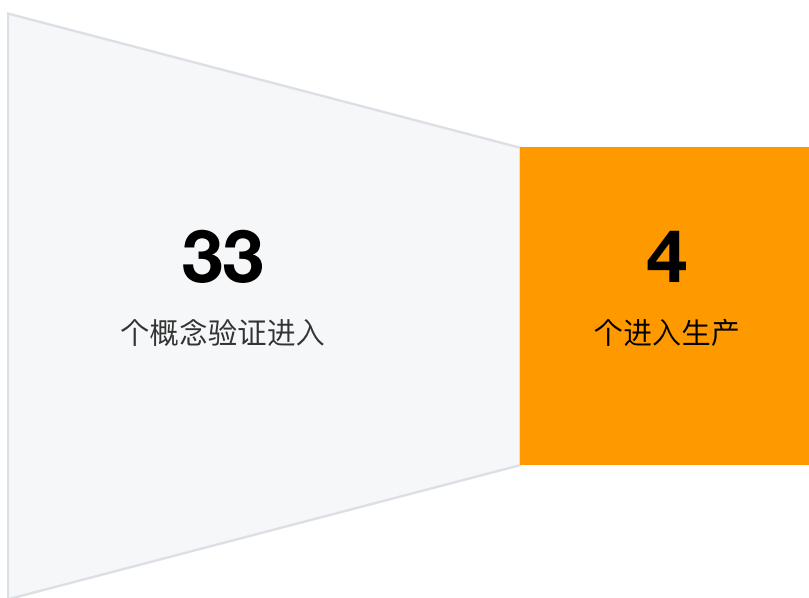
02

第二篇 · 认知

为什么 95% 的项目 死在试点与投产之间

上一篇的 20 个数字是幸存者。这一篇讲多数派：demo 惊艳、立项顺利、然后悄悄死掉的那些项目——以及它们共同的死法。

demo 与账本之间，隔着一道 95% 的鸿沟



IDC：平均每 33 个 AI 概念验证，只有 4 个上产

95%

的组织没有从 GenAI 试点中获得可衡量的损益回报（MIT，2025）

能力 ≠ 一致性

demo 展示的是能力——一两个亮眼样例就够；生产需要的是一致性——同类请求十次有十次做对。多步推理、工具调用、外部系统串在一起，任何一环的随机性都会被链式放大。（AWS《企业生产级智能体开发部署指南》，2026）

带走的一句话 上不了产的 AI，等于没有 AI。

解剖 95%：四个死因，没有一个是模型

死因一

没嵌进核心系统

模型悬在业务外面。答案再漂亮，进不了订单、风控、工单系统，就永远只是一个聊天窗口。

死因二

没人算得清账

没人能回答 CFO 一句「省了多少钱」。算不清账的预算，第二年就被砍。

死因三

数据出不了域

合规环境用不了公有云 API。试点永远跑在脱敏玩具数据上，永远「再等等」。

死因四

没有验收体系

从没人定义过什么叫「好」。上了线才发现：赢没赢，都没法算。

带走的一句话 **死因从来不是模型。**

自建、买工具，还是共建

	全自建	与 TokenPower 共建	只买 SaaS 工具
首个生产价值	12-18 个月起，组队常占一半时间	4-8 周第一条工作流投产	上手快，停在通用场景
三年总成本	约为共建路线的 2-3 倍	里程碑计费，金额提前写清楚	订阅便宜，落地改造另计
AI 人才	自招自养（全国缺口数百万级）	驻场工程师 + 你的业务专家	没有人对落地负责
模型更替	每年 4-8 次大版本自行消化	多模型网关统一吸收，不锁定	绑定单一厂商
合规与审计	审计体系从零自建	平台自带全链路审计与归因	通常没有
数据与 IP	归你	归你，数据不出域	视厂商条款而定

MIT，2025：外购与伙伴路线的成功率约 67%，自建约为其三分之一。Menlo Ventures：企业自建 AI 用例占比一年内从 47% 降至 24%。

带走的一句话 市场已经投过票了。

四个死因，四个解法



右边这一列长什么样、按什么顺序发生、凭什么可信——下一篇逐个拆开。

带走的一句话

每个死因都有解法——这就是下一篇。

03

第三篇 · 方法

怎么把 AI 落到账上

四个解法，一节对一个死因。

四步交付，落在四层平台上

交付环 · 循环而非流水线



四层平台 · 一条审计链



交付是环，不是线；平台是地基，不是工具箱。上一篇的四个死因，在这一页都有对应的解法。

带走的一句话 交付有流程，信任有系统——都不靠感觉。

开工第一件事不是写代码，是定义什么叫「好」

Bootcamp 第一周的产出不是代码，是下面这张双方签字的表。

指标	目标阈值	谁验收
工具 / 系统调用选对率	≥ 95%	双方联合
关键参数提取准确率	≥ 98%	双方联合
越权请求拒答率	100%	客户合规
响应延迟	P50 < 2 秒 · P95 < 5 秒	客户 IT
单次处理成本	< 共同商定上限	客户财务
业务指标（示例）	单均处理时长 -X% · 自助解决率 ≥ Y% · 误报 -Z%	客户业务方

技术指标	业务指标	两行同时出现
回答「AI 做得对不对」	回答「AI 做得值不值」	缺一行都不叫验收

带走的一句话 先定义「好」，再动工——像盖楼先出图纸。

三步交付：每一步可验收，每一步可叫停

① Bootcamp

2-4 周 · 成本价，全额抵首期

驻场工程师在你的数据上跑通
第一个真实用例，验收基线共
同签字

你拿到：能演示给管理层的运行
系统

每级之间有门：阈值是通行证

② 门控上产

4-8 周 · 里程碑计费

第一条工作流接入核心系统，
审计与计量开启。阈值不达，
不上线、不收费

你拿到：开始累计的降本数字

验收由基线表说话，不靠感觉

③ 规模化

按季 · 平台订阅

方案资产复用，扩展到相邻工
作流，单位成本逐季递减

你拿到：按季验收的扩张节奏

不继续时，系统、数据与黄金集都归你

带走的一句话

每个里程碑都可验收——也都可叫停。

审计不是日志，是产品



第 90 天你拿到的那份报表——省了多少人力、省了多少钱、多赚了多少钱——每一行都能点开，看到是哪些动作产生的。

带走的一句话 信任，在这里是一个可以查询的数据库。

治理不靠模型自觉，靠网关

✗ 写进提示词

「请不要访问无权限的数据」

靠模型自觉执行。概率性的，注入可绕过，出了事无法复现。

✓ 放进网关

请求到达工具之前完成鉴权

确定性的。权限、限流、策略、审计都发生在模型看到请求之前——模型没有机会「决定」要不要遵守。

「正确的做法是把安全控制外部化：放在独立的 Gateway 和 Policy 层，在工具被调用之前就完成鉴权——LLM 根本没有机会『决定』要不要遵守规则。」

— AWS《企业生产级智能体开发部署指南》· 2026

我们的网关层已在生产环境运行，每天处理真实流量。

带走的一句话 **治理不靠模型自觉，靠网关。**

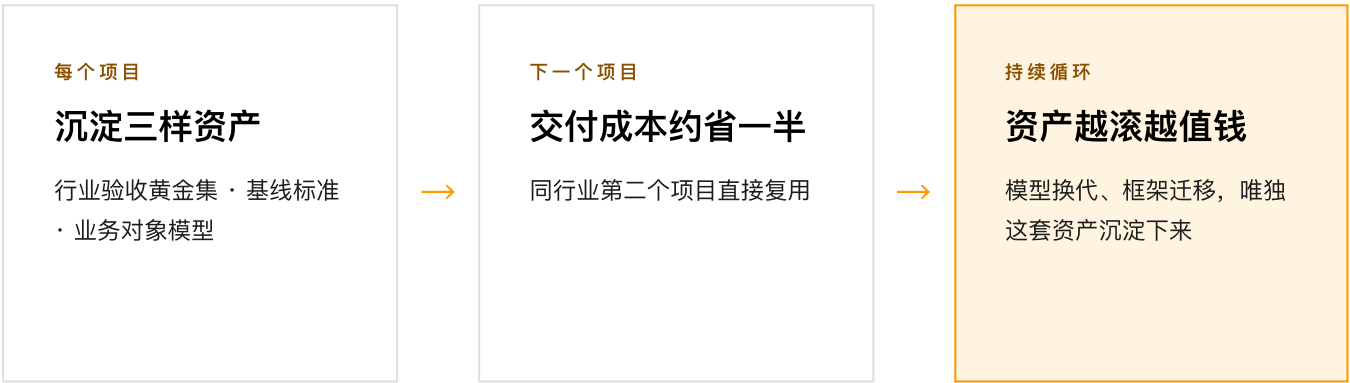
数据不出域，模型进得来



域内闭环：云上的评估与监控工具进不了隔离网络——我们把整套工具搬进你的域内：模型在域内跑，报表在域内出，验收在域内做。

带走的一句话 **数据不出域，模型进得来。**

模型会过时，黄金集越用越值钱



🔄 数据回流，黄金集持续增补

AWS 把这类资产称为「评估知识产权」

「评估平台可以买，但『什么算好』的定义权买不到——外包了黄金集与评分标准，等于把定义权交给了别人。」评估内容必须自主掌控。

资产归属，写在合同里

你的黄金集归你——数据、标注、验收标准都是你的资产。我们沉淀的是方法与工程件，所以第二个同行业项目，交付成本约省一半。

带走的一句话 **模型会过时，黄金集越用越值钱。**

我们敢按结果收费，因为结果可以被审计

主路径 · 里程碑计费

每一步的金额，开工前写清楚

- 01 Bootcamp：成本价，全额抵扣首期
- 02 上产：按共同签字的里程碑逐步计费
- 03 规模化：平台订阅，按季续订

可选项 · 效果分成

不见效，不收费

见效后按审计数据分享节省额。我们敢这么收，因为审计平台本身就是测量仪器。

Futurum, 2026：效果计费已占企业 AI 合同的 21.7%。

带走的一句话

我们敢按结果收费，因为结果可以被审计。

04

第四篇 · 行动

从一场 Bootcamp 开始

不需要一次决定全部。第一步是一场 2-4 周的 Bootcamp——驻场工程师用你的数据跑通第一个真实用例，成本价收取，签约后全额抵扣首期。

驻场 90 天，你桌上会多出什么



Bootcamp 定价：成本价，费用全额抵扣首期。第一周先和你共同签字验收基线——算出来不划算，就到此为止。

带走的一句话 90 天后，你多的不是一份报告，是一个系统和一本账。

三个最常见的担心，答在纸上

你拥有

领域知识 · 数据 · 决策权
知识产权（含你的黄金集）

我们负责

平台 · 模型运维
治理与审计 · 交付责任

数据安全

数据不出域，模型进得来

本地机房 / 专属云 / 隔离网络，出口侧
合规已完成

ROI 算不清

每个动作可审计、每分钱可归因

账本随时可查，报表逐笔可溯源

供应商依赖

不锁定，合同分段

按里程碑分段签约，每段验收后由你决定
是否继续；已交付的系统、数据与黄金集
都归你

合作伙伴

AWS（云与模型产品授权代理）· Microsoft · 火山引擎 · 通义千问 · DeepSeek · 智谱 · 月之暗面——两个模
型生态，一个货架，中立选型。

带走的一句话 担心不用被说服，用条款和账本回答。

落地自查清单（可复印给团队）

☐ 01 有没有一张双方签字的验收基线表？

☐ 03 AI 接进核心系统了，还是悬在外面当聊天窗口？

☐ 05 每一分钱成本能不能归因到具体业务？

☐ 07 模型一年换代 4—8 次，谁负责消化？

☐ 09 上线是阈值门控的，还是拍脑袋的？

☐ 11 如果明天叫停，资产归谁？

☐ 02 第一条 workflows，选的是不是代价最高的那条？

☐ 04 每一次 AI 动作有没有留痕？

☐ 06 数据出域了吗？合规部门知道吗？

☐ 08 误报、漏报有没有明确阈值？

☐ 10 生产里的失败案例，有没有回流改进？

☐ 12 第 90 天，CFO 能不能看到一本账？

12 条里有 3 条答不上来，就值得约一场 Bootcamp。

约一场 Bootcamp——2—4 周后，管理层看到的是运行中的系统。

TOKENPOWER.AI

每一个数字的出处

档位：Ⓐ 媒体 / 官方 / 研究机构报道 | Ⓑ 厂商发布的具名客户案例（未经独立审计）。完整链接备案。

案例	来源	档位
Klarna	Klarna 官方新闻 2024-02（注：2025-05 恢复部分人工坐席——本书据此主张有治理的人机协同，而非无人化）	Ⓐ
摩根大通 COiN	Bloomberg, 2017-02	Ⓐ
汇丰 × Google Cloud	Google Cloud 新闻室, 2023-06	Ⓐ
工商银行 × 盘古	证券时报 / 华为官网	Ⓐ
Emirates NBD × GitHub	GitHub 官方客户案例	Ⓑ
ADNOC	ADNOC 官方新闻稿, 2024-01	Ⓐ
山东能源 × 盘古	证券时报（济宁二号煤矿）	Ⓐ
Jacobs × Palantir	Palantir Impact 官方页	Ⓑ
TCL 华星	雷锋网 / 中华网科技, 2025	Ⓐ
和硕 × NVIDIA	NVIDIA 官方案例	Ⓑ
中科视语（验布）	纺织服装周刊, 2025	Ⓐ
空客 × Palantir	Palantir / Airbus Skywise 公开资料	Ⓑ
沃尔玛 × Pactum	哈佛商业评论 / Bloomberg, 2022-2024	Ⓐ
DHL × Locus	DHL 集团官方新闻稿, 2024-2026	Ⓐ
中通	澎湃新闻	Ⓐ
万物云	读特新闻 / 36 氪, 2025	Ⓐ
亚朵集团	澎湃新闻（案例评选）, 2025	Ⓐ
百胜中国	投资界 / 公司公开披露, 2025	Ⓐ
迪拜政府	迪拜数字局 DCAI 公开报告	Ⓐ
Synthesia × Intercom	Intercom (fin.ai) 官方客户案例	Ⓑ
Neighborhood CU	interface.ai 官方客户案例	Ⓑ
Eliassen	Eliassen 官方案例	Ⓑ
金石机器人	量子位, 2025-07	Ⓐ

方法论参考：AWS《企业生产级智能体开发部署指南》2026 · MIT NANDA《The GenAI Divide》2025 · IDC × 联想《AI CIO Playbook》2025 · Menlo Ventures 2025 · Forrester TEI（SMB）2024 · Futurum Group 2026（n=838）
· 麦肯锡预测性维护跨行业研究

TokenPower

TokenPower

看清，就决断。

TOKENPOWER.AI · DUBAI · 2026