



求智启新人工智能研究院

QIUZHI QIXIN AI RESEARCH INSTITUTE

— 企业AI转型落地 · 培训 · AGENT定制 —

企业人工智能落地实践指南

行业白皮书（2026）

—— 82 个可对照场景 · 前后成本对比 · 落地前提与避坑清单 ——

求智启新人工智能研究院 编著

二〇二六年八月

版权声明与免责声明

版权声明

本白皮书版权归求智启新人工智能研究院所有。任何单位或个人转载、摘编、引用本白皮书全部或部分内容，须注明来源为“求智启新人工智能研究院《企业人工智能落地实践指南——行业白皮书（2026）》”，且不得对内容进行歪曲、篡改。违反上述声明者，本研究院保留追究其法律责任的权利。

资料来源与编制立场

本白皮书的事实性素材来源于公开发布的政策文件、政府部门与权威研究机构的公开报告、上市公司公开披露信息及权威媒体公开报道。本研究院对上述公开信息进行了独立的整理、交叉校验、结构化归类与再分析；文中的场景拆解、成本对照口径、分析框架、判断结论与行动建议，为本研究院独立研究成果，不代表任何被引用机构或被提及企业的立场。

凡引用第三方数据或案例之处，均在正文或附录中标注了来源主体。本白皮书未获得任何被引用机构或被提及企业的授权或背书，亦未与其形成合作关系。所提及的企业名称、产品名称仅为说明行业现象，其商标权归各权利人所有。

关于案例数据的重要提示

本白皮书大量引用了企业公开披露的效率与成本数据。读者在使用这些数据时，请务必注意以下四点：

1. **口径由披露方自定。** 同一个“效率提升 50%”，可能指单环节耗时，也可能指端到端周期；可能含前期数据准备，也可能不含。本白皮书尽量还原了原始表述，但无法统一口径。
2. **未经独立审计。** 除政府部门发布的统计数据外，企业自行披露的数据未经第三方核验。
3. **存在幸存者偏差。** 失败项目极少被公开，因此基于公开案例得出的提效幅度系统性偏乐观。本白皮书为此专设“常见坑”栏目与第七章的失败模式分析。
4. **不可直接外推。** 单个企业的收益取决于其原有流程的效率水平、数据基础与组织能力，不宜作为本企业的预算依据。附录 D 提供了自主测算模板。

免责声明

本白皮书所载信息截至 2026 年 8 月，人工智能技术与产业演进迅速，部分数据可能随时间推移而失效。本白皮书内容仅供研究参考与决策辅助之用，不构成任何投资建议、法律意见、技术选型承诺或商业保证。任何单位或个人依据本白皮书内容作出的决策及由此产生的后果，本研究院不承担责任。

目 录

CONTENTS

版权声明与免责声明	2
版权声明	2
资料来源与编制立场	2
关于案例数据的重要提示	2
免责声明	2
前言：这是一本用来对照着做决定的手册	6
本手册的结构	6
怎么读一张场景卡片	6
摘要	8
一、本白皮书的核心发现	8
二、各行业已被验证的改写幅度（节选）	8
三、七条落地法则	9
四、本白皮书的原创工具	9
编制说明与研究方法	10
研究范围	10
场景的定义与拆分粒度	10
资料来源分层	10
对照表数据的处理原则	10
本白皮书的局限	11
第一章 现在处在什么阶段	12
1.1 采用率触顶，价值仍在少数人手里	12
1.2 政策从倡导性文件转向带指标的专项行动	12
1.3 从“对话式助手”走向“任务型智能体”	13
1.4 安全与治理成为规模化的首要约束	14
第二章 四个判断工具	15
2.1 SVA-3D：场景价值三维评估模型	15
2.2 L0—L4：场景成熟度阶梯	16
2.3 价值四层：收益该怎么考核	17
2.4 提效账本模型：省下的东西怎么变成钱	17
第三章 六大行业与获客职能的场景卡片	19
3.1 电商与零售	19
3.2 制造业	28
3.3 金融业	35
3.4 医疗健康	42
3.5 教育	48
3.6 政务与公共服务	53
3.7 自媒体获客与创始人 IP	58
第四章 十二个行业场景速览	68
4.1 物流与供应链	68
4.2 能源与电力	70
4.3 汽车与出行	72

4.4 电信运营	75
4.5 法律与合规服务	77
4.6 人力资源与组织管理	79
4.7 保险	82
4.8 农业	84
4.9 建筑与房地产	86
4.10 文化与娱乐与内容产业	88
4.11 餐饮与生活服务	90
4.12 科学研究与专业服务	92
第五章 七条落地法则	95
5.1 法则一：窄而深，胜过宽而浅	95
5.2 法则二：把七成资源投给人和流程	95
5.3 法则三：先定度量衡，再谈技术方案	95
5.4 法则四：区分"AI 提示"与"AI 决定"	96
5.5 法则五：数据治理独立立项	96
5.6 法则六：L3 级场景采购，L1—L2 级场景自建	96
5.7 法则七：为"退出"做设计	97
第六章 这笔账到底怎么算	98
6.1 收益侧：四类节省，只有两类能直接入账	98
6.2 投入侧：六类成本，后三类最容易被漏掉	98
6.3 一个可操作的测算流程	99
第七章 失败模式与风险治理	100
7.1 九种反复出现的失败模式	100
7.2 风险图谱：从内容风险到行动风险	101
7.3 国内外研究的共同指向	101
7.4 一个可操作的治理起步方案	102
第八章 趋势展望（2026—2030）	103
8.1 智能体渗透进入企业应用主流	103
8.2 价值计量方式从"节省工时"转向"闭环单量"	103
8.3 行业专用模型与通用模型形成分层	103
8.4 数据资产成为能力差距的主因	103
8.5 治理能力从成本项转为竞争力	104
第九章 给不同角色的行动建议	105
9.1 给企业决策者	105
9.2 给业务部门负责人	105
9.3 给技术负责人	105
9.4 给行业组织与政策研究者	106
附录 A 参考信息源清单	107
A.1 A 级：政策文件与官方发布	107
A.2 B 级：研究机构与咨询机构报告	107
A.3 C 级：企业公开案例与媒体报道	108
附录 B 场景速查总表	111
附录 C 场景自评量表	115
C.1 结果解读	115
C.2 使用提示	116

附录 D ROI 测算表模板 117

D.1 第一步：基线记录表	117
D.2 第二步：收益测算表	117
D.3 第三步：投入测算表	117
D.4 第四步：三情景测算	118
D.5 第五步：复盘节点约定	118

附录 E 术语表 119

关于求智启新人工智能研究院 120

我们是谁	120
我们的服务体系：四阶段递进	120
重点赋能行业与交付案例	120
联系我们	121
关于本白皮书	121

前言：这是一本用来对照着做决定的手册

市面上关于人工智能的报告已经足够多。基础模型的参数竞赛、算力集群的规模比拼、智能体架构的路线之争，都被反复书写。但当我们把这些材料放到企业管理者面前时，最常听到的一句话是：

"我知道 AI 很重要，但我不知道它在我这个环节上，究竟能省下多少人、多少天、多少钱。"

这是一个关于"最后一公里"的问题，也是本白皮书试图回答的唯一问题。

举一个具体的例子。一个服装品牌要做冬季羽绒服的商品图，希望呈现"孩子穿着羽绒服站在雪山下"的场景。传统路径是：找模特、找摄影团队、租场地或外出取景、修图、审片——一次拍摄的直接成本在数万元量级，周期以周计；如果要出海外雪山实景，成本还要再翻数倍。今天的另一条路径是：用一张平铺产品图加一段文字描述，由生成模型直接产出成片，单张成本降到以元为单位，出片周期以分钟计。

这不是一个"提升了 30%效率"的故事，而是一个**成本结构被整体改写**的故事。同样的改写，正在客服、质检、信贷调查、病历书写、政务热线、合同审查等几十个环节中同时发生——有的已经改写完毕，有的刚刚开始，有的看上去很美但一算账并不成立。

本白皮书要做的，就是把这几十个环节一个一个拆开，把账算清楚：**原来怎么做、花多少、多久；现在怎么做、花多少、多久；差在哪里；要具备什么前提才能实现；以及有哪些坑会让这笔账算不回来。**

本手册的结构

全书分为四个部分：

- **第一部分（第一至二章）：判断与框架。**说明当前所处的产业阶段，并给出本研究院的场景评估工具——你可以用它来判断自己的候选场景值不值得做。
- **第二部分（第三至四章）：场景卡片。**全书的主体。十八个行业加一个跨行业获客职能，共 82 个场景，每个场景一张卡片，格式统一，可以直接翻到对应位置对照使用。
- **第三部分（第五至七章）：方法与风险。**从场景卡片中提炼的跨行业规律、ROI 测算方法与失败模式分析。
- **第四部分（第八至九章及附录）：行动与工具。**分角色的行动建议，以及可直接使用的自评量表、测算模板与场景速查表。

怎么读一张场景卡片

每张场景卡片包含六个固定栏目，建议按以下方式使用：

表 0-1 场景卡片的六个栏目与使用方式

栏目	内容	建议用法
价值一句话	这个场景到底改变了什么	判断是否值得继续往下读
关键指标对照表	传统做法 vs 引入 AI 后 vs 变化幅度	把你自己的现状填进"传统做法"列，看看差距是否成立
公开案例与数据	有名有姓的企业与其披露的数字	用于内部立项材料的佐证，

栏目	内容	建议用法
		但须注明来源与口径
落地前提	要拿到上述效果，必须先具备什么	最重要的一栏。 前提不具备时，表中的效果不会出现
常见坑	已被反复验证的失败方式	在方案评审时逐条对照
适用与规模判断	什么体量、什么类型的组织适合做	判断自己是不是该场景的目标用户

提示：如果时间有限，建议优先读"落地前提"与"常见坑"两栏。对照表里的数字决定这件事值不值得做，而这两栏决定你能不能做成。

求智启新人工智能研究院
二〇二六年八月

摘要

一、本白皮书的核心发现

- 1. **AI 带来的不是"效率提升百分之几"，而是成本结构的改写。** 在电商内容生产、授信调查报告、门诊病历书写等已经跑通的场景中，公开披露的改善幅度普遍是一个数量级以上：内容生产成本降幅约 90%、百页授信调查报告从约一周压缩至 5 分钟内、门诊病历从 5—10 分钟降至十几秒。这类改变无法用"优化"来描述，只能用"重做"来描述。
- 2. **但这种改写只发生在少数环节。** 同一个行业内，不同环节的改善幅度可以相差几十倍。判断一个环节能否被改写，取决于三个条件：数据是否已在手、错了是否能改、省下的东西是否能被计量。本白皮书用 SVA-3D 模型把这三个条件工具化。
- 3. **效果的差距主要来自前提条件，而非模型选择。** 在本白皮书收集的案例中，同一类场景在不同企业的效果差异，绝大多数可以追溯到数据基础、流程配套与责任设计的差异，而非所用模型的差异。这也是每张卡片都设"落地前提"栏的原因。
- 4. **采用已经普及，价值仍然集中。** 斯坦福大学以人为本人工智能研究院 2026 年《AI 指数报告》显示约 88%的企业已在使用 AI，但生产力收益集中在少数领先群体；国际咨询机构的调研亦显示仅约四分之一的高管认为公司已从 AI 获得显著价值。"用了"和"算得过账"之间，仍隔着一道鸿沟。

二、各行业已被验证的改写幅度（节选）

表 0-2 部分场景的公开披露改善幅度速览

行业	场景	公开披露的改善
电商零售	商品图与详情页生成	内容生产成本降幅可达约 90%； 30—40 人团队的工作量由 1—2 人承担
电商零售	门店智能补货	缺货率由最高约 7%降至 2%以内； 库存周转由最高约 35 天降至普遍 21 天以内
金融	授信调查报告生成	百页报告由平均约 1 周缩短至 5 分钟内
金融	信贷资料核验	进件由小时级压缩至分钟级； 人工返工率下降约 80%；单户运营成本降低一半以上
医疗	门诊病历生成	单份病历由 5—10 分钟降至 15—20 秒； 平均 44 秒生成完整病历，看诊时长下降约 40%
制造	视觉质检	漏检率降至 0、过检率稳定在 0.5%以内（施耐德电气公开案例）
政务	12345 热线智能派单	派单准确率 90%以上； 处理效率提升约 25%，响应时间缩短约 30%
法律	合同审查	复杂合同审查由 4—8 小时缩短至 20—30 分钟； 条款遗漏风险由约 20%降至约 2%
汽车	空气动力学仿真	单次计算由约 4 小时缩短至约 1 分钟（捷豹路虎公开案例）
电信	基站与机房智能节能	年均节电约 8 亿度、节约电费约 5.2 亿元（中国电信公开披露）

资料来源：求智启新人工智能研究院根据公开信息整理。各项数据的口径由披露方自行定义，未经独立审计，不同案例之间不具备直接可比性，不应作为本企业的预算依据。完整来源见附录 A。

三、七条落地法则

从 82 个场景的横向比对中，本白皮书提炼出七条不依赖具体行业与技术路线的规律，详见第五章：窄而深胜过宽而浅；七成资源投给人和流程；先定度量衡再谈方案；区分"AI 提示"与"AI 决定"；数据治理独立立项；L3 级场景采购、L1—L2 级场景自建；为"退出"做设计。

四、本白皮书的原创工具

本白皮书提出并全书贯彻使用四个自有工具：**SVA-3D 场景价值三维评估模型**（判断场景值不值得做）、**L0—L4 场景成熟度阶梯**（判断场景走到哪一步）、**价值四层分类**（判断收益该怎么考核）、**AI 提效账本模型**（把节省折算成可入账的数字）。附录 C 与附录 D 提供了可直接使用的量表与测算模板。

编制说明与研究方法

研究范围

本白皮书聚焦"人工智能技术在具体业务环节中的应用方式与效率改善"，不涉及基础模型技术路线评测、算力基础设施建设、芯片与硬件产业链等议题。行业划分参照国民经济行业分类的大类逻辑，并结合 AI 应用场景的实际聚类做了适度合并与拆分，最终形成六个深度行业、十二个扫描行业，外加一个跨行业的"自媒体获客与创始人 IP"职能专题，共 82 个标准场景。

场景的定义与拆分粒度

本白皮书对"场景"的定义是：**一个有明确输入、明确产出物、且能被单独计量的业务环节**。例如"AI 赋能电商"不是一个场景，"商品主图生成"才是；"AI 赋能银行"不是一个场景，"授信调查报告自动生成"才是。

采用这一粒度的原因很实际：只有拆到这个粒度，才能回答"原来花多少、现在花多少"。粒度过粗的描述（如"AI 提升运营效率 30%"）无法被验证，也无法被复制。

资料来源分层

表 0-3 信息源分层与使用规则

层级	来源类型	在本白皮书中的用途
A 级	国务院及各部委正式印发的政策文件、 国家统计局口径数据	用于界定政策边界、时间表与量化目标， 直接引用并标注文号
B 级	国家级研究院所、高校研究机构、 国际权威咨询机构的公开报告	用于总体趋势判断与跨行业对比， 引用时标注机构与发布时间
C 级	上市公司公开披露信息、权威媒体 公开报道、企业公开案例	用于场景卡片中的案例与提效数据， 标注为"公开案例"，不作为统计推断依据

对照表数据的处理原则

每张场景卡片的"关键指标对照表"是本白皮书的核心，其数据处理遵循四条原则：

1. **"传统做法"列尽量取行业常规水平，而非最差水平。** 用最差水平做基线会人为放大改善幅度。若公开材料只提供了极端值，本白皮书会在表下注明。
2. **"引入 AI 后"列只填公开披露或可交叉验证的数字，不做推算。** 无可靠数据的单元格标注为"—"，而非填入估计值。
3. **变化幅度不做跨案例平均。** 每个数字都对应具体的披露主体与口径。
4. **区分"单环节"与"端到端"。** 单环节耗时下降 90%，不等于端到端周期下降 90%——中间的等待、审批、交接往往才是主要耗时。本白皮书在表注中标明了口径。

本白皮书的局限

- **幸存者偏差。** 可获取的案例绝大多数是成功案例，因此对“提效幅度”的呈现可能系统性偏乐观。第七章专门讨论失败模式以部分抵消这一偏差。
- **口径不一致。** 不同机构对“采用 AI”“落地”“规模化”的定义差异较大，跨报告的数字不宜直接比较。
- **时间点差异。** 部分案例的披露时间早于 2026 年，其数据反映的是当时的技术与成本水平，当前实际水平可能已发生变化。本白皮书在附录 A 中标注了各来源的时间。
- **未覆盖成本侧的完整信息。** 多数公开案例只披露收益不披露投入。本白皮书通过第六章的隐性成本清单与附录 D 的测算模板部分弥补这一缺口。

第一章 现在处在什么阶段

本章不讨论技术，只回答一个问题：如果你现在要为组织制定 AI 落地计划，外部环境给了你什么条件与约束。四个判断，构成后面所有场景卡片的共同背景。

1.1 采用率触顶，价值仍在少数人手里

衡量 AI 产业化进度有两个指标：多少企业在用，多少企业算得清账。这两个指标在 2026 年出现了明显背离。

采用广度上，斯坦福大学以人为本人工智能研究院（Stanford HAI）2026 年《AI 指数报告》显示约 88% 的受访企业已在业务中使用人工智能——“是否使用 AI”已不再是有区分度的问题。价值深度上，同一报告指出生产力收益集中于少数领先梯队；波士顿咨询公司（BCG）的调研发现仅约四分之一的高管认为公司已从 AI 创造了显著价值；麻省理工学院相关研究团队 2025 年的一份报告更以“约 95% 的生成式 AI 试点未产生可衡量的损益表影响”引发广泛讨论。

研究院观点：不要把“试点失败率”读成“技术不成熟”

公开讨论中被归因为“失败”的项目，绝大多数具备三个共同特征：一是选择了容错空间小但数据准备不足的场景；二是把 AI 作为旧流程的加速器，而非流程重设计的契机；三是立项时没有定义可归因的度量指标，因而验收时无法证明收益。

换言之，这更接近一个**项目管理与组织变革问题**，而非模型能力问题。本白皮书用大量篇幅写“落地前提”和“常见坑”，而不是写模型选型，原因就在这里。

1.2 政策从倡导性文件转向带指标的专项行动

2025 年 8 月，国务院印发《关于深入实施“人工智能+”行动的意见》，确立六大重点行动（科学技术、产业发展、消费提质、民生福祉、治理能力、全球合作），并给出三阶段目标：到 2027 年新一代智能终端与智能体应用普及率超 70%，到 2030 年超 90%，到 2035 年全面步入智能经济和智能社会发展新阶段。此后半年内，各主管部门的专项实施意见密集出台，且普遍带有可考核指标。

表 1-1 “人工智能+”政策体系的主要落地文件（截至 2026 年 8 月）

领域	文件与发布方	时间	关键量化目标
总体	《关于深入实施“人工智能+”行动的意见》 国务院	2025.08	2027 年智能终端与智能体应用普及率超 70%； 2030 年超 90%
能源	《关于推进“人工智能+”能源高质量发展的实施意见》 国家发展改革委、国家能源局 (国能发科技〔2025〕73 号)	2025.09	2027 年前推动 5 个以上专业大模型应用、 挖掘 10 个重点示范项目、探索百个应用场景； 2030 年相关技术与应用达到世界领先水平
医疗	《关于促进和规范“人工智能+医疗卫生”应用发展的实施意见》 国家卫生健康委等五部门	2025.11	到 2030 年基层诊疗智能辅助应用 基本实现全覆盖
制造	《“人工智能+制造”专项行动实施意见》	2026.0	到 2027 年：推动 3~5 个通用大模型深度应

领域	文件与发布方	时间	关键量化目标
	工业和信息化部等八部门	1	用、 推出 1000 个高水平工业智能体、 打造 100 个工业领域高质量数据集、 推广 500 个典型应用场景、选树 1000 家标杆企业
教育	《"人工智能+教育"行动计划》 教育部等五部门（教科信〔2026〕1 号）	2026.0 4	到 2030 年形成 AI 与教育深度融合格局； 建设国家教育智算服务平台、 基础语料库与教育专用大模型
消费	《关于加快"人工智能+消费"发展的实施意见》 商务部等八部门（商建发〔2026〕89 号）	2026.0 6	以定性目标为主，聚焦商品消费、 服务消费、商业创新三大方向

资料来源：求智启新人工智能研究院根据各部门公开发布的政策原文整理。

对企业的实际含义有三条：**其一，指标下沉**——制造领域的目标已细化到智能体数量与标杆企业数量，地方主管部门与国企将面临可考核的落地压力，相关的政策资源与示范项目机会随之出现；**其二，节奏一致**——多数文件把 2027 年作为第一个考核节点，这意味着未来一年半是政策资源最密集的窗口期；**其三，覆盖面从生产端扩展到消费端与民生端**。

1.3 从"对话式助手"走向"任务型智能体"

2025 年之前，企业侧的典型 AI 形态是"助手"：员工提问，模型给出内容，员工判断并使用。2026 年的显著变化，是"智能体"开始承担端到端的任务执行。研究机构 Gartner 的预测是：具备任务型 AI 智能体的企业应用占比将从 2025 年的不足 5% 上升到 2026 年底的约 40%。

这个转变直接影响你怎么算账。在助手阶段，提效的计量单位是"节省的人时"，收益依赖员工的使用意愿；在智能体阶段，计量单位变成"自动闭环处理的业务单量"与"人工干预率"，收益直接体现在流程吞吐与人力配置上。

表 1-2 两种 AI 形态在提效计量上的差异

对比维度	对话式助手（2023—2025 主流）	任务型智能体（2026 起加速）
介入方式	员工主动调用，逐次交互	嵌入流程节点，被事件触发
产出物	文本、代码、图像等中间产物	完成状态的业务单据或系统动作
核心指标	使用率、人均节省时长	自动闭环率、人工干预率、单位处理成本
收益归因	难，依赖员工自我报告	较易，可从系统日志直接统计
主要风险	内容准确性、员工不用	越权操作、错误动作的连锁扩散
对管理者的直接含义		
如果贵单位当前的 AI 成效汇报仍以"员工使用率"和"人均节省小时数"为主，说明应用仍停留在助手阶		

对比维度	对话式助手（2023—2025 主流）	任务型智能体（2026 起加速）
段。向智能体阶段过渡的第一个动作不是换模型，而是 重新定义指标 ：找出一条完整的业务流程，定义它的自动闭环率基线，然后再谈技术方案。		

1.4 安全与治理成为规模化的首要约束

当 AI 从生成内容转向执行动作，风险性质也随之改变。麦肯锡公司 2026 年 3 月的负责任 AI 成熟度调研显示：近三分之二的受访者将安全与风险问题列为全面扩展智能体应用的首要障碍；74%的企业认为模型输出不准确是高度相关的风险；72%将网络安全视为关键风险。中国信息通信研究院在《人工智能产业发展研究报告（2025 年）》中同样将安全治理单列成章。

本研究院的判断是：**2026—2027 年，制约行业 AI 规模化的主要瓶颈将从“能不能做”转为“敢不敢放”**。尤其在医疗、金融、政务、能源等高监管行业，决定部署节奏的是权限设计、审计留痕与责任划分机制，而非模型能力。这一点在后面的场景卡片中会反复出现——凡是成熟度停在 L2 的场景，几乎都卡在这里。

第二章 四个判断工具

要在十八个行业、七十多个场景之间做有意义的比较，必须先有统一的语言。本章介绍本研究院自建的四个工具，它们贯穿全书：**SVA-3D 模型**回答“值不值得做”，**成熟度阶梯**回答“走到哪一步了”，**价值四层**回答“该怎么考核”，**提效账本模型**回答“省下的东西怎么变成钱”。

2.1 SVA-3D：场景价值三维评估模型

常见的 AI 场景评估方法多沿用“价值—可行性”二维矩阵。这一方法用于 AI 场景时会漏掉一个关键变量：**同样的技术能力，在容错空间不同的场景里，可用性天差地别。**

一个具体例子：同样是“从非结构化文本中抽取要素”这一能力，用于电商商品描述生成时，95% 的准确率完全可用——错了改一下即可；用于临床诊断报告生成时，95% 意味着不可接受的风险，必须叠加医师复核。二维矩阵会把这两个场景放在相近位置，从而给出误导性的优先级建议。因此我们引入第三个维度。

维度一：数据可及性

指开展该场景所需的数据，在多大程度上已经沉淀、可获取、可用于训练或检索。

- **高**：数据已在企业内部系统中结构化沉淀，历史充足，无跨主体调取障碍。如制造业设备运行数据、电商交易与商品数据。
- **中**：数据存在但分散在多个系统或以非结构化形式存在，需投入治理成本。如银行信贷档案、医院病历文书。
- **低**：数据尚未系统性采集，或受法规限制难以汇聚。如农业田间实时数据、跨机构个人健康数据、跨企业供应链数据。

维度二：容错空间

指单次错误输出所造成后果的严重程度与可逆性。这是本模型区别于传统框架的核心维度。

- **高**：错误可被下游轻易发现与纠正，后果限于返工成本。如营销文案生成、内部知识问答。
- **中**：错误可能造成经济损失或客户体验损害，但可通过复核机制拦截。如合同条款审查、客服工单处理。
- **低**：错误可能危及人身安全、资金安全或行政合法性，且难以事后追回。如临床诊断结论、授信审批决定、行政处罚决定。

维度三：价值可计量性

指收益能否被清晰归因并转化为财务或运营指标。这一维度直接决定项目能否通过内部立项与验收。

- **高**：存在明确基线指标，改善可直接换算为人力、成本或收入。如质检漏检率、客服一次解决率、内容生产单件成本。
- **中**：收益真实存在但与其他变量耦合，需设计对照组或分阶段观测。如研发提效、营销转化提升。
- **低**：收益体现为体验改善、风险降低或能力储备，短期难以量化。如员工学习效率提升、品牌形象改善。

表 2-1 SVA-3D 优先级判定规则

数据可及性	容错空间	价值可计量性	判定结论与建议动作
高	高	高	首批必做。 应在 3—6 个月内上线并产出可对外汇报的成果，用于建立组织内部信心。
高	高	中/低	快速试点。 适合作为能力练兵场，但不宜作为汇报主线，需提前设计观测口径。
高	中/低	高	重点投入，配套建设。 价值大但风险高，必须同步建设复核机制、权限边界与审计留痕，投入周期 6—18 个月。
中	高	高	先治数据。 场景本身成立，瓶颈在数据治理；应把首期预算的多数投向数据工程而非模型。
中/低	中/低	高	分步拆解。 不要整体推进，应拆出其中数据完备、容错较高的子环节先行，逐步扩展。
低	低	低/中	暂缓。 当前投入产出比不成立，宜保持技术跟踪，等待外部条件变化。

2.2 L0—L4：场景成熟度阶梯

SVA-3D 用于决策“做不做”，成熟度阶梯用于描述“现状如何”。本白皮书每张场景卡片都标注了该场景的行业整体成熟度。

表 2-2 场景成熟度五级阶梯

级别	名称	判定标准	对采购决策的含义
L0	概念验证	仅有实验室演示或概念性宣传，无真实业务数据支撑	不要买。演示效果与真实效果无关
L1	单点试点	在单一部门、产线或业务单元跑通，样本量有限	可共建，不可大规模采购；要求供应商披露试点样本量
L2	单位内推广	在同一组织内跨部门或跨厂区复制成功，形成标准化交付物	可小规模采购，但需自建运维能力
L3	跨组织复制	不同企业采用相同方案均获成功，形成可采购的成熟产品或服务	应当采购而非自建。 价格与交付周期可事先议定
L4	行业基础设施	成为行业默认配置，不采用即形成劣势	不再作为“AI 项目”讨论，按常规 IT 采购流程处理
三道鸿沟 L1→L2 是 可复制性鸿沟 ：跨过它需要标准化交付物与运维体系，而不是更强的模型。 L2→L3 是 商业化鸿沟 ：跨过它需要产品化与定价能力，多数企业自建团队止步于此。			

级别	名称	判定标准	对采购决策的含义
L3→L4 是 制度化鸿沟 ：跨过它通常需要行业标准、监管要求或头部企业示范共同推动。 由此得到一条实用结论： 在 L3 级场景上采购，在 L1—L2 级场景上自建 。把自研资源投向已经商品化的场景，是国内不少企业 AI 预算浪费的主要来源。			

2.3 价值四层：收益该怎么考核

"提效"是一个被过度使用的词。本白皮书把 AI 带来的价值改善拆为四层，不同层次对应完全不同的立项逻辑与考核方式。

表 2-3 AI 价值四层分类

层次	名称	含义	典型场景	考核方式
第一层	替代型	用 AI 承担原本由人完成的重复性动作，流程本身不变	视觉质检替代目检、商品图生成、标准工单分类	直接对比人力工时与差错率
第二层	增强型	人仍是决策主体，AI 提供更快更全的输入，提升决策质量	智能投研辅助、影像辅助阅片、合同风险提示	对比决策周期与决策质量指标
第三层	重构型	流程被重新设计，原有环节被合并或取消	预测驱动的自动补货、端到端智能客服闭环、一网通办智能审批	对比端到端周期与总人力配置
第四层	创造型	产生原本不可能提供的产品或服务	千人千面学习路径、全天候基层诊疗辅助、24 小时数字人直播	对比新增业务量或覆盖人群

这套分类有一个重要的实践含义：**四层的投入产出比是递增的，但落地难度也是递增的**。第一层通常在半年内可验证，第四层往往需要三年以上。多数组织失败的原因，是用第一层的预算与耐心去追求第三、四层的效果。

2.4 提效账本模型：省下的东西怎么变成钱

这是本白皮书新增的第四个工具，用于回答一个几乎所有 AI 项目都会被追问的问题：“你说省了 30%，那到底省下了多少钱？”

本研究院观察到，AI 项目的收益之所以难以入账，是因为节省往往分散在四种性质完全不同的“节省”里，而只有其中两种能直接进损益表。

表 2-4 四类节省及其入账方式

节省类型	典型表现	能否直接入账	计量方式
外采成本节省	不再需要外部拍摄、外部律师、外包标注、外部翻译	能，且最直接	直接对比采购合同金额；是最容易被财务认可的收益
产能扩张	同样人数处理更多业务量、新增 24 小时服务时段	能，体现为收入侧	对比单位时间处理量×单位业务毛利
人力时间释放	员工每天少花 2 小时写报告	通常不能 (除非编制真的减少)	只能作为"可再投入产能"记录；直接折算为工资节省会被财务质疑
风险成本降低	差错率下降、合规风险下降、事故率下降	间接，需长周期观测	用历史差错造成的损失×差错率下降幅度估算

账本模型的三条使用规则

规则一：先算外采与产能，再算人力与风险。 立项材料中如果只有"节省人力工时"一项收益，通过财务评审的难度很大。优先寻找有外采支出可替代的场景（这也是电商内容生产、法务外部律师费这类场景最容易立项的原因）。

规则二：人力释放要写清楚"释放去做什么"。 "客户经理每天少花 3 小时做材料"本身不是收益，"因而人均日处理业务量提升 60%"才是收益。前者是投入的减少，后者才是产出的增加。

规则三：把投入侧也列全。 多数公开案例只披露收益。真实的账必须包含：模型与算力费用、数据准备与标注、系统集成与改造、培训与变更管理、持续运维（通常被低估）。附录 D 提供了完整的测算模板与隐性成本清单。

第三章 六大行业与获客职能的场景卡片

本章是全书的主体，共 46 张场景卡片。前六节按行业划分：制造业、电商与零售、金融业、医疗健康、教育、政务与公共服务；第七节按职能划分——"自媒体获客与创始人 IP"是一件几乎所有行业都在做、且做法高度趋同的事，因此单独成节。每张卡片格式统一，可以直接翻到对应位置对照使用。

阅读前请注意三件事：

- 1. 对照表中"传统做法"一列，请替换为你自己的真实数据。表中填的是行业常规水平，你的组织可能更高或更低。差距成立与否，取决于你自己的基线。
- 2. "落地前提"一栏比对照表更重要。前提不具备时，表中的效果不会出现——这是本白皮书从公开案例的成功与失败中反复验证的结论。
- 3. 成熟度标注的是行业整体水平，不是领先企业水平。具体到单个企业，实际位置可能显著高于或低于标注。

3.1 电商与零售

电商零售是本白皮书六个行业中 AI 渗透最深的一个，也是唯一在"内容生产"环节已经完成成本结构改写的行业。政策侧，2026 年 6 月商务部等八部门印发《关于加快"人工智能+消费"发展的实施意见》（商建发〔2026〕89 号），把零售、电子商务和物流配送的智能化转型列为重点；研究侧，中国信通院产业与规划研究所与中国信息消费推进联盟 2026 年 7 月发布的《"人工智能+消费"发展研究报告（2026 年）》提出消费提效期—消费重构期—消费共创期的三阶段路径。

但对企业而言，更有用的观察是：**这个行业里不同环节的改善幅度相差极大**。内容生产环节的成本已经降了一个数量级，而库存与履约环节的改善幅度通常在 30%—50%区间。下面八张卡片按改善幅度从大到小排列。

场景 3.1.1 商品图与模特图生成（AI 商拍）

成熟度 L4 | 价值层次 替代型。 这是目前所有行业中成本结构改写最彻底的场景。原来需要"找模特+找场地+摄影团队+后期修图"才能产出的一张场景图，现在由一张平铺产品图加一段文字描述即可生成，单张成本从百元至千元量级降到元级，出片周期从周降到分钟。

关键指标	传统做法	引入 AI 后	变化幅度
单张成片直接成本	摄影约 500 元/张； 模特费约 2000 元/次； 两天拍摄合计约 1.6 万元	百元以内可生成数十款商品图	约 90%降幅
出片周期	排期+拍摄+修图， 通常以周计	分钟至小时级	数量级压缩
团队规模	30—40 人内容团队 （含摄影、模特协调、修图、文案）	1—2 人操作+少量审校	约 95%人力压缩

关键指标	传统做法	引入 AI 后	变化幅度
场景灵活度	换场景＝重新拍摄， 异地实景成本再翻数倍	换场景＝换一段描述	边际成本趋近于零
详情页文案	外包撰写约 300—500 元/条	批量生成，单条成本可忽略	约 90%以上降幅

公开案例与数据

- 据公开报道，服务安踏、施耐德电气、比音勒芬、水星家纺、朗姿等品牌的 AI 内容服务商披露：传统方式下撰写一条电商详情页需 300—500 元、拍摄一张照片约 500 元、模特费约 2000 元/次，两天拍摄约 1.6 万元；改用 AI 后百元以内可生成数十款商品图，1—2 人可完成原 30—40 人团队的工作量，**综合成本降幅可达约 90%**。该服务商同时披露其 AI 替代率已从创业初期的 30%—40%提升至 90%—95%。
- 同一报道提及，某头部运动鞋服品牌在应用后月营收额外增加"两三千万"元；相关 MCN 业务的月度 GMV 从 2024 年 6 月的约 1099 万元增至约 3000 万元。（注：营收增长受多重因素影响，不宜全部归因于 AI）

落地前提

- **产品平铺图的拍摄质量仍需保证。** AI 改写的是"场景与模特"这部分成本，产品本身的形态、材质、颜色仍需真实素材作为输入。省掉产品拍摄这一步是最常见的错误。
- **建立品牌视觉规范库。** 生成模型需要明确的风格约束（色调、构图、模特风格、背景类型），否则产出物在同一店铺内视觉不统一，反而损害品牌感。
- **建立审校环节与不合格清单。** 高频问题包括：手指与肢体结构异常、logo 变形、面料纹理错乱、配饰与产品不符。应形成书面的检查清单交给审校岗。
- **核对平台规则。** 各电商平台对 AI 生成图的标注要求、对"虚假宣传"的界定存在差异且在持续更新，上架前须核对当期规则。

常见坑

- **用 AI 图呈现产品本身不具备的特性。** 例如把实际不挺括的面料生成出挺括效果，会直接推高退货率与差评率，是把成本从内容端转移到了售后端。这是本场景最严重也最常见的失败方式。
- **只算生成成本，不算审校成本。** 若审校环节仍需人工逐张检查，实际人力节省会显著低于宣称值。应在试点期记录"每百张图的审校工时"。
- **一次性替换全部内容团队。** 建议保留具备视觉判断力的核心人员转做审校与风格把控——这部分能力恰恰是 AI 最难替代的。
- **忽视肖像权与版权风险。** 生成的人物形象若与真实人物高度相似，或训练素材来源不清晰，存在法律风险。应要求供应商就素材来源作出书面承诺。

适用与规模判断：几乎所有 SKU 数量在数百以上、且有视觉展示需求的零售商家。SKU 越多、上新越频繁、场景越多样，收益越大。服饰、家纺、家居、美妆是收益最高的品类；食品、生鲜等对真实感要求极高的品类需更谨慎。

场景 3.1.2 数字人直播

成熟度 L3 | 价值层次 替代型+创造型。这个场景的价值不只是"比真人主播便宜", 更在于**打开了原本不存在的时段**——凌晨到清晨的长尾流量, 此前因为人力成本不成立而被整体放弃。

关键指标	传统做法	引入 AI 后	变化幅度
单场直播人力	主播+助播+场控+运营, 通常 3 人以上	1 名运营可同时管理多个直播间	人力大幅压缩
可开播时长	受主播体力与排班限制, 通常每日 6—12 小时	7×24 小时持续开播	时段覆盖翻倍以上
边际开播成本	每增加一个时段 需增加一班人力	接近于零	边际成本结构改变
单场转化效率	真人主播在互动、 应变、信任建立上占优	标品、复购品、低决策成本品类 差距较小	因品类而异, 非普遍优于真人
内容一致性	依赖主播状态, 话术合规性需逐场把控	话术库统一, 合规性可前置校验	显著提升

公开案例与数据

- 据 2026 年数字人电商直播相关白皮书披露, 数字人直播已成为电商零售行业的"标配能力", 呈现规模化普及态势, 深度渗透至各规模商家及主流品类; 国内 AI 数字人市场 2024 年增速达 85.3%, 布局企业超 114 万家。典型模式包括京东与联想的"人机协同、日不落直播间", 以及京东与东阿阿胶在全时覆盖下的高客单转化探索。
- 公开材料普遍将数字人直播的核心价值描述为"轻量化数据驱动"与"极低边际成本的 7×24 小时开播", 而非单场 GMV 超越真人。

落地前提

- **先想清楚要打的是哪个时段。**数字人最确定的收益来自真人不开播的时段（深夜至清晨）。用数字人替换黄金时段的真人主播, 通常得不偿失。
- **话术库比形象更重要。**观感自然度已不是主要瓶颈, 转化率的差距主要来自话术设计、商品讲解逻辑与互动应答的覆盖度。应把预算重点放在话术工程而非形象定制。
- **接入实时互动与兜底转人工。**完全无互动的循环播放会被平台限流, 也无法承接高意向用户。需设计"高意向弹幕触发转人工"的机制。
- **核对平台合规要求。**各直播平台对虚拟形象标识、AI 主播备案、话术合规的要求不同且在变化, 须逐平台确认。

常见坑

- **把数字人当成"省下主播工资"的工具。**若只做替代而不做时段扩展, 收益有限且容易伤害既有转化。正确的定位是"增量时段"而非"存量替代"。
- **忽视流量分配规则。**平台对低互动直播间的流量倾斜有限, 无人值守的数字人直播间常出现"开播时长很长但曝光极低"的情况。开播时长本身不产生价值。
- **用未授权的真人形象。**使用他人形象生成数字人存在明确的法律风险, 须取得书面授权并留存。

- **对高客单、高决策成本品类期望过高。** 珠宝、大家电、教育课程等需要建立信任的品类，数字人目前的转化能力与真人差距明显。

适用与规模判断：标品、快消、复购型商品的中大型商家收益最明显。SKU 单价低、决策链路短、有明显长尾时段流量的类目优先。单店月 GMV 过低时，运营与话术维护成本可能吃掉全部收益。

场景 3.1.3 智能补货与销量预测

成熟度 L2—L3 | 价值层次 重构型。 这是零售业真正的利润所在，也是最难做的场景。它改变的不是某个岗位的效率，而是“缺货”与“压货”这对结构性矛盾——而这对矛盾直接吃掉零售企业的毛利。

关键指标	传统做法	引入 AI 后	变化幅度
缺货率	人工经验补货， 门店缺货率最高可达约 7%	降至 2% 以内	下降约 70%
库存周转天数	最高约 35 天	普遍 21 天以内	缩短约 40%
补货管理人效	数百家门店需 相应规模的补货团队	近 30 人管理近千家门店	人效大幅提升
年度人力成本	—	年节约约 5000 万元	可直接入账
决策频率	按周或按月人工盘点决策	按日甚至按小时自动生成补货单	响应速度数量级提升

公开案例与数据

- 据公开案例披露，多点（Dmall）的 AI 自动补货方案通过“智能预测—智能补货—订单可视—零供协同”流程，将合作商超的**缺货率从最高约 7% 降至 2% 以内，库存周转天数从最高约 35 天降至普遍 21 天以内**；管理端实现近 30 人管理近千家门店，年节约人力成本约 5000 万元。
- 该案例同时指出，传统人工补货的典型症状是“畅销品缺货与滞销品高库存并存”——这正是经验决策无法同时优化两端的结果。

落地前提

- **至少需要 12—24 个月的完整销售历史，且要覆盖一个完整的季节周期。** 数据不足时模型无法学到季节性与节假日规律，效果会显著劣于有经验的采购人员。
- **商品主数据必须先治理干净。** 同品多码、单位不一致、包装规格混乱是补货模型失效的首要原因，且这类问题在多渠道经营的企业中极为普遍。
- **要打通库存、销售、在途、促销四类数据。** 只有销售数据的预测模型无法处理促销带来的需求跃变，而促销恰恰是零售最重要的变量。
- **需要与供应商协同。** 补货建议再准确，若供应商交付周期不稳定也无法兑现。这一场景的收益上限由供应链协同能力决定。

常见坑

- **只看预测准确率，不看业务指标。** 预测准确率提升 5 个百分点，可能对缺货率毫无影响。考核指标必须是缺货率与周转天数，而不是模型精度。

- **忽略新品与长尾商品。** 模型在有历史数据的商品上表现良好，在新品上几乎无能为力。必须保留人工决策通道，并明确划分模型负责的商品范围。
- **促销期直接套用日常模型。** 大促期间需求分布与日常完全不同，必须单独建模或人工接管，否则大促当天的缺货会抵消全年的收益。
- **低估门店执行的偏差。** 系统生成的补货单如果门店不执行或随意修改，全部收益归零。上线时须同步建立执行率监控。

适用与规模判断：门店数在数十家以上、SKU 数在数千以上的连锁零售、商超、便利店。生鲜品类因损耗成本高，单位收益最大。单店或少量门店的商家，人工经验通常已经足够。

场景 3.1.4 智能客服与售后工单处理

成熟度 L3—L4 | 价值层次 替代型→重构型。 从"帮客服写回复"（助手）到"独立完成整单退换货处理"（智能体），这个场景正在跨越本白皮书第一章所说的形态分界线，收益的计量方式也随之从"节省工时"变为"自动闭环单量"。

关键指标	传统做法	引入 AI 后	变化幅度
咨询响应时间	排队等待，高峰期数分钟	秒级响应，无排队	等待消除
服务时段	受排班限制	7×24 小时	全时段覆盖
标准问题解决	需人工逐条应答	自动闭环处理，人工只处理例外	人工干预率显著下降
会话质检覆盖率	人工抽检，通常仅覆盖百分之几	全量自动质检	从抽检到全检
多语言支持	按语种配置人力，小语种成本极高	统一模型覆盖多语种	跨境场景成本大幅下降

公开案例与数据

- 据沙丘智库《2026 年中国银行业大模型应用跟踪报告》，智能客服是金融与零售场景中落地最早、覆盖最广的方向之一；电商侧，AI 客服已从"关键词匹配的机器人"演进为"可调用订单、物流、库存系统完成处置动作的智能体"。
- 本研究院提示：该场景公开披露的具体数字口径差异极大（有的统计"机器人接待量占比"，有的统计"完全无需人工介入的比例"），跨案例比较意义有限，建议以自身基线做前后对照。

落地前提

- **先把知识库与规则治理干净。** 效果瓶颈几乎总是在知识库——过期政策、多版本并存、口径冲突。这部分工作量通常占项目总投入的一半以上。
- **明确划定 AI 可自主执行的动作边界。** 查询类可全放开；退款、改地址、补发等涉及资金或履约的动作，须设金额阈值与二次确认。
- **保留清晰的转人工入口。** 隐藏转人工按钮以抬高自动率，是最容易引发投诉与舆情的做法。

- **建立"AI 答错"的回收机制。** 每一次人工修正都应回流为知识库更新的输入，否则同样的错误会持续发生。

常见坑

- **用"机器人接待率"作为 KPI。** 这个指标可以通过让用户找不到人工入口来轻松刷高，与真实服务质量负相关。应改用"一次解决率"与"服务后满意度"。
- **在情绪激烈的会话中坚持由 AI 处理。** 投诉、维权、纠纷类会话应快速识别并转人工，AI 的标准化应答在此类场景中会显著激化矛盾。
- **忽视对话记录的合规留存。** AI 承担的服务承诺同样具有法律效力，会话记录须完整留存备查。

适用与规模判断：日咨询量在千次以上的电商、零售、服务型企业。咨询量越大、问题标准化程度越高、跨时区经营（跨境）的场景收益越大。

场景 3.1.5 跨境本地化与多语言内容

成熟度 L3 | 价值层次 替代型。 跨境电商的语言成本原本随目标市场数量线性增长——每进入一个新市场就要增加一套翻译与本地化人力。AI 把这条线性曲线压平了。

关键指标	传统做法	引入 AI 后	变化幅度
单语种商品信息翻译	人工翻译按字计费，专业本地化成本更高	批量机器翻译+人工校审	成本大幅下降
新市场上线周期	招募译员+逐条翻译，通常以月计	批量生成后校审，以天计	周期数量级压缩
可覆盖语种数	受译员招募能力限制，小语种尤难	模型覆盖的语种均可尝试	长尾市场变得可行
本地化质量	专业译员质量高但成本高	常见语种接近可用，小语种仍需人工把关	质量与语种强相关

公开案例与数据

- 本场景在跨境电商行业已属基础配置。公开材料普遍将其价值描述为"让原本因语言成本不成立的小语种市场变得可尝试"，即价值主要体现在市场覆盖的扩展，而非既有市场的成本节省。

落地前提

- **建立术语表与品牌词库。** 品牌名、型号、专有名词的翻译必须固定，否则同一商品在不同页面出现不同译法，直接影响搜索与信任。
- **区分"可机翻直发"与"必须人工校审"的内容类型。** 商品参数可直发；营销文案、售后政策、法律条款必须人工校审。
- **注意合规文本的属地要求。** 各国对商品标签、成分说明、警示语的法定表述有强制要求，不可依赖翻译生成。

常见坑

- **直译文化敏感内容。** 颜色、数字、图案、节日相关的营销内容在不同市场含义差异极大，机翻无法处理。
- **忽视本地搜索词习惯。** 直译的商品标题往往不包含当地消费者实际使用的搜索词，导致翻译“正确”但没有流量。

适用与规模判断：已进入或计划进入三个以上语种市场的跨境卖家。单一英语市场的卖家收益有限。

场景 3.1.6 AI 导购与对话式购物

成熟度 L2—L3 | 价值层次 重构型。 这是本节中**最不确定但潜在影响最大**的场景。它改变的可能不是效率，而是消费者触达商品的路径——即流量分发的入口本身。

关键指标	传统做法	引入 AI 后	变化幅度
选购决策路径	搜索→筛选→比价→详情页→下单	描述需求→直接获得推荐方案	路径重构
复杂需求满足	需消费者自行拆解需求并多次搜索	一次描述即可综合多维约束	显著改善
对商家的影响	流量通过关键词与付费投放分配	流量可能通过智能体的理解与引用分配	分配逻辑可能改变
当前转化贡献	—	尚无稳定可比的公开数据	价值可计量性低

公开案例与数据

- 多方研究观察到 2026 年 AI 购物助手使用出现明显加速，行业内有“AI 购物普及元年”的提法。中国信通院相关研究将其归入“消费重构期”的典型形态。
- 本研究院提示：该场景目前缺乏可靠的转化率对照数据，公开材料多为趋势判断而非效果验证。

落地前提

- **优先做“被理解”而不是“自建入口”。** 对绝大多数品牌与商家而言，当前更务实的动作是让自己的商品信息结构化、属性完整、描述准确，使其能被各类智能体正确理解和引用。
- **梳理商品结构化数据。** 属性缺失、参数不规范的商品在对话式检索中会被系统性忽略。
- **自建导购智能体前先算清楚流量来源。** 没有自有流量的商家自建导购入口，通常无人使用。

常见坑

- **把资源投向自建对话入口而非商品数据治理。** 这是当前该场景最普遍的资源错配。
- **用现有指标体系衡量该场景。** 它改变的是流量结构，而流量结构的变化在早期难以被现有指标捕捉，据此得出“效果不好”的结论可能为时过早。

适用与规模判断：平台型企业与拥有大量自有流量的头部品牌可考虑自建；中小商家应聚焦商品数据质量，保持观察。

场景 3.1.7 投放素材生成与智能出价

成熟度 L3 | 价值层次 增强型。 广告投放的核心矛盾是"素材衰减"——同一素材投放数日后点击率必然下降，而人工产出速度跟不上衰减速度。AI 把素材供给从瓶颈变成了非瓶颈。

关键指标	传统做法	引入 AI 后	变化幅度
素材日产能	设计师人均每日数条至十余条	批量生成，数量级提升	供给瓶颈解除
A/B 测试规模	受素材数量限制，同时测试组数有限	可同时测试大量变体	测试效率提升
素材衰减应对	发现衰减后需数日才能补充新素材	可近实时补充	响应周期压缩
单条素材成本	设计与拍摄成本	接近于零	大幅下降

公开案例与数据

- 该场景与 3.1.1 商品图生成技术同源，在广告投放场景中的应用已相当普遍。公开材料普遍将其价值定位为"解除素材供给瓶颈"，从而使高频测试与快速迭代成为可能。

落地前提

- **必须配套完善的投放数据回流机制。** 素材产能提升的价值，完全取决于能否快速识别哪些素材有效。没有归因体系，大量素材只是浪费预算。
- **建立素材合规审查前置。** 广告法对功效宣称、绝对化用语、比较广告有明确限制，生成内容须过合规校验后才可投放。

常见坑

- **用素材数量代替素材质量思考。** 大量同质化变体不会改善效果，只会增加测试成本。变体应围绕明确的假设设计。
- **忽视广告合规风险。** 生成内容中出现的功效宣称与绝对化用语是监管处罚的高发区。

适用与规模判断：有持续投放预算、且投放渠道支持快速素材更换的商家。投放预算规模越大，素材迭代速度的边际价值越高。

场景 3.1.8 门店客流分析与陈列识别

成熟度 L3 | 价值层次 增强型。 线下零售长期缺少与线上等价的"行为数据"——顾客看了什么、停留多久、为什么没买。视觉 AI 把这部分数据补上了。

关键指标	传统做法	引入 AI 后	变化幅度
客流统计	人工计数或红外计数，仅有总量	分时段、分区域、分动线的完整数据	数据颗粒度大幅提升

关键指标	传统做法	引入 AI 后	变化幅度
陈列合规检查	督导巡店抽查，覆盖频次低	摄像头自动识别，可每日核查	覆盖率从抽检到全检
货架缺货发现	依赖店员巡视	自动识别并告警	发现时效大幅提升
人员排班依据	经验判断	基于分时客流数据	人效可优化

公开案例与数据

- 视觉客流分析与货架识别在连锁零售中已属成熟应用。公开材料显示其主要价值在于把线下经营从“经验决策”转为“数据决策”，尤其在陈列合规与排班优化两个环节的收益较为确定。

落地前提

- **摄像头点位与角度需专门规划。** 沿用安防摄像头通常无法满足客流与陈列识别的角度要求，改造成本须计入预算。
- **必须处理个人信息合规问题。** 涉及人脸的采集与处理受严格法律约束，建议采用不采集人脸的客流统计方案，并做好告知与标识。
- **数据要接入具体决策。** 只做看板不改排班、不改陈列，该场景不产生任何价值。

常见坑

- **采集人脸做顾客识别。** 这一做法的合规风险极高，且在多数场景中并非必需。
- **只上系统不改流程。** 客流数据若不进入排班与陈列的决策流程，投入即沉没。

适用与规模判断：门店数在数十家以上的连锁零售、餐饮、专卖店。单店面积大、动线复杂、陈列标准化程度高的业态收益更大。

电商零售的一个自检问题

本节八个场景中，前两个（内容生产、数字人直播）已经被广泛采用，第三个（智能补货）才是真正影响利润表的。

因此本研究院建议零售企业用一个问题自检：**过去 12 个月，你的库存周转天数改善了多少？**

如果答案是“没有明显变化”，那么当前的 AI 应用仍停留在内容与客服这类“看得见但不进利润表”的环节。内容成本降 90% 听上去很大，但对多数零售企业而言，内容成本在总成本中的占比远低于库存与履约成本。

3.2 制造业

制造业是"人工智能+"政策体系中指标最密集的领域。2026 年 1 月工业和信息化部等八部门印发《"人工智能+制造"专项行动实施意见》，提出到 2027 年推动 3~5 个通用大模型深度应用、推出 1000 个高水平工业智能体、打造 100 个工业领域高质量数据集、推广 500 个典型应用场景、选树 1000 家标杆企业。

从实践看，制造业 AI 呈现明显的"哑铃结构"。一份基于 167 个已验证案例的行业白皮书显示：案例在生产制造环节占 62.3%（104 例），研发设计占 21.6%，而供应链仅占 5.4%、销售不足 2%；技术类型上，传统机器学习（约 46.7%）与机器视觉（约 20.4%）仍是落地主力，AI 智能体占比约 3%。

这个结构说明了什么

供应链环节的价值密度并不低于生产环节，为什么落地案例只有 5.4%？

答案不在技术，在**数据可及性**：生产环节的数据产生于企业自有设备，可控可得；供应链数据分散在上下游多个法人主体手中，汇聚需要商业谈判而非技术投入。

按 SVA-3D 判断，供应链场景属于"价值高、容错中、数据可及性低"的组合，对应动作是"先治数据"。这也解释了为什么供应链 AI 被普遍视为下一阶段的蓝海，但落地节奏取决于产业链协同机制，而非算法进步。

对企业的直接含义：如果你现在就想做供应链 AI，先问自己能否拿到上下游的数据。拿不到，这个项目的天花板已经确定。

场景 3.2.1 表面缺陷视觉质检

成熟度 L3—L4 | 价值层次 替代型。这是制造业投入回收周期最短、账最好算的场景，也是绝大多数制造企业 AI 能力建设的正确起点。它替代的是一项人眼无法长期稳定完成的工作——连续数小时高强度目视检查。

关键指标	传统做法	引入 AI 后	变化幅度
漏检率	人工目检受疲劳影响，后半班次显著上升	公开案例披露可降至 0	关键质量指标改善
过检率（误报）	传统机器视觉需大量人工二次复检	公开案例披露稳定在 0.5%以内	复检工作量大幅下降
检测一致性	因人而异、因时段而异	标准统一，不受疲劳影响	波动消除
质检人力角色	全员逐件目检	从"主力"转为"复核"，干预越来越少	岗位性质改变
回本周期	—	公开案例披露约 4—6 个月（中小企业投入 5 万—30 万元，回本 2—6 个月）	普遍在一年以内

公开案例与数据

- **施耐德电气（公开案例）：**AI 视觉检测方案实现漏检率降至 0、过检率稳定在 0.5%以内，人员角色从检验"主力"转为"候补"；已在北京、上海、天津等地的 **10 个工厂、20 条生产线**部署。其规模化方法值得注意——优先选择有强烈质量提升需求的工厂，在同类型产线横向推广以降低试错成本，并采用"中央平台+分布式架构"实现跨工厂复用。
- **长虹（公开案例）：**AI 质检年节省成本达千万元级，**回本周期约 4—6 个月**。
- **中小制造企业（公开案例汇总）：**轻量化 AI 质检的投入区间约 5 万—30 万元，回本周期约 2—6 个月。

落地前提

- **缺陷样本是第一瓶颈，不是模型。** 需要收集足量的缺陷样本，尤其是低频缺陷类型。多数项目在这一步耗时最长。若缺陷率本身极低，需考虑合成数据或长期采集。
- **光源与工位改造必须计入预算。** 成像质量决定效果上限，而成像质量取决于光源、镜头、工位固定方式。这部分硬件与工程改造成本在不少项目中超过软件成本。
- **先与一线约定可接受的过检率上限。** 过检（误报）意味着一线要做额外复核。若过检率过高，系统会被绕过或关闭。
- **试点产线选"中位水平"而非最先进产线。** 在最好的产线上试点会得到过于乐观的结论，无法反映复制难度。

常见坑

- **用大语言模型做本该由机器视觉解决的问题。** 本场景的主力技术是机器视觉与传统机器学习，不是大模型。技术选型错位会导致成本高且效果不稳定。
- **忽视产品换型带来的模型失效。** 每次产品或工艺变更都可能需要重新采样与训练，运维成本必须计入长期账。
- **只算质检员工资，不算质量成本。** 本场景更大的收益往往在流出不良品的减少（客诉、退货、返工、品牌损失），这部分应纳入测算。
- **低估标注工作量。** 缺陷标注需要有经验的质检员参与，占用的是稀缺人力，这项隐性成本常被忽略。

适用与规模判断：有稳定批量生产、缺陷类型可视化、且质检环节人力占比明显的制造企业。产量越大、缺陷成本越高、产品换型越少，收益越确定。

场景 3.2.2 设备预测性维护

成熟度 L3 | 价值层次 替代型+增强型。 从"坏了再修"或"定期换件"转向"快坏之前修"。收益的大头不在维修成本，而在**非计划停机的减少**——对连续生产型企业，一次非计划停机的损失可能超过全年维护预算。

关键指标	传统做法	引入 AI 后	变化幅度
维护策略	事后维修或按固定周期预防性更换	按实际状态与预测寿命维护	策略重构
非计划停机	突发故障导致，	提前预警，可安排在计划停机窗口	主要收益来源

关键指标	传统做法	引入 AI 后	变化幅度
	损失以停产时长计		
备件库存	为应对突发需保有较高安全库存	按预测需求备货	占用资金下降
过度维护	未到寿命的部件被定期更换	按实际状态判断	耗材成本下降

公开案例与数据

- **通威（公开案例）**：部署 50 余个机器学习模型用于生产优化，公开披露**降本约 37%**。
- **南钢（公开案例）**：峰谷发电 AI 调度，公开披露相关效益提升约 4.24 倍。
- 本研究院提示：预测性维护的收益高度依赖设备类型与停机成本，不同企业间差异极大，上述数据不宜横向比较。

落地前提

- **需要足量的故障历史数据**。这是本场景最大的门槛——运行良好的设备恰恰缺少故障样本。数据积累通常需要 12 个月以上。
- **老旧设备的数据采集改造成本必须前置评估**。加装传感器、打通不同品牌 PLC 协议的成本，在多数项目中超过算法开发成本。这是预测性维护项目最常见的预算超支来源。
- **先算清楚单次非计划停机的损失**。如果这个数字算不出来，本场景的 ROI 就无法论证。
- **要与维修排程系统打通**。预警若不能转化为工单与排程，就只是一个告警灯。

常见坑

- **在停机成本低的设备上做预测性维护**。对可快速更换、有冗余、停机影响小的设备，本场景的投入产出比通常不成立。
- **告警过多导致“狼来了”**。误报率过高会使维护人员逐渐忽略告警，系统实际失效。上线初期应设较保守的阈值。
- **只做模型不做流程**。预警的价值取决于响应机制，没有明确的响应责任人与处置流程，预警等于噪声。

适用与规模判断：连续生产型企业（钢铁、化工、水泥、造纸）、设备停机成本高的离散制造企业、以及大型公用设施。设备数量少、停机可容忍的企业优先级较低。

场景 3.2.3 工艺参数寻优

成熟度 L2—L3 | 价值层次 增强型。制造业中大量工艺参数依赖老师傅的经验设定，且往往停留在“能用”而非“最优”。AI 在参数空间中的搜索能力，可以找到人工经验难以触达的组合。

关键指标	传统做法	引入 AI 后	变化幅度
参数设定依据	老师傅经验+历史沿用	基于历史数据与实时反馈的模型优化	决策依据改变

关键指标	传统做法	引入 AI 后	变化幅度
优化维度	人工只能同时调整少数参数	可同时优化数十个参数的组合	搜索空间扩大
良率与能耗	稳定但存在优化空间	公开案例显示可有可观改善	因工艺而异
经验传承	依赖师徒传授， 人员流动即流失	沉淀为模型与数据资产	知识资产化

公开案例与数据

- **通威**（公开案例）：以 50 余个机器学习模型驱动生产优化，公开披露降本约 37%，属于本场景的典型应用。
- **海尔**（公开案例）：全链路 AI 应用公开披露年增效超 1.1 亿元、库存周转提升约 60%，其中工艺与生产优化是重要组成部分。

落地前提

- **需要能够安全试错的机制。** 参数优化必须能在真实产线上验证，而产线试错有成本与风险。通常需要先建立仿真环境或在小批量试产中验证。
- **工艺数据与质量数据必须能够对齐。** 即某一批次的参数设定要能追溯到该批次的质量结果。这一点在很多企业的信息化现状下并不成立。
- **需要工艺专家深度参与。** 模型给出的参数组合必须由工艺专家判断其物理合理性，否则可能给出统计上有效但工艺上危险的建议。

常见坑

- **把模型建议直接下发到产线执行。** 工艺参数直接影响产品安全与设备安全，必须保留人工确认环节。
- **忽视原材料批次差异。** 同样的参数在不同批次原料上效果不同，模型若未纳入原料特性会不稳定。
- **用短期良率提升掩盖长期设备损耗。** 某些参数组合短期提升良率但加速设备磨损，评估周期需足够长。

适用与规模判断：工艺参数多、良率或能耗有明显优化空间、且有数据采集基础的流程工业与精密制造。工艺简单或已高度标准化的行业收益有限。

场景 3.2.4 能耗与峰谷用电优化

成熟度 L2—L3 | 价值层次 增强型。 这是制造业中**价值可计量性最高**的场景之一——收益可以直接换算成电费，无需任何归因争议。在电价波动与峰谷价差扩大的背景下，其价值还在上升。

关键指标	传统做法	引入 AI 后	变化幅度
用电调度依据	按生产计划用电， 较少考虑电价时段	在满足生产约束下 优化用电时段分布	成本结构优化
能耗监测粒度	按厂区或车间总表	按设备、按工序	颗粒度提升

关键指标	传统做法	引入 AI 后	变化幅度
异常能耗发现	月度对账时发现	实时告警	发现时效提升
收益计量	—	直接换算为电费节省	可直接入账

公开案例与数据

- 南钢（公开案例）：峰谷发电 AI 调度，公开披露相关效益提升约 4.24 倍。
- 跨行业参照——中国电信在基站与机房侧的 AI 节能系统公开披露年均节电约 8 亿度、节约电费约 5.2 亿元，其技术逻辑（实时监控设备运行与环境信息、智能关断闲置设备、智能调节制冷系统）对制造业厂务能耗管理同样适用。

落地前提

- 需要分项计量的电表体系。只有总表数据无法定位优化空间，分项计量改造是前置投入。
- 生产约束必须准确建模。优化用电时段不能牺牲交付。约束条件建模不准确，会给出无法执行的方案。
- 了解本地电价政策与需求响应机制。峰谷价差、需量电费、可中断负荷补贴等政策直接决定收益规模。

常见坑

- 为省电费牺牲设备寿命或产品质量。频繁启停部分设备的隐性成本可能高于电费节省。
- 只做监测看板不做调度优化。能耗看板本身不省电，必须连接到实际的调度动作。

适用与规模判断：电费在总成本中占比较高的高耗能行业（钢铁、有色、化工、建材、玻璃），以及大型厂区的厂务系统。

场景 3.2.5 工业知识问答与工艺检索

成熟度 L2—L3 | 价值层次 增强型。制造企业积累了大量的工艺文件、设备手册、故障处理记录、质量案例，但它们分散在文件服务器与老员工的记忆里。这个场景把“找资料”的时间压缩掉。

关键指标	传统做法	引入 AI 后	变化幅度
查找工艺资料	翻阅纸质或电子文档，或询问老师傅	自然语言提问，直接定位	检索时间大幅压缩
新员工上手周期	依赖带教，周期以月计	可自助查询，缩短依赖期	培训成本下降
经验知识流失	老员工离职即流失	沉淀在知识库中	风险降低
故障处置参考	凭经验或翻查历史工单	自动检索相似历史案例	处置速度提升

公开案例与数据

- 该场景在制造业中属于典型的“低风险高频次”应用，公开案例普遍将其价值定位为缩短查找时间与降低对个别专家的依赖，而非直接的成本节省。

落地前提

- **文档治理是主要工作量。** 版本混乱、格式不一、扫描件无法检索是最常见的障碍。这部分投入通常占项目总量的六成以上。
- **必须标注答案来源。** 工业场景中"AI 说的"没有效力，"文件第 X 页说的"才有。答案必须可追溯到原文位置。
- **要有明确的知识更新责任人。** 知识库不更新，三个月后就会开始给出过期答案。

常见坑

- **期望它能回答文档中没有的内容。** 这类系统只能检索已有知识，无法产生新知识。宣传口径若超出这一边界，会迅速失去用户信任。
- **把敏感工艺文件放入外部服务。** 核心工艺是企业命脉，部署方式须与保密要求匹配。

适用与规模判断：工艺复杂、文档积累多、人员流动较快的制造企业。文档基础薄弱的企业需先做文档治理。

场景 3.2.6 生产排程与产线平衡

成熟度 L2 | 价值层次 重构型。 排程是制造业中最典型的"复杂约束优化"问题，也是多数企业至今仍靠 Excel 和老计划员经验完成的环节。

关键指标	传统做法	引入 AI 后	变化幅度
排程编制时间	计划员手工编排，通常以小时至天计	分钟级生成可行方案	编制效率提升
插单与变更响应	需重新手工排，响应慢	快速重排	柔性提升
设备利用率与换型损失	依赖计划员经验平衡	可在多目标间寻优	综合优化
方案可解释性	计划员可解释	需专门设计解释机制	下降，需补偿

公开案例与数据

- 本场景的主力技术是运筹优化而非大模型，大模型主要用于交互层（用自然语言描述排程约束与调整意图）。公开案例中的收益幅度差异极大，取决于原有排程水平。

落地前提

- **基础数据必须准确：工时定额、换型时间、设备能力、物料齐套状态。** 这些数据不准，排出的方案再优也无法执行。这是本场景失败率高的首要原因。
- **要能接受"次优但可执行"的方案。** 数学最优解常常忽略了车间的隐性约束，需要预留人工调整空间。
- **先解决计划员的信任问题。** 计划员是本场景的直接使用者，若其不信任系统输出，会持续手工覆盖，项目实际失效。

常见坑

- 在基础数据不准的情况下上排程系统。这是制造业信息化的经典教训，在 AI 时代同样适用。
- 追求全自动排程。建议定位为"辅助排程"——系统给方案，计划员调整确认，逐步提高自动采纳率。

适用与规模判断：产品种类多、订单批量小、换型频繁的离散制造企业收益最大。单一品种大批量生产的企业排程本身不复杂，优先级低。

场景 3.2.7 安全生产行为识别

成熟度 L3 | 价值层次 替代型。 用视频理解自动识别未戴安全帽、违规进入危险区域、违规操作等行为。这个场景的价值不体现在效率，而体现在**避免的事故成本与监管处罚**。

关键指标	传统做法	引入 AI 后	变化幅度
违章发现方式	安全员巡检抽查	摄像头全时段自动识别	覆盖率从抽检到全时
发现时效	事后或巡检时发现	实时告警	可干预窗口出现
记录与追溯	依赖人工记录	自动留存证据	管理闭环改善
安全员配置	按区域配置巡检人力	转为处置与管理角色	岗位性质改变

公开案例与数据

- 本场景在建筑、化工、电力、矿山等高危行业已广泛部署，多数由监管要求驱动而非企业自发。公开材料显示其技术成熟度较高，主要挑战在于告警的处置闭环。

落地前提

- 明确告警后的处置流程与责任人。只告警不处置，系统就成了一个记录违章的摄像机，反而可能在事故后成为不利证据。
- 摄像头覆盖与角度需按识别需求规划。安防摄像头的角度通常不适合行为识别。
- 做好员工告知与沟通。涉及员工行为的监控须依法告知，且应把定位说明为"保护"而非"监视"，否则会引发抵触与规避行为。

常见坑

- 把系统用于考核处罚而非风险预防。一旦被员工认为是扣钱工具，就会出现各种规避行为，安全目标反而落空。
- 误报率过高。高误报会淹没真实告警，需在上线初期投入时间调优。

适用与规模判断：建筑施工、化工、电力、矿山、金属冶炼等高危行业，以及有明确安全监管要求的场所。

3.3 金融业

金融业是国内大模型应用推进最快的行业之一。据沙丘智库《2026 年中国银行业大模型应用跟踪报告》，已采用大模型的银行占比从 2025 年的 39.0% 上升至 53.5%；正在采用智能体的银行占比从 25.0% 升至 32.3%；明确表示暂无采用计划的银行仅占 2.4%。麦肯锡 2026 年 3 月的负责任 AI 成熟度调研亦显示，金融服务业与科技、传媒、电信同处各行业前列。

但观察金融业的场景分布会发现一条清晰的界线：**成熟度最高的场景，几乎全部集中在“面向内部”或“不直接产生法律后果”的环节；而与授信决策、投资建议直接相关的场景，成熟度普遍停留在 L2。**

这条界线是怎么划出来的

用 SVA-3D 可以完整解释。金融业的数据可及性极高、价值可计量性也很高，唯一的强约束是**容错空间**：一次错误的授信决定或投资建议，不仅带来经济损失，还涉及监管合规与消费者权益保护责任。因此在监管对 AI 辅助决策的责任边界作出明确规定之前，机构的理性选择必然是“让 AI 做输入，让人做决定”。

这带来一个明确的行动建议：不要等监管明确后再动手，而应在当前阶段完成三件事——建立 AI 参与决策的全链路留痕、沉淀人工修正 AI 输出的样本库、设计可分级放权的权限体系。这三件事完成得越早，监管口径明确后的放量速度就越快。

场景 3.3.1 授信调查报告自动生成

成熟度 L2—L3 | 价值层次 替代型。这是金融业中改善幅度最惊人的场景之一。一份百页的授信调查报告，客户经理原本需要一周左右完成资料收集、整理、撰写；引入大模型后可在数分钟内生成初稿，客户经理的角色从“写报告”变为“审报告”。

关键指标	传统做法	引入 AI 后	变化幅度
百页授信报告生成	客户经理平均约 1 周	自动生成平均耗时不到 5 分钟	约 7 天→5 分钟
客户经理角色	大量时间用于资料整理与文字撰写	转为核实、判断与客户沟通	岗位价值重心转移
报告格式一致性	因人而异	模板统一，要素齐全	质量下限提升
覆盖规模	—	服务一线客户经理近 2000 名，覆盖 29 家一级分行	已达规模化部署

公开案例与数据

- **中国光大银行（公开案例）：**上线大模型授信调查智能报告功能。公开披露显示，拟写一份百页授信报告平均需 **1 周左右**，而系统自动生成一份百页调查报告平均耗时**不到 5 分钟**；该系统已服务一线客户经理近 **2000 名**，覆盖 **29 家** 一级分行。

落地前提

- **必须建立可靠的数据源接入。** 报告的价值在于数据准确，而非文字流畅。工商、司法、征信、财报、内部台账等数据源的接入质量决定成败。
- **要有明确的"AI 生成"标识与人工确认环节。** 报告最终由客户经理签字负责，系统须留痕哪些内容为 AI 生成、哪些经人工修改。
- **分行业配置模板。** 制造业、房地产、贸易类企业的调查要点差异极大，通用模板会导致关键风险点遗漏。
- **沉淀人工修改记录。** 客户经理对 AI 初稿的每一次修改，都是训练数据与规则优化的输入。

常见坑

- **用生成速度作为核心 KPI。** 报告写得快没有意义，风险识别准确才有意义。考核应看"经审核后的实质性修改率"与"事后风险暴露情况"。
- **AI 生成的分析结论未经核实即采信。** 生成模型可能产出看似合理但无数据支撑的判断。所有结论性表述必须可追溯到具体数据来源。
- **忽视客户经理的能力退化风险。** 长期只审不写，年轻客户经理的尽调能力可能无法养成。建议保留一定比例的手工撰写训练。

适用与规模判断：有一定授信业务规模的商业银行。业务量越大、报告标准化程度越高、数据源接入越完整，收益越大。

场景 3.3.2 信贷资料智能核验与进件处理

成熟度 L2—L3 | 价值层次 重构型。 这个场景改变的不是某个人的工作效率，而是整条进件流程的形态：从"客户交材料—人工核对—退回补正—再提交"的多轮往返，变为"上传即校验、缺项即提示"的一次性通过。

关键指标	传统做法	引入 AI 后	变化幅度
业务进件速度	小时级	分钟级	数量级压缩
资料收集整理时间	基准	缩短 70%以上	大幅压缩
客户经理人均日处理量	基准	提升约 60%	产能扩张
人工返工率	基准	下降约 80%	往返消除
后台审核人员工作量	基准	减负超 65%	人力释放
资料识别合规准确率	人工核验，易疲劳出错	超 99%	质量提升
资料漏采、错采、模糊上传问题	基准	减少 95%	源头治理
单户运营成本	基准	降低一半以上	可直接入账

公开案例与数据

- **江阴银行**（公开案例，2026 年 8 月）：大模型信贷系统上线后公开披露的数据包括——资料收集整理时间缩短 70%以上，业务进件从小时级压缩至分钟级，客户经理人均日处理业务量提升约 60%，人工返工率下降 80%，后台审核人员工作量减负超 65%，资料识别合规准确率超 99%，资料漏采错采模糊上传等问题减少 95%，用信资料自动核验效率提升 70%，贷后影像资料智能识别率超 98%，单户运营成本降低一半以上。

落地前提

- **要件清单必须先标准化。** 系统能校验的前提是明确定义了"什么业务需要哪些要件、每个要件的合格标准是什么"。这部分制度梳理工作往往比技术实施更耗时。
- **影像质量要求需前置到客户端。** 在客户上传的瞬间给出"照片模糊、缺角、反光"的即时反馈，才能实现"减少 95%的模糊上传"这一效果。事后校验的价值低得多。
- **要与核心业务系统深度集成。** 独立的校验工具无法改变流程形态，必须嵌入进件流程节点。

常见坑

- **只做识别不做流程改造。** 识别准确率再高，如果流程仍是"人工再核一遍"，返工率不会下降。本场景的收益主要来自流程重构而非识别本身。
- **对非标资料期望过高。** 各类非标准格式的证明材料、手写材料的识别效果显著低于标准件，需保留人工通道。
- **忽视客户端体验。** 若前端提示过于严苛导致客户反复上传，会损害获客转化，需在严格性与体验间调优。

适用与规模判断：零售信贷、小微信贷业务量大的银行与消费金融机构。单笔金额小、笔数多、要件标准化程度高的业务收益最大。

场景 3.3.3 编码辅助与测试用例生成

成熟度 L3—L4 | 价值层次 替代型。 金融机构是国内软件开发人力投入最密集的行业之一。这个场景的特点是：容错空间高（代码有测试与评审兜底）、数据可及性高（代码库就在手上）、价值可计量性中等（研发效率的归因有争议）。

关键指标	传统做法	引入 AI 后	变化幅度
基础代码编写	开发人员手工编写	AI 生成后人工审核修改	编写环节耗时下降
单元测试用例编写	常因工期压力被压缩	批量生成，覆盖率可提升	质量环节受益明显
代码注释与文档	常年欠账	可批量补齐	技术债缓解
新人上手周期	依赖代码走读与带教	可通过 AI 理解存量代码	缩短
端到端交付周期	受需求、评审、测试、上线窗口约束	改善幅度显著低于编码环节的改善幅度	注意口径差异

公开案例与数据

- 据沙丘智库《2026 年中国银行业大模型应用跟踪报告》，编码与测试智能体是银行业典型应用案例之一，与智能运维、数字员工建设同属推进最快的方向。
- 本研究院提示：本场景公开披露的"效率提升"数字口径差异极大——多数统计的是"编码环节耗时"或"代码采纳率"，而非端到端交付周期。两者不可混淆（见下方"常见坑"）。

落地前提

- **代码资产与规范必须先梳理。** 生成代码要符合本机构的框架、命名规范与安全要求，需要把这些规范投喂给工具。
- **必须配套代码安全扫描。** AI 生成代码可能引入安全漏洞或不合规的依赖，须纳入既有的安全流水线。
- **评审环节不能弱化。** 生成代码的可读性通常不错，容易让评审者放松警惕，反而增加隐性缺陷风险。

常见坑

- **用"编码效率提升 X%"论证"交付周期缩短 X%"。** 这是本场景最常见的论证错误。在多数金融机构，编码只占端到端交付周期的一小部分，需求澄清、评审、测试、投产窗口才是主要耗时。压缩编码时间未必缩短交付周期。
- **对存量系统改造期望过高。** AI 对老旧系统、私有框架、缺少文档的存量代码的理解能力明显下降。
- **把代码上传至不合规的外部服务。** 金融机构的代码属敏感资产，须采用符合监管要求的部署方式。

适用与规模判断：有自建研发团队的金融机构。新建系统、标准技术栈、有完善测试体系的团队收益最大。

场景 3.3.4 智能运维与故障根因定位

成熟度 L3 | 价值层次 替代型→重构型。 金融系统的告警量极大而真实故障极少，运维人员长期处于"告警疲劳"状态。AI 在这个场景的价值是把海量告警收敛为少量根因判断。

关键指标	传统做法	引入 AI 后	变化幅度
告警处理	海量告警人工逐条研判	自动聚合收敛，输出根因假设	噪声大幅降低
故障定位时间	依赖专家经验，跨系统排查耗时长	自动关联日志、链路与变更记录	定位时间压缩
夜间值守	需安排人力轮班	自动处置常见问题，人工处理例外	值守压力下降
知识沉淀	故障处理经验存于个人	沉淀为可检索的处置知识库	能力资产化

公开案例与数据

- 智能运维是《2026 年中国银行业大模型应用跟踪报告》收录的典型案例方向之一。跨行业参照可见电信业的同类实践（见第四章 4.4 节）——网络与 IT 运维数据规模大、结构化程度高，是 AI 落地条件最好的领域之一。

落地前提

- 日志、监控、配置、变更四类数据必须打通。故障根因分析的本质是跨源关联，缺任一类数据都会显著降低准确率。
- 需要积累历史故障与处置记录作为训练素材。没有历史故障标注，模型无法学会判断根因。
- 明确自动处置的动作白名单。重启服务、切换流量、扩容等动作应分级授权，涉及资金交易链路的动作须人工确认。

常见坑

- 在监控体系不完善的情况下上智能运维。数据缺失时，AI 只会给出更快的错误判断。
- 过早开放自动处置权限。错误的自动处置动作可能把小故障放大为大数据，应从"建议"起步，逐步放权。

适用与规模判断：系统规模大、告警量大、有 7×24 运行要求的金融机构与大型企业 IT 部门。

场景 3.3.5 反欺诈与异常交易识别

成熟度 L4 | 价值层次 替代型。 这是金融业 AI 应用中历史最久、最成熟的场景，早在生成式 AI 之前就已是行业标配。收益直接体现为避免的欺诈损失，价值可计量性极高。

关键指标	传统做法	引入 AI 后	变化幅度
识别方式	规则引擎为主， 新型欺诈手法响应滞后	机器学习+图计算， 可识别团伙关联	覆盖能力提升
响应时效	批量筛查或事后发现	交易实时拦截	从事后到事中
误拦截率	规则过严影响正常客户体验	模型可在拦截率与 误伤率间调优	体验与安全平衡
新型手法适应	需人工分析后新增规则	模型可持续学习	响应周期缩短

公开案例与数据

- 反欺诈是金融 AI 最成熟的应用，各主要银行、支付机构与保险公司均已部署。因涉及风控细节，公开披露的具体数据较少，属于典型的"效果确定但不便公开"的场景。

落地前提

- 需要持续的对抗样本更新。欺诈手法快速演化，模型需持续迭代，这是长期运维成本而非一次性投入。
- 必须设置人工复核与申诉通道。误拦截对正常客户的伤害极大，须有快速申诉与解冻机制。
- 注意模型可解释性要求。监管与客户投诉处理均可能要求说明拦截理由。

常见坑

- 只追求拦截率不看误伤率。过度拦截造成的客户流失成本可能高于欺诈损失。
- 模型上线后长期不迭代。反欺诈是持续对抗，静态模型的效果会随时间衰减。

适用与规模判断：所有涉及资金交易的金融机构与支付平台。交易量越大，模型效果越好。

场景 3.3.6 研报摘要与投研信息聚合

成熟度 L3 | 价值层次 增强型。 投研人员每天面对的信息量远超阅读能力上限。AI 在这个场景的价值不是替代判断，而是扩大可覆盖的信息面。

关键指标	传统做法	引入 AI 后	变化幅度
研报阅读处理	人工逐篇阅读，覆盖量受限	批量摘要提取要点	覆盖面扩大
跨源信息聚合	手工汇总多个数据源	自动聚合并标注来源	汇总时间压缩
数据检索问答	需熟悉数据库结构与查询语法	自然语言提问	使用门槛下降
分析结论	由分析师判断	仍须由分析师判断	不变，且不应改变

公开案例与数据

- 投研辅助是各券商、基金与银行资管部门推进较早的方向。公开材料普遍强调其定位为“信息处理”而非“投资建议”，以规避适当性管理与投资顾问资质相关的合规风险。

落地前提

- 必须标注信息来源与时间戳。投研场景中信息的时效性与来源可靠性是第一位的。
- 明确区分“事实提取”与“观点生成”。前者可用，后者须极度谨慎并明确标注。
- 建立幻觉检测机制。生成模型可能编造数据、错配公司名称与财务数字，投研场景对此零容忍。

常见坑

- 对外提供 AI 生成的投资建议。涉及投资顾问资质与适当性管理的合规风险，须谨慎处理。
- 用 AI 摘要替代原文阅读。关键决策依据必须回溯原文，摘要可能丢失重要限定条件。

适用与规模判断：券商研究所、基金公司、银行资管与私行部门。信息处理量越大，收益越明显。

场景 3.3.7 智能客服与远程银行

成熟度 L3—L4 | 价值层次 替代型。 与电商客服（场景 3.1.4）技术同源，但合规约束显著更强：金融产品的每一句介绍都可能构成销售行为，受适当性管理规则约束。

关键指标	传统做法	引入 AI 后	变化幅度
咨询响应	排队等待坐席	秒级响应	等待消除
服务时段	受坐席排班限制	7×24 小时	全时覆盖
标准业务办理	需坐席逐单操作	自动闭环处理	人工干预率下降
话术合规性	依赖坐席培训与抽检	话术库统一，可前置校验、全量质检	合规风险下降

公开案例与数据

- 智能客服与"数字员工"建设是《2026 年中国银行业大模型应用跟踪报告》收录的主要方向之一，多家银行已在客服中心与远程银行场景规模化部署。

落地前提

- 产品话术必须经合规审核后入库。与电商不同，金融客服的每一句产品介绍都需符合适当性管理与信息披露要求。
- 严格区分"咨询"与"销售"。涉及产品推荐的对话须触发适当性评估流程，不可由 AI 直接完成。
- 全量会话留存与质检。这既是合规要求，也是本场景相对人工坐席的重要优势。

常见坑

- 让 AI 直接推荐理财产品。涉及适当性管理的合规风险，须由具备资质的人员完成或经严格的规则约束。
- 在投诉与纠纷场景中坚持 AI 处理。应快速识别并转人工，避免矛盾升级。

适用与规模判断：有客服中心或远程银行业务的金融机构。咨询量越大、标准化程度越高，收益越明显。

3.4 医疗健康

医疗是六大行业中制度约束最强、同时政策推动也最明确的领域。2024 年 11 月，国家卫生健康委等部门发布《卫生健康行业人工智能应用场景参考指引》，一次性列出 84 个典型应用场景；2025 年 11 月，五部门联合发布《关于促进和规范“人工智能+医疗卫生”应用发展的实施意见》，提出到 2030 年基层诊疗智能辅助应用基本实现全覆盖。

基础设施侧的进展为 AI 下沉提供了条件：据公开报道，全国已有约 80% 的县（市、区）初步建成县域影像、心电、检验资源共享中心，2025 年县域远程医学影像诊断服务量超过 6800 万人次。

医疗 AI 的三个结构性特征

- 一、成熟度与容错空间高度负相关。成熟度最高的是影像检出（AI 提示、医师确认），最低的是基层全科诊疗辅助（AI 直接影响诊疗路径）。这个排序几乎完全由容错空间决定，与技术难度关系不大。
- 二、价值兑现的主体与投入主体不一致。医院投入 AI，收益的很大一部分体现为患者等待时间缩短、误诊率下降等社会效益，而非医院自身的财务收益。这一错位是医疗 AI 商业模式长期不清晰的根本原因。
- 三、基层是最大增量，也是最难的部分。三甲医院的 AI 价值是“提效”，基层的 AI 价值是“从无到有”——属于第四层创造型价值。政策把 2030 年目标定在基层全覆盖，正是着眼于此。

场景 3.4.1 门诊病历与预问诊生成

成熟度 L2—L3 | 价值层次 替代型。医生把大量时间花在敲键盘上，而不是看病人。这个场景通过语音识别与结构化生成，把病历书写时间压缩了一个数量级，直接释放的是最稀缺的医生时间。

关键指标	传统做法	引入 AI 后	变化幅度
单份病历书写	5—10 分钟/份	15—20 秒/份 (另有案例披露平均 44 秒生成完整病历)	约 20 倍效率提升
日均病历工作量	约 8 小时	约 25 分钟	数量级压缩
单次看诊时长	约 6.97 分钟	约 4 分钟	下降约 40%
病历内容准确率	人工书写， 受时间压力影响完整性	准确率超 90% (需医师确认)	质量下限提升
科室覆盖与产出	—	覆盖 6 个专科并推进至 9 个专科， 累计生成预问诊病历 6.6 万余份	已达规模化

公开案例与数据

- **上海市第一人民医院（公开案例）：**在眼科白内障手术病历生成场景中，病历书写从 **5—10 分钟/份** 缩短至 **15—20 秒/份**，每日相关工作从 **8 小时** 缩短至 **25 分钟**，效率提升近 20 倍；日均处理近 100 台手术的电子病历。意图识别与诊断抽取准确率均在 90% 以上。
- **北京世纪坛医院（公开案例）：**AI 辅助病历生成项目分立项调研、试点运行、推广扩展、运行维护四阶段推进，历时约半年完成从试点到多科室推广。公开披露成效为：**平均 44 秒生成一**

份完整病历，内容准确率超 90%，看诊时间从 6.97 分钟缩短至 4 分钟左右（下降约 40%），全院累计生成预问诊病历 6.6 万余份，覆盖 6 个专科并向 9 个专科推进。

落地前提

- **必须分专科定制模板与术语。** 通用模板在专科场景下的可用性很低。上述两个案例都采取了"按专科逐个定制"的路径，而非一次性全院上线。
- **诊室的声学环境需要评估。** 多人诊室、嘈杂环境会显著降低语音识别准确率，部分场景需要定向麦克风改造。
- **医师确认环节不可省略。** 病历是法律文书，AI 生成内容必须经医师审核确认后方可归档，且系统须留痕。
- **采用"试点一定制一推广"的分阶段路径。** 两个公开案例都用了约 3—6 个月完成单科试点，再逐科推广。跳过试点直接全院上线的失败率很高。

常见坑

- **把 90%的准确率理解为"可以不审核"。** 剩余 10%中可能包含关键的阴性症状遗漏或用药错误，医师确认是不可让渡的环节。
- **忽视病历质控与医保审核要求。** 生成的病历需同时满足病案质控标准与医保结算规则，这些规则须前置嵌入模板。
- **只算医生时间不算实施成本。** 分专科定制的工作量、与 HIS/EMR 系统的集成成本、持续的模板维护，都需计入。
- **用"生成速度"作为主要考核指标。** 应考核"医师实质性修改率"——修改率过高说明模板不适配，节省的时间会被修改时间抵消。

适用与规模判断：门诊量大、病历书写负担重的二级以上医院。专科特征明显、诊疗流程标准化程度高的科室（眼科、皮肤科、口腔科等）优先。

场景 3.4.2 医学影像辅助检出与分诊

成熟度 L3—L4 | 价值层次 增强型。 这是医疗 AI 中最成熟的方向。它的定位始终清晰：**AI 负责"不漏"，医师负责"定性"。** 价值体现在两处——减少阅片工作量，以及把急症从排队队列中提前捞出来。

关键指标	传统做法	引入 AI 后	变化幅度
肺结节筛查工作量	医师逐层阅片	AI 预标注后医师复核	减少医生 30%—50%工作量
影像诊断整体效率	基准	—	整体提升约 30%
主动脉夹层等急症影像判读	15—20 分钟	约 3 分钟	判读时间压缩约 85%
微小病灶检出	受疲劳与经验影响	可识别毫米级微小结节	早期检出率提升
基层可及性	基层缺乏影像诊断能力	约 80%的县（市、区）已建成县域影像等共享中心；	从无到有

关键指标	传统做法	引入 AI 后	变化幅度
		2025 年县域远程影像诊断超 6800 万人次	

公开案例与数据

- 据公开报道（新华网，2026 年 3 月）：在肺结节筛查场景，AI 可减少医生 **30%—50%** 的工作量，影像诊断效率整体提升约 **30%**；主动脉夹层的影像诊断时间可从 **15—20 分钟压缩至约 3 分钟**。全国约 **80%** 的县（市、区）已初步建成县域影像、心电、检验资源共享中心，2025 年县域远程医学影像诊断服务量超 **6800 万人次**。

落地前提

- 须使用取得相应医疗器械注册证的产品。用于临床诊断的影像 AI 属医疗器械管理范畴，这是不可绕过的前提。
- 要与 PACS/RIS 系统深度集成。需要医师切换系统才能查看的 AI 结果，实际使用率会很低。
- 明确 AI 结果在报告中的呈现方式与责任界定。报告最终由医师签发，AI 结果的引用方式须符合规范。
- 急症分诊场景需要打通排队逻辑。“3 分钟判读”的价值只有在能真正把该患者提前到队列前面时才兑现。

常见坑

- 自动化偏见。医师可能因信任 AI 提示而降低独立判断强度，尤其对 AI 未标注的区域放松警惕。这是国际研究中反复观察到的现象，须通过流程设计与培训干预。
- 只看 AI 单独的敏感度/特异度指标。真实场景中决定结果的是“AI+医师”组合的表现。建议增加两类观测：人机组合的最终诊断质量相比人工基线的变化；医师对 AI 提示的采纳率与推翻率分布——后者是识别自动化偏见的早期信号。
- 跨设备、跨机构的泛化能力不足。在 A 医院设备上训练的模型，在 B 医院不同厂商设备上的表现可能显著下降，须做本地验证。

适用与规模判断：有影像科室的各级医院，以及承担区域影像诊断中心职能的机构。基层机构的边际价值最大。

场景 3.4.3 智能导诊与预问诊

成熟度 L3 | 价值层次 重构型。把“到院后排队询问挂哪个科”这一环节，前移到就诊前的手机端完成，同时把患者主诉结构化后传递给医生。改变的是就诊流程的形态。

关键指标	传统做法	引入 AI 后	变化幅度
分诊方式	导诊台人工引导 或患者自行判断	症状描述后自动推荐科室	前移至院外
挂错科室率	患者自判易出错， 造成二次挂号	推荐准确率可显著改善	就诊路径优化

关键指标	传统做法	引入 AI 后	变化幅度
医生获取信息	诊室内从零问起	就诊前已获得结构化主诉与病史	诊室时间释放
导诊人力	需配置导诊台人员	转向处理复杂与特殊需求患者	人力重配置

公开案例与数据

- 智能导诊与预问诊在三甲医院中已较为普及，是《卫生健康行业人工智能应用场景参考指引》列出的典型场景之一。北京世纪坛医院的公开案例显示，预问诊与病历生成的联动是其项目的核心设计（见场景 3.4.1）。

落地前提

- 必须与医院的科室设置与专家排班对齐。推荐了科室但没有号源，反而增加患者挫败感。
- 须设置明确的急症识别与红旗提示。对可能的急危重症，系统应立即提示直接就医而非继续问询流程。
- 保留人工导诊通道。老年患者与不熟悉智能设备的人群必须有替代路径。

常见坑

- 让 AI 给出诊断性结论。导诊系统的边界是"建议去哪个科"，不是"你可能得了什么病"。越界会带来医疗风险与合规风险。
- 问询流程过长。患者的耐心有限，问题过多会导致中途放弃，反而不如直接挂号。

适用与规模判断：门诊量大、科室设置复杂的三级医院。科室少、患者熟悉的基层机构收益有限。

场景 3.4.4 病案首页编码与质控

成熟度 L3 | 价值层次 替代型。 病案编码直接关系 DRG/DIP 结算，编码错误既影响医院收入也带来合规风险。这是医疗 AI 中**价值可计量性最高**的场景——收益可以直接换算成结算金额。

关键指标	传统做法	引入 AI 后	变化幅度
编码方式	编码员逐份人工编码	AI 预编码+编码员审核	工作量下降
编码一致性	因人而异	规则统一	波动减少
质控覆盖率	人工抽检	全量自动质控	从抽检到全检
结算影响	编码错误导致结算损失或违规风险	错误率下降	可直接换算为金额

公开案例与数据

- 病案编码与质控是医院信息化中 AI 应用较早且效果确定的方向，在 DRG/DIP 支付方式改革推进后，其价值进一步凸显。

落地前提

- **依赖病历书写质量。** 病历记录不完整、诊断书写不规范时，AI 编码同样无法准确。因此本场景常与场景 3.4.1 联动实施。
- **编码规则需随版本更新。** ICD 编码与本地 DRG/DIP 分组规则会更新，系统须有及时的规则维护机制。
- **保留编码员终审。** 编码结果直接影响结算，须由具备资质的编码员确认。

常见坑

- **用 AI 编码追求结算金额最大化。** 这已越过合规红线，属于高风险行为。系统设计应以"准确"而非"有利"为目标。
- **忽视病历质量这一上游问题。** 只优化编码环节而不改善病历书写，效果有限。

适用与规模判断：实施 DRG/DIP 支付的各级医院。病案量越大，收益越明显。

场景 3.4.5 慢病管理与智能随访

成熟度 L2—L3 | 价值层次 重构型。 随访是公认重要但长期做不到位的工作——人力成本高、随访率低。AI 把随访的边际成本降到接近于零，使"全员随访"从理想变为可行。

关键指标	传统做法	引入 AI 后	变化幅度
随访方式	护士电话逐个随访	自动外呼或消息随访	边际成本趋近于零
随访覆盖率	受人力限制，覆盖率通常有限	可实现全量覆盖	覆盖面数量级提升
随访频次	按规定节点，难以加密	可按需高频随访	颗粒度提升
异常识别	依赖随访人员判断	规则+模型自动识别异常并转人工	人力聚焦于异常

公开案例与数据

- 慢病随访是《卫生健康行业人工智能应用场景参考指引》列出的典型场景。北京世纪坛医院的公开案例中，术后随访是其试点验证的场景之一。

落地前提

- **必须有明确的异常转人工规则。** 随访发现异常后若无人接手，系统就是在制造风险而非降低风险。
- **患者知情同意与个人信息保护须合规。** 涉及健康数据的处理受严格法律约束。
- **内容须由临床团队审定。** 随访问题与健康指导内容属于医疗建议范畴，不可由技术团队自行设定。

常见坑

- **用随访完成率作为唯一指标。** 打通电话不等于随访有效，应关注异常发现率与后续干预率。
- **AI 给出用药调整建议。** 这已进入诊疗行为范畴，须由医师完成。

适用与规模判断：有慢病管理职能的医院、社区卫生服务中心与家庭医生团队。管理人群规模越大，收益越明显。

场景 3.4.6 药物研发辅助

成熟度 L2 | 价值层次 创造型。这是医疗领域少数属于第四层价值的场景——AI 不是加快了原有流程，而是让某些原本不可行的探索路径变得可行（如超大规模化合物空间的筛选）。

关键指标	传统做法	引入 AI 后	变化幅度
靶点发现与化合物筛选	实验筛选， 通量受限于实验资源	计算筛选可覆盖 远超实验能力的化合物空间	探索范围扩大
临床试验文档	大量人工撰写	结构化生成后人工审核	撰写环节提效
研发周期	新药研发周期以年计	早期环节可压缩， 但临床阶段受法规约束不变	总周期改善有限

公开案例与数据

- 药物研发是 AI 在生命科学领域应用最受关注的方向。公开材料普遍指出，AI 在早期发现阶段的价值已较为确定，但由于临床试验阶段的周期由法规与受试者招募决定，**总研发周期的压缩幅度显著小于早期环节的改善幅度。**

落地前提

- **需要高质量的实验数据积累。** 计算结果必须由实验验证形成闭环，缺少湿实验能力的纯计算团队难以形成价值。
- **要建立计算与实验的协同机制。** 本场景的成败取决于两支团队能否高效配合。

常见坑

- **用早期环节的改善幅度论证总周期缩短。** 与场景 3.3.3（编码辅助）的常见坑同理——局部提效不等于端到端提效。
- **低估验证成本。** 计算筛选产出的候选化合物仍需大量实验验证，这部分成本可能抵消筛选环节的节省。

适用与规模判断：有自研管线的制药企业与生物科技公司、科研院所。无实验验证能力的机构不适合独立开展。

3.5 教育

2026 年 4 月，教育部、国家发展改革委、工业和信息化部、科技部、国家数据局五部门联合印发《“人工智能+教育”行动计划》（教科信〔2026〕1 号），从人才培养与素养提升、教育深度融合、基础环境建设、生态系统优化四方面部署，提出建设国家教育智算服务平台、基础语料库与教育专用大模型，目标是到 2030 年形成人工智能与教育深度融合的格局。

但从落地角度看，教育是本白皮书六个深度行业中**价值可计量性最低**的一个。这不是数据问题，而是学科属性问题：学习效果的显现存在长周期滞后，且受家庭、动机、同伴等大量混杂因素影响，很难把某段成绩变化归因于某个 AI 工具。

这一特性带来的两个后果，以及本研究院的建议

- 后果一：教育 AI 项目的立项与验收长期缺乏公认标准，采购决策容易被演示效果主导。
- 后果二：最容易被量化的场景（教师备课时长、批改工作量）反而不是教育的核心价值所在，容易造成“AI 减轻了教师负担，但学生学习效果没有变化”的局面。
- 建议：优先解决**“负担”而非“效果”**。不是因为效果不重要，而是因为效果类场景的收益无法在采购周期内被证明，容易导致项目被中途否定。更稳妥的路径是先解决教师的事务性负担，收益在一个学期内即可观察；教师时间释放之后，再由教师本人决定如何投入到因材施教上——这是一条“通过解放人来改善效果”的间接路径，比让 AI 直接干预学习过程更符合当前的技术成熟度。

场景 3.5.1 备课与教案、试卷生成

成熟度 L3 | 价值层次 替代型。教师工作时间中有相当比例花在备课材料准备、课件制作、习题选配上。这个场景是教育 AI 中收益最确定、争议最小的一个——因为它不触碰教学判断本身。

关键指标	传统做法	引入 AI 后	变化幅度
教案初稿准备	查找资料、参考同类教案、逐节撰写	按教学目标生成初稿后修改	准备时间大幅压缩
课件制作	手工排版与素材寻找	结构化生成后调整	制作时间压缩
组卷	从题库人工挑选、平衡难度与知识点	按知识点与难度分布自动组卷	组卷时间压缩
分层教学材料	因工作量大而难以实现	可为不同层次学生批量生成差异化材料	从不可行到可行

公开案例与数据

- 教育部《中国智慧教育发展报告（2025—2026）》与《“人工智能+教育”行动计划》均把教师教学工具作为重点方向。相关公开调研显示，教师群体对 AI 工具的使用已相当普遍，备课与资料准备是使用频率最高的场景。

落地前提

- 必须对齐课程标准与本地教材版本。**通用模型生成的内容常与本地教材的编排顺序、术语体系不一致，需要接入教材知识图谱或本地语料。

- **教师审核是必要环节。** 生成内容可能包含知识性错误或不适龄表述，必须经教师确认。
- **要沉淀本校优质教案作为参考素材。** 直接使用通用模型的产出，质量通常低于本校骨干教师的水平；结合本校教案库才能产出可用材料。

常见坑

- **用生成材料直接进课堂而不审核。** 学科知识性错误在课堂上的负面影响远大于节省的时间。
- **以"生成数量"考核使用效果。** 应关注教师的实际采纳率与修改幅度。
- **忽视对青年教师成长的影响。** 备课过程本身是教师专业成长的重要途径，完全依赖生成可能影响教学基本功的养成。建议对新入职教师保留一定的手工备课要求。

适用与规模判断：各学段学校。班级数多、教师工作量大、需要分层教学的学校收益最明显。

场景 3.5.2 主观题批改与作文评价

成熟度 L2—L3 | 价值层次 替代型。 客观题的自动批改早已成熟，主观题（尤其是作文）的批改一直是教师负担的最大来源。AI 在这里的定位是"初评+反馈"，而非"定分"。

关键指标	传统做法	引入 AI 后	变化幅度
作文批改	教师逐篇精批， 一个班需数小时	AI 初评生成反馈， 教师复核调整	批改时长大幅压缩
反馈及时性	受批改周期限制， 通常隔数日	可即时反馈	反馈闭环加快
反馈颗粒度	受时间限制， 评语往往简略	可给出分维度的具体建议	反馈质量提升
练习频次	因批改负担而受限	可提高练习频次	训练量提升
评分权威性	教师评分具权威性	**AI 评分不宜直接作为 正式成绩**	边界必须明确

公开案例与数据

- AI 作文批改在中小学与语言培训领域应用较广。公开材料普遍强调其价值在于"提高反馈频次"而非"替代教师评分"——因为评分涉及教育公平与家长信任，边界较为敏感。

落地前提

- **必须明确 AI 评分的效力边界。** 建议定位为"练习反馈"，正式考试评分仍由教师完成，并向学生与家长明确说明。
- **评价维度需与教学要求对齐。** 不同学段、不同文体的评价标准差异大，需分别配置。
- **保留教师调整通道并记录调整幅度。** 调整幅度是评估模型适配度的关键指标。

常见坑

- **用 AI 评分作为正式成绩或升学依据。** 涉及教育公平，风险极高。

- **评语模板化导致学生失去信任。** 千篇一律的 AI 评语会被学生迅速识别并忽略，反而降低反馈效果。
- **只评分不给具体改进建议。** 分数对学习的帮助有限，具体的、可操作的修改建议才是本场景的价值所在。

适用与规模判断：写作训练量大的语文、英语学科，以及语言类培训机构。

场景 3.5.3 口语陪练与语言学习

成熟度 L3—L4 | 价值层次 替代型+创造型。 这是教育 AI 中**唯一已经完成成本结构改写**的场景。一对一口语外教的成本使其无法普及；AI 陪练把边际成本降到接近于零，让“每天练 20 分钟口语”从少数人的选项变成所有人的选项。

关键指标	传统做法	引入 AI 后	变化幅度
一对一口语练习	外教一对一，按小时计费，成本高	边际成本接近于零	成本结构改写
可练习时间	受约课与时差限制	随时可练	时间约束消除
开口量	课堂中人均开口时间极少	一对一持续开口	练习量数量级提升
心理门槛	当众开口存在畏难情绪	无社交压力	参与度提升
反馈精度	外教反馈质量高但主观	发音评测可给出量化反馈	可量化

公开案例与数据

- 口语陪练是语言学习领域应用最成熟的 AI 场景，市场化产品已相当普及。其成功的核心原因符合 SVA-3D 的判断：数据可及性高（语音与文本数据充足）、容错空间高（说错了改就是了，且犯错本身是学习过程）、价值可计量性中等（发音准确度与流利度可量化）。

落地前提

- **评测标准要与学习目标匹配。** 应试口语与实用交流口语的评价重点不同，须分别配置。
- **要设计练习内容的进阶路径。** 单纯的自由对话缺乏结构，学习效率低于有设计的练习序列。
- **保留真人交流环节。** AI 陪练解决的是练习量问题，真实交流中的语用能力仍需真人场景。

常见坑

- **过度追求发音评测分数。** 评测算法有其偏好，刷分可能偏离真实的交流能力。
- **用 AI 陪练完全替代真人课程。** 两者解决的是不同问题，替代关系不成立。

适用与规模判断：语言学习者、语言培训机构、开设外语课程的学校。学习者基数越大，边际成本优势越明显。

场景 3.5.4 学情分析与教学建议

成熟度 L2 | 价值层次 增强型。把分散在作业、测验、课堂互动中的数据汇总，识别出班级共性薄弱点与个体差异。定位是"给教师提供依据"，而非"替教师做决定"。

关键指标	传统做法	引入 AI 后	变化幅度
学情掌握方式	教师凭经验与批改印象判断	基于全量作答数据的知识点掌握分析	依据从经验转向数据
薄弱点定位	定位到"这个班数学差"	定位到具体知识点与题型	颗粒度提升
个体差异识别	受班额限制难以覆盖全员	可覆盖每个学生	覆盖面提升
分析报告生成	教师手工统计	自动生成	统计工作量消除

公开案例与数据

- 学情分析是《"人工智能+教育"行动计划》中"教育智能大脑"的重要组成。公开材料显示，其效果高度依赖作业与测验的数字化程度——纸笔作业占主导的学校难以获得所需数据。

落地前提

- **作业与测验必须数字化且知识点标注准确。**这是本场景的硬前提，也是多数学校的实际障碍。
- **分析结果必须转化为可操作的教学动作。**"这个知识点掌握率 62%"本身没有价值，"建议下节课用 X 方式重讲 Y 内容"才有价值。
- **注意学生数据的保护要求。**未成年人个人信息受特别保护，数据采集范围与使用方式须严格合规。

常见坑

- **用学情数据给学生贴标签或排名公示。**涉及未成年人权益，且可能造成心理伤害。
- **只做数据看板不做教学干预。**看板本身不改善学习效果。
- **为获取数据而增加学生作业负担。**与减负要求相悖。

适用与规模判断：作业与测验已数字化的学校。纸笔作业为主的学校需先解决数据采集问题。

场景 3.5.5 企业培训与技能课程生成

成熟度 L2—L3 | 价值层次 替代型。企业内训的核心矛盾是"课程开发周期长、更新慢，跟不上业务变化"。AI 把课程开发从"项目"变成了"日常动作"。

关键指标	传统做法	引入 AI 后	变化幅度
课程开发周期	从需求到上线通常以月计	基于既有材料快速生成，以天计	周期数量级压缩
课程更新频率	因开发成本高而更新滞后	可随业务变化及时更新	时效性提升

关键指标	传统做法	引入 AI 后	变化幅度
岗位覆盖	只为大岗位开发课程，长尾岗位无课程	长尾岗位也可低成本覆盖	覆盖面扩大
模拟演练	需组织角色扮演，成本高频次低	AI 扮演客户/面试官，可反复演练	从稀缺到充足

公开案例与数据

- 企业培训场景与教育场景技术同源，但价值可计量性更高——培训效果可通过岗位绩效指标间接验证。公开材料显示，销售话术演练、客服场景演练、合规知识考核是应用最集中的方向。

落地前提

- 必须以企业自有的业务知识与规范为素材。通用课程内容对企业价值有限，核心是把内部经验与制度转化为课程。
- 要与实际业务指标建立关联观测。培训场景的价值论证依赖后续绩效改善，需提前设计观测方案。

常见坑

- 生成大量无人学习的课程。课程数量不是价值，完课率与行为改变才是。
- 用 AI 生成合规培训内容而不经法务审核。合规内容的准确性要求极高。

适用与规模判断：员工规模较大、岗位类型多、业务变化快的企业。零售、餐饮、金融、客服等岗位标准化程度高的行业收益最明显。

3.6 政务与公共服务

政务是国务院《关于深入实施"人工智能+"行动的意见》六大重点行动中"人工智能+治理能力"的直接对应领域。研究侧，北京大学政府管理学院与中国信息通信研究院于 2026 年世界人工智能大会期间联合发布《政务智能体发展研究报告（2026 年）》，汇编 30 项典型经验，覆盖高频服务类、运行协同类、治理研判类、裁量决策辅助类四大场景，并提出"高频服务先行、运行协同稳步、治理研判探索"的梯次推进路径。

这个推进顺序与 SVA-3D 的判断高度吻合：高频服务类场景数据可及性高、容错空间大（答错可纠正）、价值可计量（办件量、一次办结率），因此最先落地；裁量决策辅助类场景容错空间极小（涉及行政相对人权益），因此最后推进。

政务 AI 的三个独特约束

- 一、合法性优先于效率。行政行为的合法性要求意味着，AI 输出如果直接构成行政决定的依据，必须能够被完整解释和追溯。这是裁量类场景推进缓慢的根本原因，而非技术不足。
- 二、数据汇聚的行政成本高于技术成本。跨部门数据共享涉及职责边界与责任划分，往往需要专项协调机制推动。这决定了政务 AI 项目的实际周期通常远长于技术实施周期。
- 三、服务对象的能力差异大。政务服务面向全体公民，包括数字能力较弱的群体。这要求必须保留人工兜底通道，不能以"AI 已能处理"为由削减线下服务——这一约束在企业场景中并不存在。

场景 3.6.1 政务热线智能受理与派单（12345）

成熟度 L3 | 价值层次 替代型→重构型。政务热线是全国范围内 AI 落地最深、案例最多的政务场景。它的价值链条很清晰：话务承接—诉求理解—责任主体匹配—派单—督办—质检，AI 可以贯穿全链条。

关键指标	传统做法	引入 AI 后	变化幅度
智能派单准确率	人工派单受话务员经验影响	90%以上	准确率显著提升
整体处理效率	基准	—	提升约 25%
诉求响应时间	基准	—	缩短约 30%
话务承接	高峰期排队，受坐席数量限制	智能应答分流常见咨询	峰值压力缓解
质检覆盖率	人工抽检	全量自动质检	从抽检到全检
数据规模与复用	海量工单数据难以有效利用	基于 4 亿余条诉求数据支撑"未诉先办"研判	数据资产化

公开案例与数据

- **辽宁省 12345 热线**（公开案例，2026 年 6 月）：构建涵盖"智能应答、智能受理、智能派单、智能办理、智能质检"的全流程闭环体系。公开披露数据为：**智能派单准确率 90%以上，整体处理效率提升约 25%，诉求响应时间缩短约 30%，**"小宁智能问答"线上咨询达 5.7 万次，数

据规模 4 亿余条诉求。系统已向沈阳、大连等 5 个市级平台及 6 个县區部署，"辽宁模式"并复制至海南、青海、长春、杭州等地。

- 清华大学数字政府与治理研究院《AI 驱动政务热线发展研究报告（2025）》对该领域的技术路径与治理要求作了系统梳理，可作为方案设计的参考框架。

落地前提

- **责任主体清单必须先梳理清楚。** 派单准确率的上限由"事项—责任部门"对照关系的清晰程度决定，而非模型能力。跨部门职责交叉的事项是准确率的主要失分点。
- **需要积累足量的历史工单作为训练素材。** 上述案例的 4 亿条诉求数据是其效果的基础，数据积累不足的地区效果会明显下降。
- **要打通督办与考核环节。** 派单准确不等于办理到位，价值兑现依赖后续的督办闭环。
- **方言与口音的识别需专门优化。** 政务热线的来电人群覆盖面广，通用语音模型在方言场景下准确率会明显下降。

常见坑

- **用智能应答率作为主要考核指标。** 这会激励系统把复杂诉求也拦在 AI 环节，导致群众反复来电，实际服务质量下降。
- **忽视"兜底率"。** 企业场景常用"自动闭环率"衡量成效，直接移植到政务场景是危险的——在公共服务中，一次未被妥善兜底的失败交互，社会成本远高于企业客服场景。**本研究院建议增设两个指标：兜底成功率**（智能体无法处理时成功转接人工并最终办结的比例）**与弱势群体服务达成率**（老年人、残障人士等群体的办结率与整体办结率的差值），以防止"为提高自动率而抬高服务门槛"的技术性排斥。
- **AI 直接对群众作出办理承诺。** 承诺的履行涉及多个部门，AI 不应作出超出其能力范围的答复。

适用与规模判断：各级政务热线平台。话务量越大、历史工单积累越多、区域内部门协同机制越完善，收益越大。

场景 3.6.2 公文起草与校核辅助

成熟度 L2—L3 | 价值层次 替代型。 机关单位的文字工作量极大，且格式规范严格。AI 在这个场景的价值集中在"格式合规"与"初稿成型"两个环节，而非内容判断。

关键指标	传统做法	引入 AI 后	变化幅度
公文初稿	从零撰写，反复修改格式	按文种生成规范初稿	成稿时间压缩
格式规范校核	人工逐项对照规范检查	自动校核格式、用语、体例	校核工作量下降
引用准确性	手工核对政策文件原文	自动检索并标注来源	核对效率提升
内容实质判断	由起草人与审核人负责	仍由人负责，且不应改变	边界明确

公开案例与数据

- 公文辅助是《政务智能体发展研究报告（2026 年）》所归纳的"运行协同类"场景的典型代表，在各级机关中应用较广。

落地前提

- **必须内置本单位的公文格式规范与常用体例。** 通用模型对党政机关公文格式的掌握不足，需专门配置。
- **政策引用必须可追溯到原文。** 生成模型可能编造或错引政策条款，这在公文中是严重问题。所有引用须标注文号与条款位置。
- **部署方式须符合保密要求。** 涉密与内部敏感信息不得进入外部服务。

常见坑

- **用 AI 起草涉及重要决策与政策表述的文件。** 此类文件的实质内容必须由责任人把关，AI 仅可辅助格式与文字。
- **生成内容中出现编造的政策依据。** 这是本场景风险最高的失败方式，必须建立强制的引用核验环节。
- **忽视写作能力的传承。** 公文写作是机关干部的基本功，完全依赖 AI 可能影响年轻干部的能力养成。

适用与规模判断：各级党政机关与事业单位。文字工作量大、公文种类多的单位收益明显。

场景 3.6.3 办事材料智能预审与导办

成熟度 L2—L3 | 价值层次 重构型。群众办事最常见的痛点是"跑了两趟才知道材料不全"。这个场景把材料校验前移到提交之前，改变的是办事流程的形态。

关键指标	传统做法	引入 AI 后	变化幅度
材料校验时点	窗口受理时人工核对	线上提交时即时校验	前移至提交前
补正次数	常需一次或多次补正	缺项即时提示，减少往返	往返次数下降
办事指南理解	条文式指南，群众理解成本高	自然语言问答式导办	理解门槛下降
窗口人员工作	大量时间用于材料核对	转向复杂事项与特殊情形处理	人力重配置

公开案例与数据

- 材料预审与智能导办属于《政务智能体发展研究报告（2026 年）》定义的"高频服务类"场景，是梯次推进路径中最先落地的一类。

落地前提

- **事项标准化是硬前提。** 事项名称、申请条件、材料清单、审查标准必须先统一规范，否则 AI 无法给出确定的校验结果。这部分工作属于政务服务标准化范畴，工作量大且不可跳过。
- **要与政务服务平台深度集成。** 独立的预审工具无法改变办事流程。

- **必须保留线下与人工通道。** 数字能力较弱的群体必须有替代路径，且不应因线上办理率考核而被弱化。

常见坑

- **预审标准严于实际受理标准。** 会导致群众按 AI 提示准备了材料却仍被退回，或反之，损害政府公信力。预审规则必须与受理规则严格一致。
- **用线上办理率考核挤压线下服务。** 这是政务数字化中反复出现的问题，在 AI 场景中同样需要警惕。

适用与规模判断：各级政务服务中心与线上政务平台。高频事项、材料要求复杂的事项收益最大。

场景 3.6.4 基层网格事件智能识别与派单

成熟度 L2—L3 | 价值层次 重构型。 城市治理中的大量问题（占道经营、垃圾堆放、设施损坏、违章停车）依赖网格员巡查发现。视觉 AI 把“人巡”变为“机巡+人处置”。

关键指标	传统做法	引入 AI 后	变化幅度
问题发现方式	网格员定时巡查	视频自动识别+群众上报	发现渠道扩展
发现时效	受巡查频次限制	实时或准实时	时效大幅提升
事件分类与派单	人工判断归属部门	自动分类并匹配责任主体	流转加快
网格员角色	大量时间用于巡查发现	转向现场处置与群众工作	岗位价值转移

公开案例与数据

- 基层网格智能治理属于《政务智能体发展研究报告（2026 年）》定义的“治理研判类”场景。多地城市运行管理平台已部署，但各地成熟度差异较大。

落地前提

- **摄像头资源与点位需统筹。** 多数城市已有大量摄像头，但归属不同部门、点位与角度未按治理需求规划，整合的行政成本高于技术成本。
- **事件分类标准与责任清单必须先明确。** 与热线派单同理，准确率上限由责任划分的清晰度决定。
- **要控制识别范围，避免过度采集。** 公共空间的视频分析须符合个人信息保护要求，应避免采集与治理无关的个人信息。

常见坑

- **只发现不处置，形成“事件积压”。** 发现能力提升会使事件数量激增，若处置力量未同步，会导致大量事件超期未办，反而形成考核压力与舆情风险。**这是本场景最典型的失败方式。**
- **用事件发现量作为绩效指标。** 会激励重复上报与过度识别。应考核办结率与群众满意度。

适用与规模判断：已建成城市运行管理平台、且有一定摄像头基础的城市城区。处置力量不足的地区应先解决处置能力再上发现能力。

场景 3.6.5 政策问答与政策文件检索

成熟度 L2—L3 | 价值层次 增强型。政策文件数量庞大、层级复杂、更新频繁，无论是企业找政策还是干部用政策，检索成本都很高。这个场景的定位是"精准找到条文"，而非"解释政策"。

关键指标	传统做法	引入 AI 后	变化幅度
政策检索	按标题或关键词检索，难以按需求场景检索	按场景与条件语义检索	检索方式改变
惠企政策匹配	企业需自行判断是否符合条件	按企业画像主动匹配可申报政策	从人找政策到政策找人
政策解读	需咨询主管部门	AI 给出条文依据与适用条件	自助获取
答案权威性	主管部门解释具权威性	**AI 答复不具法律效力，须标注来源并提示以原文为准**	边界必须明确

公开案例与数据

- 政策问答与惠企政策精准推送在多地已有实践，是《政务智能体发展研究报告（2026 年）》所收录经验的常见类型之一。

落地前提

- **政策库的版本管理是核心工作。**失效、修订、废止的政策若未及时下架，会给出错误答复。需建立明确的政策入库与失效流程。
- **答案必须标注文件名称、文号与条款位置。**政务场景中"AI 说的"没有效力，必须可追溯到原文。
- **须明示 AI 答复的性质。**应明确提示"本答复仅供参考，以政策原文和主管部门解释为准"。

常见坑

- **用 AI 答复替代正式的政策解释。**政策解释权属于制定机关，AI 答复不具效力，若被理解为官方答复会引发争议。
- **政策库更新滞后。**这是本场景最主要的风险来源，且失效政策造成的误导后果严重。

适用与规模判断：各级政务服务平台、营商环境服务平台、机关内部办公系统。政策数量多、层级复杂的地区收益更大。

3.7 自媒体获客与创始人 IP

这一节与前面六节不同。前面六节按行业划分，这一节按**职能**划分——因为“通过内容获客”是一件几乎所有行业的企业都在做、且做法高度趋同的事：无论你卖的是工业设备、法律服务还是护肤品，在小红书发笔记、在抖音做口播、把评论区的意向客户接进私域，这套动作的骨架是一样的。

把它单独成节，还有一个更实际的理由：**这是本白皮书中最适合中小企业和个人品牌起步的一组场景**。它不需要打通 ERP，不需要改造产线，不需要跨部门数据协同——一个人、一台电脑就能开始且效果在一个月内可见。

3.7.0 先看清楚这条链路

创始人 IP 获客不是“发内容”这一件事，而是一条有六个环节的链路。AI 在每个环节的成熟度和价值完全不同，混为一谈是这个方向最常见的错误。

表 3-9 内容获客链路的六个环节与 AI 的介入程度

环节	要解决的问题	AI 当前能做到什么	成熟度
① 选题	发什么内容有人看	对标账号拆解、爆款要素归纳、选题库批量生成	L3
② 生产	内容怎么做出来	脚本、图文、封面、口播视频、剪辑，成本降一到两个数量级	L3—L4
③ 分发	如何覆盖多平台多账号	一稿多平台改写、矩阵账号批量分发	L3
④ 承接	有人问了怎么接住	评论区与私信秒级应答、意向识别、自动留资	L2—L3
⑤ 培育	线索怎么变成客户	私域 SOP 自动化、个性化跟进内容生成	L2
⑥ 被搜到	AI 搜索时代如何被推荐	GEO 生成式引擎优化，让品牌被 AI 正确理解与引用	L1—L2（快速上升）

资料来源：求智启新人工智能研究院。

这条链路上，最容易被忽略的是第④环

绝大多数企业把全部注意力放在②生产环节——因为那里最容易看到“AI 好厉害”。但从获客结果看，**决定成败的往往是第④环（承接）**。

一个公开数据可以说明问题：据 2026 年小红书运营相关公开分析，**超 70% 的成交转化集中在私信咨询的前 3 分钟**，且平台算法已将私信回复率与响应时效纳入流量分配权重——回复滞后会直接影响内容曝光。

换句话说：**你用 AI 把内容产能提高了 10 倍，但如果多出来的咨询在 3 分钟内没人回，这 10 倍产能大部分是浪费的**。本节的卡片顺序按链路排列，建议按顺序读，不要跳到生产环节就停。

场景 3.7.1 选题挖掘与对标账号拆解

成熟度 L3 | 价值层次 增强型。 内容做不起来，八成不是执行问题而是选题问题。这个场景把"凭感觉想选题"变成"从数据里找选题"，是整条链路中投入产出比最高、却最常被跳过的一环。

关键指标	传统做法	引入 AI 后	变化幅度
选题来源	凭个人经验与灵感， 产出不稳定	批量分析对标账号的 爆款内容，归纳共性要素	从灵感驱动到数据 驱动
对标分析	人工逐条翻阅竞品内容， 一天看几十条	批量抓取后自动归类、 提炼标题结构与钩子类型	分析样本量数量级 提升
选题库储备	常处于"明天发什么"的 临时状态	一次生成数十至上百条 可执行选题	从被动到有储备
用户需求洞察	靠猜	从评论区、问答区、 搜索词中提取真实问题	需求来源真实化

公开案例与数据

- 据 2026 年小红书运营相关公开分析，平台算法的意图识别能力持续增强，能快速捕捉评论区的高意向留言（询价、求推荐等）——这意味着**评论区本身就是最好的选题来源**：用户在评论里问什么，就是下一条内容该讲什么。
- 本研究院提示：该场景公开的量化数据极少，多为方法论层面的经验总结。其价值主要体现在"减少无效内容的比例"，而这一指标各家统计口径差异极大。

落地前提

- **必须先定义清楚"你要的客户长什么样"。** 选题的目标不是流量，是**目标客户的流量**。定位不清时，AI 会推荐泛流量选题，看起来数据好但不转化。这是本场景的第一前提。
- **建立自己的爆款素材库。** 把行业内跑通的内容存下来，作为 AI 分析的输入。通用模型对垂直行业的内容规律理解有限。
- **把评论区与私信中的真实提问作为选题主来源。** 这是免费且最准的需求数据。
- **区分"引流内容"与"转化内容"。** 前者追覆盖面，后者追精准度，两者的选题逻辑完全不同，不能用同一套标准评价。

常见坑

- **追逐平台热点而偏离专业定位。** 蹭热点能带来一次流量，但会稀释账号标签，导致后续内容的推荐精准度下降。对做 B 端获客的创始人 IP，这个损失通常大于短期流量收益。
- **用点赞量而不是留资量评价选题。** 一条 10 万赞的泛娱乐内容，可能带来 0 条有效线索；一条 500 赞的专业内容，可能带来 20 个客户。**考核指标错配是这个方向最普遍的失败方式。**
- **直接照搬对标账号的选题。** 同样的选题在不同人设、不同粉丝结构下效果完全不同。AI 给的是参考不是答案。

适用与规模判断：所有做内容获客的企业与个人。B 端、高客单、决策周期长的业务尤其需要——因为这类业务的内容必须精准，泛流量几乎无价值。

场景 3.7.2 图文笔记与短视频脚本生成

成熟度 L3—L4 | 价值层次 替代型。内容生产是整条链路中 AI 改写最彻底的环节。原本"一个人一天写 2—3 条"的产能天花板被直接掀掉，内容供给从瓶颈变成非瓶颈。

关键指标	传统做法	引入 AI 后	变化幅度
图文笔记撰写	人工构思+撰写，一天数条	按选题批量生成初稿后人工润色	产能数量级提升
短视频脚本	编导撰写，一条需数小时	批量生成分镜脚本	撰写时间大幅压缩
封面与配图	设计师制作或素材库购买	生成式图像批量产出	参见场景 3.1.1，成本降幅约 90%
100 条电商短视频	约 1500 元	约 300 元	降幅约 80%
单个分镜生成	100—300 元	约 1.55 元	降幅 98%以上
A/B 测试能力	产能不足，无法同时测多个版本	同一选题可产出多版标题、封面、开头	测试规模提升

公开案例与数据

- 据公开报道的行业成本测算：**100 条电商短视频的制作成本从约 1500 元降至约 300 元（降幅约 80%）**；**单个分镜生成从 100—300 元降至约 1.55 元（降幅 98%以上）**；3 分钟漫剧单集从约 180 元降至约 36 元。视频制作领域整体测算显示"单条视频成本降低了 60%"。
- 横向参照场景 3.1.1：电商内容生产领域公开披露的综合成本降幅可达约 90%，1—2 人可完成原 30—40 人团队的工作量。
- **注意口径：**上述成本对比多针对"批量、标准化、非精品"的内容。精品内容与品牌级内容的 AI 替代率显著更低。

落地前提

- **必须先沉淀"你自己的话"。**直接用通用模型生成的内容有明显的模型腔，在小红书、抖音这类对真实感敏感的平台会上会被用户识别并划走。正确做法是先把创始人的口头表达、常用案例、观点主张整理成语料，让 AI 在此基础上生成。**这是本场景成败的关键前提。**
- **建立平台差异化的改写规则。**小红书重"我的真实经历"，抖音重"前 3 秒钩子"，视频号重"关系链共鸣"，公众号重"结构化深度"。同一内容需按平台改写而非直接复用。
- **保留人工润色环节。**AI 产出的是初稿不是终稿。建议的分工是：AI 负责结构与铺陈，人负责观点、案例与情绪。
- **建立敏感词与合规词库。**各平台对功效宣称、绝对化用语、医疗与金融相关表述有严格限制，须前置校验。

常见坑

- **用 AI 直接量产同质化内容。**平台算法对低质重复内容有明确的识别与限流机制，量产反而会拖累账号权重。**产能提升必须配合质量把关，否则是负收益。**

- **内容没有"人味"。** 创始人 IP 的核心资产是信任，而信任来自个人化的表达、真实的经历与明确的观点。全 AI 生成的内容恰恰在这三点上最弱。
- **忽视平台合规。** 教育、医疗、金融、法律等行业的内容受额外监管，违规轻则限流重则封号，前期积累归零。
- **只算生产成本，不算审校与合规成本。** 与场景 3.1.1 同理，若审校仍需逐条人工把关，实际节省会大幅低于宣称值。

适用与规模判断：所有做内容获客的企业与个人。内容需求量越大、更新频率越高、账号数量越多，收益越明显。

场景 3.7.3 创始人口播与数字分身

成熟度 L3 | 价值层次 替代型+创造型。 创始人 IP 最大的瓶颈是"创始人只有一个，且时间最贵"。数字分身把"出镜"这件事从创始人的时间约束中解放出来——但它同时是本节中**风险最需要控制**的场景。

关键指标	传统做法	引入 AI 后	变化幅度
口播视频制作	创始人本人出镜拍摄，需布景、化妆、多次重录	文本驱动生成，无需本人到场	时间约束解除
单条制作周期	拍摄+剪辑，以小时至天计	分钟级产出	周期数量级压缩
多语种版本	需重新拍摄或配音	同一形象生成多语种版本	出海成本大幅下降
内容更新	改一句话需重拍整条	改文本即可重新生成	修改成本趋近于零
专业短片（10 分钟）	约 350 万元（影视级）	约 3000 元	极端案例降幅 99%以上
真实感与信任度	本人出镜信任度最高	当前仍有可感知差距	这是主要制约

公开案例与数据

- 据公开报道的影视与短视频行业成本测算，AI 视频在极端案例中的降幅可达 99%以上（如 5 秒专业特效镜头从约 3000 元降至约 3 元）；80 集真人短剧制作成本从 30—50 万元降至 2—20 万元。这些数字说明的是**技术能力的上限**，而非日常口播场景的典型水平。
- 参照场景 3.1.2（数字人直播）：公开材料普遍将数字人的核心价值定位为"极低边际成本的时段覆盖"，而非"单场效果超越真人"。**口播分身的定位应与此一致——用于扩展产能，而非替代创始人的核心内容。**

落地前提

- **必须取得本人的书面授权并妥善保管形象与声音素材。** 这既是法律要求，也是防止素材被滥用的必要措施。
- **建议明确标注为 AI 生成。** 部分平台已有强制标识要求，且主动标注对信任的损害通常小于被用户发现后的反弹。
- **核心观点内容仍应本人出镜。** 建议的分工是：**建立信任的内容本人做，交付信息的内容分身做。** 例如行业观点、价值主张由本人出镜，产品说明、课程讲解、常见问答由分身完成。

- **语料库比形象更重要。** 分身的说服力来自内容质量与表达风格，而非形象逼真度。

常见坑

- **用数字分身讲需要建立信任的内容。** 这是本场景最大的错误。创始人 IP 的价值基础是"这个人可信"，而分身恰恰无法传递这一点。用它讲价值观、讲行业判断、讲承诺，会直接伤害账号的核心资产。
- **未经授权使用他人形象或声音。** 涉及明确的人格权侵权风险，已有大量公开案例。
- **分身与本人的表达风格不一致。** 老客户能立刻察觉，会引发"这个人变了"的负面感受。
- **全部内容改用分身。** 一旦账号的真实感消失，粉丝互动率与转化率会同步下滑，且难以恢复。

适用与规模判断：已有一定内容基础、创始人时间成为瓶颈的企业。内容需求量大、需要多语种或多平台覆盖的场景收益最明显。**刚起步、尚未建立个人信任度的账号不建议使用。**

场景 3.7.4 账号矩阵与多平台分发

成熟度 L3 | 价值层次 替代型。 一稿多发的核心难点不是"发"，而是"改"——每个平台的内容偏好、字数限制、话术风格都不同。AI 把改写成本压到接近于零，矩阵运营才真正变得可行。

关键指标	传统做法	引入 AI 后	变化幅度
一稿多平台改写	人工按平台逐个改写	一次生成多平台版本	改写成本趋近于零
可运营账号数	受人力限制，通常个位数	矩阵规模可显著扩大	覆盖面提升
矩阵渠道流量占比	—	公开分析显示 KOS 导购账号矩阵流量占比突破 60%	已成主流渠道
本地化与多语种	每语种需独立团队	批量生成多语种版本	出海门槛下降

公开案例与数据

- 据 2026 年小红书运营相关公开分析：**账号矩阵运营已经普及，KOS（金牌导购）账号矩阵引流模式成熟，此类渠道流量占比突破 60%**；平台鼓励品牌绑定专业号与 KOS 导购账号形成合规矩阵。

落地前提

- **必须使用平台认可的合规矩阵形式。** 各平台对多账号运营有明确规则（如企业号+员工号的绑定关系），违规的"养号矩阵"会被批量封禁。
- **每个账号需要差异化定位。** 内容完全相同的矩阵账号会被算法识别为重复内容并限流，矩阵的价值来自"不同角度覆盖不同人群"。
- **统一的知识库与话术底座。** 矩阵账号越多，口径不一致的风险越高，须有统一的产品信息与话术来源。

常见坑

- **用同一份内容批量分发到多个账号。** 这是最常见的做法，也是最容易触发限流的做法。
- **矩阵规模超出承接能力。** 账号多了咨询就多，若第④环（承接）跟不上，流量全部浪费——参见本节开头的提示。
- **员工账号的资产归属不清。** 员工离职带走账号是矩阵运营的典型纠纷，需提前约定。

适用与规模判断：有多条产品线、多个区域市场或多名一线业务人员的企业。单一产品、单一市场的小微企业优先做好主号即可。

场景 3.7.5 评论区与私信的即时应答

成熟度 L2—L3 | 价值层次 重构型。 本节中最被低估、但对获客结果影响最大的场景。它解决的是一个纯粹的物理问题：客户在深夜 11 点问了一句“多少钱”，而你在睡觉。

关键指标	传统做法	引入 AI 后	变化幅度
首次响应时间	受人工在线时间限制，可能数小时	秒级响应	响应窗口从小时到秒
服务时段	工作时间	7×24 小时	全时段覆盖
高意向识别	人工逐条翻看，易遗漏	自动识别询价、求推荐等高意向留言	线索不漏
留资率	基准	公开案例披露某工具使品牌留资率提升 38%	转化前置
成交时间窗口	—	超 70% 成交转化集中在私信咨询前 3 分钟	响应速度直接决定成交
对流量的反向影响	回复慢不影响曝光	**平台已将响应时效纳入流量分配权重**	回复慢=曝光少

公开案例与数据

- 据 2026 年小红书运营相关公开分析：平台算法优先扶持私信回复率高、能做到秒级响应的账号，回复滞后会直接影响内容曝光；超 70% 的成交转化集中在私信咨询的前 3 分钟；用户消费决策链路缩短 45%；某 AI 私信工具公开披露帮助品牌留资率提升 38%。
- **这组数据的含义值得强调：** 在当前的平台算法下，响应速度不只影响转化，还反向影响你的内容能被多少人看到。这使得第④环从“售后环节”变成了“流量环节”。

落地前提

- **必须先建立结构化的产品与常见问题知识库。** 报价区间、服务范围、交付周期、常见异议的标准回答——这些不整理清楚，AI 只能给出模糊回复，反而降低信任。
- **明确划定 AI 可自主回复的边界。** 建议：产品介绍、常见问答、预约引导可自动；报价谈判、方案定制、投诉处理必须转人工。
- **设置高意向自动转人工的触发规则。** 出现询价、要方案、问案例等信号时，应立即通知真人接手，而不是继续由 AI 周旋。

- **遵守各平台对自动化工具的规则。** 部分平台对第三方自动回复工具有限制，须核对当期规则，避免账号受限。

常见坑

- **用 AI 应付高意向客户。** 客户能感觉出来。已经明确表达购买意向的人被机器人接待，是最伤转化的做法。**AI 的定位应是"接住并识别"，而不是"谈成"。**
- **回复内容过于模板化。** 在小红书这类强调真实互动的平台，模板回复会被用户和算法同时识别。
- **只做私信不做评论区。** 评论区是公开的，一条好的评论区回复能被后来的所有访客看到，其杠杆远高于一对一私信。
- **承接能力建设滞后于内容产能。** 这是本节反复强调的核心风险。

适用与规模判断：所有有咨询量的内容获客账号。**咨询量越大、客单价越高、竞争越激烈的行业，响应速度的价值越大。**

场景 3.7.6 直播获客与切片分发

成熟度 L2—L3 | 价值层次 增强型+创造型。 对创始人 IP 而言，直播的价值不在带货，而在**建立信任的密度**——一场 2 小时直播传递的信息量与真实感，是几十条短视频难以达到的。AI 在这里做两件事：降低开播门槛，以及把一场直播的价值榨干。

关键指标	传统做法	引入 AI 后	变化幅度
直播内容准备	人工写大纲与话术	AI 生成结构化直播大纲与问答预案	准备时间压缩
实时问答支持	主播凭记忆应对	实时检索知识库提示答案要点	专业度稳定性提升
直播切片	人工回看剪辑，一场直播出几条	自动识别高光片段，批量生成切片	单场产出数量级提升
切片分发	人工逐平台上传	批量改写标题后分发矩阵	长尾流量持续获取
长尾时段覆盖	受主播体力限制	数字人承接长尾时段（参见场景 3.1.2）	时段扩展

公开案例与数据

- 参照场景 3.1.2（数字人直播）：公开材料显示数字人直播已成为电商零售的标配能力，核心价值是“极低边际成本的 7×24 小时开播”；国内 AI 数字人市场 2024 年增速达 85.3%，布局企业超 114 万家。
- 本研究院提示：**创始人 IP 直播与带货直播的逻辑不同。** 前者的目标是信任与线索，后者的目标是成交转化，因此数字人在前者中的适用性显著低于后者——建议数字人只用于回放、切片与长尾时段，核心直播仍由本人完成。

落地前提

- **直播前必须准备结构化的问答知识库。** 现场被问住是直播中最伤专业形象的时刻，实时检索能显著降低这一风险。
- **把切片当作独立的内容产品来设计。** 直播时有意识地做“可切片的表达”——完整的观点、清晰的开头结尾，会大幅提高切片可用率。
- **建立切片到线索的转化路径。** 切片本身不产生线索，须配置引导动作（评论区置顶、主页引导、私信关键词）。

常见坑

- **用数字人做创始人 IP 的核心直播。** 与场景 3.7.3 同理，会直接损害信任资产。
- **切片只追数量不看完播。** 无头无尾的机械切片完播率极低，不如少而精。
- **直播不做承接设计。** 直播结束后咨询涌入而无人应答，是常见的资源浪费。

适用与规模判断：已有一定内容基础、且业务适合通过深度沟通建立信任的企业（B 端服务、高客单产品、专业服务、教育培训）。

场景 3.7.7 GEO：生成式引擎优化

成熟度 L1—L2（快速上升） | 价值层次 创造型。 这是本白皮书中最新、最不确定、但可能影响最远的场景。当用户越来越多地问 AI“哪家公司做这个比较好”而不是去搜索引擎翻结果时，“能不能被 AI 正确理解并引用”就成了一个全新的获客变量。

关键指标	传统做法	引入 AI 后	变化幅度
优化目标	SEO：搜索排名与点击量	GEO：被 AI 直接引用与零点击曝光	目标物改变
用户行为迁移	传统搜索为主	超 41% 用户几乎完全转向 AI 搜索； 47.8% 大幅增加 AI 搜索频率； 仅 10.6% 仍以传统搜索为首选	入口正在迁移
购买决策影响	搜索引擎影响决策	超 80% 用户购买前通过 AI 搜索辅助决策	决策入口前移
市场规模	2025 年国内 GEO 市场约 6 亿元	2030 年预测突破 500 亿元	高速增长期
见效周期与可计量性	SEO 排名可直接观测	引用率观测手段尚不成熟	价值可计量性低

公开案例与数据

- 据艾瑞咨询《2026 年 GEO 生成式引擎优化行业研究报告》：国内 GEO 市场 2025 年约 **6 亿元**，预计 2030 年突破 **500 亿元**；用户侧，**超 41% 的用户几乎完全转向 AI 搜索，47.8% 大幅增加 AI 搜索使用频率，仅 10.6% 仍以传统搜索引擎为首选；超 80% 的用户在购买前通过 AI 搜索辅助决策**，3C 数码、家电、旅游等决策周期长的品类渗透率更高。报告指出 GEO 与 SEO 的核心差异在于“SEO 追求点击量与排名，而 GEO 聚焦 AI 直接引用与零点击曝光”。

- 该报告同时指出，GEO 的市场增长来自两个方向：存量端是传统 SEO 与 PR 预算的转移，增量端是知识库构建、网页优化、API 对接等技术服务支出。

落地前提

- **先把"关于你的事实"整理清楚并公开发布。** AI 引用的前提是能找到结构化、可信、一致的信息。企业介绍、产品参数、服务范围、案例、资质在各渠道口径不一致时，AI 要么不引用，要么引用错误版本。**这是 GEO 最基础也最有效的动作。**
- **在权威、可被抓取的渠道沉淀内容。** 包括官网的结构化页面、行业媒体、专业问答社区、公开的行业报告与白皮书——AI 更倾向引用来源可靠的内容。
- **围绕真实问题组织内容，而非围绕关键词。** 用户对 AI 提的是完整问题（"预算 50 万的中小制造企业做 AI 质检该怎么起步"），而非关键词。内容结构应对应问题结构。
- **建立引用监测机制。** 定期在主流 AI 助手检索本品牌相关问题，记录是否被提及、被如何描述、有无错误信息，形成基线并持续观测。这是当前唯一可行的效果衡量方式。

常见坑

- **把 GEO 当成 SEO 的换皮版本。** 关键词堆砌、外链农场这类做法在生成式检索中基本无效，且可能被判定为低质内容。
- **期望短期可量化的 ROI。** 按 SVA-3D 判断，GEO 属于"数据可及性高、容错空间高、价值可计量性低"的组合——这意味着它值得做，但不适合作为需要短期交差的 KPI 项目。建议以"品牌资产建设"而非"获客渠道"的口径立项。
- **放任错误信息不管。** AI 可能引用过时或错误的品牌信息（旧地址、旧价格、已停产的产品），这类错误会被反复复制，且纠正成本随时间上升。**发现即纠正，是 GEO 的日常工作。**
- **只做官网不做第三方渠道。** AI 的引用来源分散，单一渠道的覆盖度有限。

适用与规模判断：决策周期长、客户会主动做功课的行业（B 端服务、专业服务、高客单消费品、教育、医疗健康、软件与工业设备）。冲动型、低客单、依赖货架流量的品类优先级较低。

场景 3.7.8 线索培育与私域跟进

成熟度 L2 | 价值层次 重构型。 内容获客的最后一环。B 端与高客单业务的成交周期常以月计，而多数线索死在"加了微信之后没有下文"。AI 在这里解决的是"持续跟进"的人力成本问题。

关键指标	传统做法	引入 AI 后	变化幅度
线索分层	凭销售个人判断	按互动行为自动打分分层	优先级可排序
跟进内容	销售各自发挥，口径不一	按客户所处阶段生成个性化跟进内容	一致性与相关性提升
跟进频次	受销售精力限制，长尾线索被放弃	低意向线索可低成本持续培育	线索生命周期延长
沉默线索激活	基本不做	按行业动态、新案例触发再触达	存量线索复活

关键指标	传统做法	引入 AI 后	变化幅度
销售动作留痕	依赖手工记录	自动记录与提醒	过程可管理

公开案例与数据

- 本场景的公开数据主要来自 SaaS 服务商的营销材料，属 C 级信息源中可靠性较低的一类，本白皮书未采用其具体数字。其价值逻辑是清晰的：**当跟进的边际成本趋近于零时，原本因“不值得花时间”而被放弃的长尾线索重新变得有价值。**

落地前提

- **必须先定义清楚线索分层标准与各层的跟进动作。** 没有明确的 SOP，AI 只会生成没有目的的问候语，反而消耗客户耐心。
- **内容要有真实价值。** 培育内容应是行业洞察、案例、方法，而不是“在吗”“考虑得怎么样了”。后者发得越勤，被删得越快。
- **必须遵守个人信息保护与营销规范。** 未经同意的商业信息推送存在合规风险，须有明确的授权与退出机制。

常见坑

- **高频群发骚扰。** AI 降低了发送成本，最容易引发的后果就是过度打扰，导致批量拉黑与投诉——这是把资产变成负债。
- **用 AI 冒充真人做深度沟通。** 被识破后信任损失远大于效率收益。建议 AI 负责触达与内容准备，真人负责对话。
- **只做触达不做转化路径设计。** 培育内容需明确的下一步动作（预约诊断、领取资料、参加直播），否则再多触达也不产生结果。

适用与规模判断：成交周期长、客单价高、需要多次触达的业务（B 端服务、专业服务、教育培训、大宗消费品）。冲动消费型业务收益有限。

给准备启动创始人 IP 获客的企业：三个月的最小可行路径

第 1 个月——先解决“说什么”和“接得住”。 用场景 3.7.1 把选题库建起来（重点从客户真实提问中提取），同时把场景 3.7.5 的知识库与自动应答配好。**不要先冲产能。** 这个月的目标是：每条内容都有明确的目标客户，且任何咨询都能在 3 分钟内被接住。

第 2 个月——放大产能。 启用场景 3.7.2 的批量生产与 3.7.4 的多平台分发，把内容频次提上去，同时严格记录“哪类选题带来了有效线索”。这个月的核心指标是**留资量**，不是播放量。

第 3 个月——补链路两端。 前端用场景 3.7.6（直播与切片）提升信任密度，后端用 3.7.8 把线索培育跑起来；同时启动场景 3.7.7（GEO）的基础工作——把关于自己的事实信息整理清楚并公开发布。

全程不建议做的事：一上来就做数字分身（场景 3.7.3）。在个人信任尚未建立时使用分身，会让账号从起点就失去最重要的资产。

本节小结：内容获客链路上，AI 的能力分布是极不均匀的——生产环节（②）已到 L3—L4，承接环节（④）在 L2—L3，培育（⑤）与 GEO（⑥）仍在 L1—L2。但**获客结果的瓶颈通常在④而非②**。把预算按能力成熟度分配是常见做法，把预算按瓶颈分配才是正确做法。

第四章 十二个行业场景速览

本章覆盖另外十二个行业，共 36 张精简卡片。与第三章相比，本章保留了"对照表+案例+落地前提+常见坑"的核心信息，但压缩了展开论述。每节末尾给出该行业的关键变量——即决定这个行业 AI 能走多远的那个瓶颈。

4.1 物流与供应链

场景 4.1.1 智能调度与订单分配

成熟度 L3 | 价值层次 增强型。 把"人找车、车找货"的撮合从人工调度转为算法实时匹配，核心指标是准点率与配对率。

关键指标	传统做法	引入 AI 后	变化幅度
订单分配	人工调度或简单规则	算法实时匹配	分配时效提升
日处理规模	受调度人力限制	日均百万级订单与 几十万名司机的即时分配	规模数量级提升
准点率	基准	95%	服务质量可量化
配对率	基准	92%	运力利用提升

公开案例与数据

- **货拉拉（公开案例）：**平台日均处理百万级订单与几十万名司机之间的即时分配，公开披露**准点率 95%、配对率 92%**；其"货运无忧大模型"的货运事实性问答准确率达 90% 以上。风控侧，2024 年封禁假单用户近 10 万（同比增长 47%）、警告提醒司机近 6 万人（同比增长一倍）。

落地前提

- **需要实时的位置、运力、订单三类数据流。**任一数据延迟都会显著降低匹配质量。
- **目标函数须多方平衡。**只优化平台效率会损害司机收入公平性，进而影响运力供给稳定。

常见坑

- **只优化单一指标。**极致压缩配送时间会推高安全风险与司机流失，需设置约束边界。
- **忽视算法对劳动者的影响。**派单算法的公平性与透明度已成为监管与舆论关注重点。

适用与规模判断：有一定规模的运力平台、网络型快递企业、大型企业自有车队。

场景 4.1.2 单据识别与自动对账

成熟度 L3 | 价值层次 替代型。物流业单据密集（运单、回单、磅单、发票、异常件凭证），识别与对账是纯粹的重重复劳动，也是最容易算清账的场景。

关键指标	传统做法	引入 AI 后	变化幅度
单据录入	人工逐单录入	OCR+结构化提取	录入环节消除
对账周期	月末集中对账，人力集中投入	日常自动核对，例外人工处理	周期分散化
网点人力成本	基准	公开案例披露乡镇网点人力成本降约 60%	可直接入账
货物体积测量	人工测量或估算	AR 识货自动测量	效率提高十倍以上

公开案例与数据

- **快递 100**（公开案例）：AI 对账系统使乡镇网点人力成本骤降约 60%。
- **AR 识货**（公开案例）：货物体积测量效率提高十倍以上。

落地前提

- **单据版式需先归类。** 版式种类过多且无规律时，识别准确率会明显下降。
- **必须设置对账差异的人工处理通道与阈值。** 小额差异自动处理、大额差异必须人工复核。

常见坑

- **忽视手写与破损单据。** 这部分占比虽小但处理成本高，需保留人工路径。

适用与规模判断：网点数量多、单据量大的快递、快运、货代企业。

场景 4.1.3 仓内机器人与智能分拣

成熟度 L3—L4 | 价值层次 替代型。 大型仓储中心的标配。改变的是"人到货"变为"货到人"，拣选效率与准确率同时提升。

关键指标	传统做法	引入 AI 后	变化幅度
拣选方式	拣货员按单行走拣货	机器人搬运货架至工位	行走时间消除
拣选准确率	受人工疲劳影响	系统引导，准确率提升	差错下降
仓储密度	需预留人行通道	可提高存储密度	空间利用提升
峰值弹性	大促需大量临时工	可通过增加机器人扩容	弹性改善

公开案例与数据

- **京东物流**（公开案例）：部署智狼飞梯机器人与搬运机器人，覆盖入库、上架、拣选、出库等环节。
- **菜鸟**（公开案例）：研发用于自动驾驶的强化学习规划器。

落地前提

- **仓库需为自动化重新设计或改造。** 在既有仓库中加装的效果通常显著低于新建自动化仓。
- **订单结构要适配。** 多品少量的电商订单适配度高；整箱整托的 B 端订单收益有限。

常见坑

- **投资回收期估算过于乐观。** 仓储自动化属重资产投入，需按 5 年以上周期测算，并考虑业务量波动风险。

适用与规模判断：日均订单量大、SKU 多、拣选作业密集的电商仓与区域分拨中心。

关键变量：跨主体数据共享机制。 物流 AI 的下一层价值（第三层重构型）需要货主、承运方、仓储方之间的数据打通，其推进速度取决于产业链协同机制与数据权属规则的明确，而非算法进步。

4.2 能源与电力

政策指引最具体的行业之一。2025 年 9 月，国家发展改革委与国家能源局印发《关于推进“人工智能+”能源高质量发展的实施意见》（国能发科技〔2025〕73 号），覆盖电网、能源新业态、新能源、水电、火电、核电、煤炭、油气八大方向，提出到 2027 年推动 5 个以上专业大模型应用、挖掘 10 个重点示范项目、探索百个应用场景。

场景 4.2.1 智能节能与能耗优化

成熟度 L3 | 价值层次 增强型。 能源与通信行业中价值最容易量化的场景——省下的电可以直接换算成钱，无任何归因争议。

关键指标	传统做法	引入 AI 后	变化幅度
设备运行控制	按固定策略运行， 闲置设备照常耗电	实时监控后智能关断 与启动闲置设备	精细化控制
制冷系统	固定风速与频率	按负载与环境智能调节	能耗下降
年节电量	基准	约 8 亿度（中国电信公开披露）	规模化收益
年节约电费	基准	约 5.2 亿元（中国电信公开披露）	可直接入账

公开案例与数据

- **中国电信（公开案例）：**自研 AI 节能系统部署于基站、IT 云主机、AB 类机房，通过实时监控设备运行与环境信息、对局部空闲设备智能关断与启动、对机房与数据中心制冷系统风速频率智能调节，公开披露**年均节电约 8 亿度、年均节约电费约 5.2 亿元。**
- **南钢（公开案例）：**峰谷发电 AI 调度，公开披露相关效益提升约 4.24 倍。

落地前提

- **需要分项计量与实时采集能力。** 只有总表数据无法定位优化空间。

- **安全与业务约束必须硬性建模。** 通信基站关断需保证覆盖不受影响，工业设备关断需保证工艺不受影响。

常见坑

- **为省电牺牲服务质量或设备寿命。** 频繁启停的隐性成本可能超过电费节省。
- **只做监测看板不做调度动作。** 看板本身不省电。

适用与规模判断：高耗能企业、大型数据中心、通信网络运营者、大型园区。

场景 4.2.2 负荷与新能源功率预测

成熟度 L3 | 价值层次 增强型。 预测精度直接影响调度成本与考核罚款，是电力系统中价值可计量性最高的场景之一。

关键指标	传统做法	引入 AI 后	变化幅度
预测方式	经验曲线与统计外推	融合气象、历史、实时数据的机器学习	精度提升
新能源出力预测	风光出力波动大，预测误差高	预测误差下降	考核罚款减少
调度成本	需保有较高备用容量	备用需求下降	成本下降
市场交易	报量报价依赖经验	基于预测优化报价策略	收益提升

公开案例与数据

- 功率预测是《关于推进“人工智能+”能源高质量发展的实施意见》明确列出的重点方向之一（新能源功率预测、电力供需预测）。该场景在新能源场站与电网调度机构中已广泛部署。

落地前提

- **气象数据源的质量决定上限。** 数值天气预报的精度是新能源功率预测的主要制约。
- **需要长周期历史数据覆盖不同气象条件。** 至少需要跨越完整年度周期。

常见坑

- **只看整体预测误差不看极端时段。** 极端天气下的预测失准造成的损失，可能超过日常精度提升的全部收益。

适用与规模判断：新能源场站、售电公司、电网调度机构、参与电力市场交易的主体。

场景 4.2.3 输变电设备智能巡检

成熟度 L3 | 价值层次 替代型。 用无人机与固定摄像头替代人工登杆、爬塔、徒步巡线，既提效也显著降低作业安全风险。

关键指标	传统做法	引入 AI 后	变化幅度
巡检方式	人工徒步或登高巡检	无人机自主巡检+AI 识别	作业方式改变
巡检覆盖频次	受人力与地形限制	可显著提高频次	覆盖提升
高危作业	需登高、跨越复杂地形	大幅减少人员高空作业	安全风险下降
缺陷识别	依赖巡检人员经验	AI 自动识别典型缺陷	一致性提升

公开案例与数据

- 无人机智能巡检在国家电网系统中已规模化应用，是电力行业 AI 落地最成熟的方向之一。中国科学院自动化研究所等机构在电网无人机 AI 自适应巡检方面亦有公开的技术成果。

落地前提

- 需建立缺陷样本库。典型缺陷（绝缘子破损、销钉缺失、异物悬挂）的样本积累是识别准确率的基础。
- 空域与飞行安全管理须合规。无人机作业受空域管理规定约束。

常见坑

- 只采集不研判，形成海量未处理影像。巡检影像量极大，若识别与工单流转不通，数据会成为负担。

适用与规模判断：输配电网运维单位、大型光伏与风电场站、长输管道运营方。

关键变量：容错空间。电力系统的安全边界由物理规律与监管要求共同决定。能源 AI 的成熟度上限，短期内取决于"AI 建议—人工确认—系统执行"这一链路的可靠性设计，而非模型精度。

4.3 汽车与出行

汽车是 AI 渗透最深的制造子行业，其特殊性在于 AI 同时出现在产品端（智能驾驶与座舱）和生产端（研发与制造），且产品端的渗透速度更快。据公开报道，工业和信息化部数据显示组合驾驶辅助功能渗透率已达约 70%，高阶领航辅助（NOA）功能渗透率突破 30%。

场景 4.3.1 整车研发仿真加速

成熟度 L2—L3 | 价值层次 增强型。用神经网络代理模型替代部分传统数值仿真计算，把"提交计算—等待数小时—一看结果"的迭代循环压缩到近实时，从而改变设计工作的方式。

关键指标	传统做法	引入 AI 后	变化幅度
空气动力学单次计算	约 4 小时	约 1 分钟	约 240 倍加速

关键指标	传统做法	引入 AI 后	变化幅度
设计迭代方式	提交—等待—评估， 每日只能试少数方案	近实时反馈， 可大量尝试	工作方式改变
概念设计	手绘到 3D 模型需 多团队耗费数月	数小时完成	周期数量级压缩
整车开发周期目标	约 60 个月（传统）	目标 30 个月（日产公开目标）	目标减半

公开案例与数据

- **捷豹路虎**（公开案例）：借助神经网络加速，空气动力学计算从原本约 **4 小时缩短至约 1 分钟**。
- **通用汽车**（公开案例）：利用工具将手绘草图转换为 3D 模型与动画，此前需多个团队耗费数月的工作现在数小时完成；并开发了"AI 驱动虚拟风洞"，可提供近实时的风阻系数预测。
- **日产**（公开案例）：将 AI 应用于代码生成与单元测试等重复性工作；公司将整车开发周期目标从传统的约 60 个月压缩至 30 个月。

落地前提

- **代理模型必须用高保真仿真结果训练，且需持续验证。** 代理模型的适用范围有限，超出训练分布的设计方案预测不可靠。
- **关键设计节点仍需高保真仿真校验。** 代理模型用于快速筛选，不用于最终验证。

常见坑

- **用代理模型结果直接作为设计定型依据。** 必须保留高保真验证环节。
- **用局部环节的加速倍数论证整体周期压缩。** 开发周期由多环节共同决定，仿真加速只是其中一环。

适用与规模判断：有自主研发能力的整车厂与一级供应商。仿真计算量大的领域（空气动力学、碰撞安全、NVH、热管理）收益最明显。

场景 4.3.2 智能座舱与车载智能体

成熟度 L3 | 价值层次 创造型。从"识别固定指令的语音助手"升级为"理解复杂自然语言并调用车内外服务完成任务的智能体"，改变的是产品体验而非生产效率。

关键指标	传统做法	引入 AI 后	变化幅度
交互方式	固定唤醒词+预设指令	自然语言，支持复杂多步任务	交互范式改变
功能范围	空调、导航、音乐等 基础车控	可联动舱驾、 调用外部服务	能力边界扩展
后装升级	旧车型无法升级	后装设备可为存量车提供智能体能力	存量市场打开

公开案例与数据

- 据 2026 世界人工智能大会相关公开报道：吉利极氪 8X 搭载"超级 Eva"智能体，支持复杂自然语言任务处理；比亚迪腾势 Z9 GT 配备智能体实现舱驾联动；后装侧，京东 JoyInside 与伴驾精灵推出的设备面向约 **3.66 亿** 辆低智能车型提供升级方案。

落地前提

- **车端算力与云端能力需合理分工。** 完全依赖云端在弱网环境下体验断崖式下降。
- **车控指令的安全边界必须硬性约束。** 涉及行车安全的操作不可由自然语言直接触发。

常见坑

- **在行驶中提供需长时间注意力的交互。** 会带来行车安全风险。
- **用演示场景衡量真实体验。** 展台演示与真实用车环境（噪声、方言、打断）差距很大。

适用与规模判断：整车厂、座舱供应商、后装设备厂商。

场景 4.3.3 Robotaxi 与限定场景无人驾驶

成熟度 L2—L3 | 价值层次 重构型。 封闭与限定场景已实现规模化，开放道路的推进节奏由法规与责任认定规则决定，而非技术能力。

关键指标	传统做法	引入 AI 后	变化幅度
运营验证规模	—	累计订单 35 万余次、安全行驶超 400 万公里（上汽享道公开披露）	已进入规模验证阶段
部署周期	传统方案需逐场景高精地图与调试	无图方案部署周期由一月缩至一天	部署成本大幅下降
算力支撑	—	吉利星睿智算中心 2.0 综合算力达 23.5EFLOPS，可支撑 700 万辆车并发计算	基础设施就位
自研芯片	依赖外部供应	蔚来 NX9031X 累计出货超 30 万颗；理想马赫 M100 单芯片算力 1280TOPS	产业链自主化

公开案例与数据

- 据 2026 世界人工智能大会相关公开报道：**上汽享道 Robotaxi** 完成订单 **35 万余次**、安全行驶超 **400 万公里**，量产定制车计划 2027 年亮相；**九识智能** L4 级无图自动驾驶方案实现规模化量产，部署周期由一月缩至一天。

落地前提

- **限定场景（园区、港口、矿区、干线）的经济性先于开放道路成立。** 场景越封闭、路线越固定，落地越快。
- **远程接管与安全员机制是当前阶段的必需配置。**

常见坑

- **低估运营成本。** 车辆维护、清洁、调度、远程监控的人力成本可能超过节省的司机成本。
- **用测试里程论证安全性。** 里程数不等于安全性，关键是接管率与场景覆盖度。

适用与规模判断：出行平台、车企、封闭场景（港口、矿区、园区）的运营方。

关键变量：责任认定规则。 辅助驾驶向更高阶自动驾驶演进的节奏，主要由事故责任划分规则与保险机制决定，技术能力已非唯一瓶颈。

4.4 电信运营

电信业既是 AI 的使用者，也是算力基础设施的提供者。据三大运营商 2025 年年报公开信息，中国联通披露人工智能收入同比增长超过 140%；中国移动披露数据算法收入同比增长 12.6%，并升级基础大模型、推出上百款 AI 产品及解决方案与数十款垂类智能体。

场景 4.4.1 网络智能运维与故障自愈

成熟度 L3 | 价值层次 替代型→重构型。 电信网络数据规模大、结构化程度高、告警密集，是 AI 落地条件最好的领域之一。

关键指标	传统做法	引入 AI 后	变化幅度
告警处理	海量告警人工逐条研判	自动聚合收敛并输出根因	噪声大幅降低
故障定位	跨专业、跨系统排查耗时长	自动关联多源数据	定位时间压缩
常见故障处置	人工执行处置动作	预案自动执行	闭环处理
值守人力	7×24 轮班值守	人工聚焦于复杂与新型故障	人力重配置

公开案例与数据

- 中国移动设计院等机构公开了基于大数据的 AI 运维服务支撑方案；三大运营商均将网络智能运维作为 AI 应用的重点方向。国际上，电信企业启用代理式 AI 优化网络运维亦已成为趋势。

落地前提

- 告警、性能、配置、变更四类数据必须打通。
- 自动处置动作须分级授权，从"建议"起步逐步放权。

常见坑

- **过早开放高风险自动处置权限。** 错误的自动动作可能把局部故障扩大为大面积中断。

适用与规模判断：电信运营商、大型企业网络运维部门、云服务商。

场景 4.4.2 基站与机房智能节能

成熟度 L3 | 价值层次 增强型。 与场景 4.2.1 同源。电费是运营商的主要运营成本之一，本场景的收益可直接入账。

关键指标	传统做法	引入 AI 后	变化幅度
年节电量	基准	约 8 亿度	可直接量化
年节约电费	基准	约 5.2 亿元	可直接入账
覆盖设施	—	基站、IT 云主机、AB 类机房	多类型覆盖

公开案例与数据

- **中国电信（公开案例）：**自研 AI 节能系统公开披露年均节电约 8 亿度、节约电费约 5.2 亿元。

落地前提

- **必须保证网络覆盖与服务质量不受影响。** 节能策略需与网络性能指标联动约束。

常见坑

- **在业务高峰时段误关断设备。** 会直接影响用户体验与网络 KPI。

适用与规模判断：电信运营商、铁塔运营方、大型数据中心。

场景 4.4.3 号码与业务风控反诈

成熟度 L3 | 价值层次 增强型。 运营商掌握通信行为数据，在诈骗电话识别与拦截上具有天然位置优势。价值主要体现为社会效益。

关键指标	传统做法	引入 AI 后	变化幅度
诈骗识别	基于举报与黑名单， 响应滞后	基于行为模式实时识别	从事后到事中
拦截时效	事后封停	通话过程中预警或拦截	时效提升
新型手法适应	需人工分析后更新规则	模型持续学习	响应加快

公开案例与数据

- 通信反诈是运营商 AI 应用的重点方向，与公安部门的协同机制已较为成熟。因涉及风控细节，具体数据公开披露较少。

落地前提

- **须严格遵守通信秘密保护相关法律要求。**
- **必须建立误拦截的快速申诉与解除机制。**

常见坑

- **误拦截正常业务号码。** 对企业客户的影响极大，需设置白名单与快速处理通道。

适用与规模判断：电信运营商。

关键变量：能否避免在 AI 产业链中被"管道化"。运营商掌握算力与连接，但在模型与应用层的生态位仍需通过合作或自研加以确立。相关公开分析指出，运营商在 AI 领域面临生态不完善的挑战。

4.5 法律与合规服务

法律是典型的"文本密集、专业壁垒高、容错空间小"行业。AI 的价值集中在文档处理与检索环节，而非法律判断本身——因为专业责任始终由执业者承担。

场景 4.5.1 合同审查与风险条款提示

成熟度 L2—L3 | 价值层次 增强型。 法务与律师工作中重复度最高、规则最明确的环节。改善幅度在本白皮书收录的所有场景中位居前列。

关键指标	传统做法	引入 AI 后	变化幅度
简单合同审查	30—120 分钟	10—20 分钟	约 3—6 倍提速
复杂合同审查	4—8 小时	20—30 分钟	约 8—16 倍提速
法律知识检索	约 3 分钟	约 15 秒	约 12 倍提速
案例检索	5—20 分钟	5—20 秒	约 60 倍提速
法律观点整合	30—60 分钟	5—10 分钟	约 6 倍提速
合同条款遗漏风险	约 20%	约 2%	下降约 90%
证据链缺陷预警覆盖	约 55%	约 92%	提升约 67%
初级法务人力投入	某集团 10 人/月	2 人/月	成本降约 80%
外部律师费用	某金融公司 100 万元	16 万元	下降约 84%

公开案例与数据

- 据法律科技服务商公开披露的客户数据：简单合同审查从 30—120 分钟缩短至 10—20 分钟，复杂合同从 4—8 小时缩短至 20—30 分钟；合同条款遗漏风险从约 20%降至约 2%；某集团初级法务人力投入从 10 人/月减少至 2 人/月（成本降约 80%），某金融公司外部律师费用从 100 万元降至 16 万元（降约 84%）；某物流集团合同流转效率提升 90%、协作审查提升 75%、合规管理提升 80%。合同审查准确率区间为 85%—90%。
- **注意：** 上述数据来自服务商披露的客户案例，属于本白皮书 C 级信息源，未经独立审计，且存在明显的选择性披露可能。使用时应作为"可能达到的上限"而非"预期值"。

落地前提

- **必须建立本企业的合同模板库与风险偏好规则。** 通用模型不知道贵司对某类条款的容忍边界。
- **审查结论必须由执业人员确认。** 85%—90%的准确率意味着每 10 份合同就可能有 1 份存在问题，专业责任不可转移。
- **优先从标准化程度高的合同类型入手**（采购、销售、劳动合同、保密协议），非标合同的效果显著下降。

常见坑

- **把"外部律师费下降 84%"当作可复制的预期。** 该数字对应特定企业的特定业务结构，不具普遍性。这类服务商披露数据是本白皮书中幸存者偏差最严重的部分。
- **依赖 AI 审查高价值、非标准的重大合同。** 此类合同的风险点往往在于商业条款的博弈，而非条文合规。

适用与规模判断：合同量大的企业法务部门、处理标准化业务的律师事务所。合同量小或以非标重大交易为主的机构收益有限。

场景 4.5.2 类案检索与法律研究

成熟度 L2—L3 | 价值层次 增强型。 检索增强生成是主流技术路线，价值在于把"找得到"变成"找得全、找得快"。

关键指标	传统做法	引入 AI 后	变化幅度
案例检索	5—20 分钟	5—20 秒	约 60 倍提速
检索方式	关键词检索， 依赖检索技巧	按案情要素语义检索	门槛下降
覆盖完整性	受检索词影响，易遗漏	语义检索覆盖更全	遗漏减少

公开案例与数据

- 类案检索是法律 AI 中应用最广的功能，各主流法律数据库均已集成。

落地前提

- **必须核对检索结果的裁判文书原文。** 生成模型可能编造案号或错误概括裁判要旨。
- **注意法律法规的时效性。** 已修订或废止的规定若被引用，后果严重。

常见坑

- **直接引用 AI 概括的裁判要旨而不核对原文。** 这是法律 AI 最危险的使用方式，国内外均已出现因引用虚构案例而受到处罚的案例。

适用与规模判断：律师事务所、企业法务、司法机关。

场景 4.5.3 尽职调查文档批量处理

成熟度 L2—L3 | 价值层次 替代型。 尽调涉及海量文档的通读与要素提取，是律师工作中人力投入最集中的环节，也是价值可计量性最高的法律 AI 场景。

关键指标	传统做法	引入 AI 后	变化幅度
文档通读	律师逐份阅读	AI 批量提取关键要素	通读环节压缩
要素提取	人工摘录制表	自动生成要素清单	制表工作消除
遗漏风险	人工疲劳导致遗漏	全量处理，覆盖一致	覆盖率提升
计费影响	按小时计费， 客户成本高	可改变服务定价模式	商业模式影响

公开案例与数据

- 尽调文档处理是律所与投行采用 AI 最积极的场景，因其价值可直接换算为节省的人时。

落地前提

- 必须与保密要求匹配的部署方式。 尽调材料高度敏感，须私有化或专属部署。
- 要素提取清单需按交易类型预先定义。

常见坑

- 把提取结果直接作为尽调结论。 提取是第一步，判断仍须律师完成。
- 忽视利益冲突与保密义务。 材料上传至第三方服务可能违反保密约定。

适用与规模判断：从事并购、投融资、IPO 业务的律所与企业法务、投行。

关键变量：专业责任的不可转移性。 法律意见的责任始终由执业者承担，这决定了 AI 在该行业的天花板是第二层增强型价值，短期难以进入第三层。

4.6 人力资源与组织管理

场景 4.6.1 简历解析与初筛

成熟度 L3 | 价值层次 替代型。 招聘漏斗最上层、重复度最高的环节。但也是本白皮书中 **合规风险最需要警惕** 的场景之一。

关键指标	传统做法	引入 AI 后	变化幅度
简历解析	人工阅读提取信息	自动结构化解析	录入环节消除
批量初筛	HR 逐份阅读， 千份简历需数日	批量匹配打分， 大幅压缩处理时间	处理量提升

关键指标	传统做法	引入 AI 后	变化幅度
筛选一致性	因人而异、受疲劳影响	标准统一	波动减少
候选人响应速度	受 HR 处理能力限制	可快速反馈	体验改善

公开案例与数据

- AI 简历筛选在中大型企业招聘中已较为普及。公开材料普遍强调其价值在于处理量的提升与响应速度的改善。
- **注意：**该领域的公开数据多来自 HR SaaS 服务商的营销材料，属 C 级信息源中可靠性较低的一类，本白皮书未采用其具体数字。

落地前提

- **必须完成算法影响评估并保留人工复核。** 招聘涉及就业平等，是各国算法监管的重点领域。
- **筛选标准须与岗位要求有明确关联。** 使用与岗位无关的特征（如毕业院校、性别相关的代理变量）筛选存在歧视风险。
- **须向候选人告知使用了自动化决策工具。**

常见坑

- **用历史录用数据训练模型。** 历史数据中隐含的偏见会被模型学习并放大，这是招聘 AI 最典型也最隐蔽的失败方式。
- **完全自动拒绝候选人。** 建议 AI 只做排序与推荐，拒绝决定由人做出并可申诉。
- **忽视非标准背景的优秀候选人。** 模型倾向于匹配典型画像，会系统性遗漏跨界人才。

适用与规模判断：年招聘量大、岗位标准化程度高的中大型企业。高端与稀缺岗位不适用。

场景 4.6.2 HR 事务智能体与员工问答

成熟度 L3 | 价值层次 替代型。 HR 部门大量时间消耗在重复性的制度咨询上（假期政策、报销标准、社保公积金）。这个场景收益确定、风险低。

关键指标	传统做法	引入 AI 后	变化幅度
制度咨询	HR 逐个答复	员工自助查询	咨询工作量下降
响应时效	受 HR 工作时间限制	7×24 可用	全时可用
答复一致性	不同 HR 口径可能不一	统一知识库	口径统一
HR 角色	大量时间用于事务性答复	转向人才发展与组织建设	价值重心转移

公开案例与数据

- HR 事务智能体是企业内部 AI 应用中落地最快的场景之一，与场景 3.2.5（工业知识问答）技术同源。

落地前提

- 制度文件必须版本清晰、及时更新。
- 涉及个人薪酬、绩效等敏感信息的查询须严格权限控制。

常见坑

- 制度更新后知识库未同步。会给出过期答复并引发劳资争议。
- AI 答复被员工理解为公司正式承诺。须明确提示以正式制度文件为准。

适用与规模判断：员工规模在数百人以上的企业。规模越大、制度越复杂，收益越明显。

场景 4.6.3 面试与情景演练

成熟度 L2 | 价值层次 替代型+创造型。 用于大规模标准岗位的初面筛选，以及内部培训中的模拟演练。前者争议较大，后者价值明确。

关键指标	传统做法	引入 AI 后	变化幅度
初面安排	需协调面试官时间	候选人可随时完成	约面成本消除
评估一致性	不同面试官标准不一	标准统一	波动减少
内部演练	需组织角色扮演，频次低	AI 扮演客户/面试官，可反复练习	从稀缺到充足
高端岗位适用性	面试官 judgment 关键	效果有限	边界明确

公开案例与数据

- AI 面试在大规模校园招聘、连锁行业门店招聘中应用较多。内部培训演练场景的接受度高于对外招聘场景。

落地前提

- 必须告知候选人并取得同意。
- 评估维度须与岗位胜任力模型对齐，且可解释。
- 建议定位为“初筛辅助”，最终录用决定由人做出。

常见坑

- 依赖表情、语调等生物特征做性格与能力判断。此类评估的科学性存在广泛争议，且涉及生物识别信息的合规风险。
- 在高端岗位使用 AI 面试。会造成候选人体验受损与雇主品牌损害。

适用与规模判断：大规模标准岗位招聘（校招、连锁门店、客服中心），以及企业内部培训演练。

关键变量：公平性与合规性。 招聘环节的 AI 应用直接涉及就业平等，是监管关注的重点领域。建议企业在部署前完成算法影响评估并保留人工复核环节。

4.7 保险

场景 4.7.1 车险智能定损与理赔

成熟度 L3 | 价值层次 替代型→重构型。 车险理赔是保险业中标准化程度最高、频次最高的作业，也是 AI 改造最彻底的环节。

关键指标	传统做法	引入 AI 后	变化幅度
定损方式	查勘员到现场拍照定损	用户自助拍照，AI 识别定损	查勘环节可省略
理赔周期	通常以天计	公开案例披露 平均结案时间约 2.16 分钟 (小额简易案件)	数量级压缩
查勘人力	按案件量配置查勘队伍	人力聚焦于复杂与 争议案件	人力重配置
欺诈识别	依赖查勘员经验	可在极短时间内识别 异常模式	风控前置

公开案例与数据

- 据公开报道，AI 技术赋能下汽车出险平均结案时间可达 **2.16 分钟**（对应小额简易案件）；行业内亦有“理赔周期从 3 天缩短至 2 小时”的公开表述。**注意：这两个数字对应的案件类型不同，不可混淆——前者是简易小额案件，后者是常规案件。**

落地前提

- 需要大量的历史定损照片与结果数据。
- 必须设置金额阈值。超过阈值的案件须人工查勘，这是控制欺诈风险的关键。
- 配件价格库需持续更新。

常见坑

- 用简易案件的处理时长宣传整体理赔时效。会造成客户预期落差与投诉。
- 取消人工查勘导致欺诈率上升。自助定损的欺诈风险高于现场查勘，须配套反欺诈模型。

适用与规模判断：车险业务规模较大的财产保险公司。

场景 4.7.2 智能核保

成熟度 L2—L3 | 价值层次 替代型。 标准件自动核保、非标件人工处理的分流机制，把核保人员的精力集中到真正需要判断的案件上。

关键指标	传统做法	引入 AI 后	变化幅度
标准件核保	核保员逐单审核	自动通过	人工介入消除

关键指标	传统做法	引入 AI 后	变化幅度
核保时效	通常需数小时至数日	标准件秒级出结果	时效大幅提升
核保员角色	处理全部案件	聚焦于非标与疑难件	人力重配置
健康险核保	依赖问卷与体检报告 人工判读	可结构化解析 医疗数据辅助判断	依赖数据可及性

公开案例与数据

- 智能核保在寿险与健康险领域推进较快，但受制于医疗数据的可及性，其自动化比例低于车险。

落地前提

- 核保规则必须先明确化、可执行化。依赖核保员经验判断的隐性规则需先显性化。
- 须严格控制自动核保的责任边界与金额上限。

常见坑

- 自动核保规则过松导致逆选择。需持续监控自动通过件的赔付率。
- 处理健康数据时忽视合规要求。医疗健康数据属敏感个人信息。

适用与规模判断：寿险、健康险、意外险业务量大的保险公司。

场景 4.7.3 代理人展业辅助

成熟度 L2 | 价值层次 增强型。 为代理人提供产品问答、方案生成与话术支持。合规话术约束是本场景最主要的设计难点。

关键指标	传统做法	引入 AI 后	变化幅度
产品知识查询	翻阅条款或询问内勤	即时问答	查询效率提升
方案建议书生成	人工制作，耗时长	按客户需求快速生成	制作时间压缩
话术合规性	依赖培训与抽检	统一话术库，可前置校验	合规风险下降
新人上手	依赖师徒带教	可自助学习与演练	培训周期缩短

公开案例与数据

- 代理人展业辅助工具在保险公司中应用广泛，其核心价值在于降低产品复杂度带来的销售门槛。

落地前提

- 所有对客话术必须经合规审核入库。
- 须与适当性管理流程衔接。产品推荐须匹配客户风险承受能力评估结果。

常见坑

- AI 生成的收益演示或产品对比未经合规审核。属销售误导的高风险区。

适用与规模判断：有代理人渠道的保险公司。

4.8 农业

农业是政策关注度上升最快、但数据基础最薄弱的行业。农业农村部于 2026 年 4 月发布通知，公开征集“人工智能+农业”典型应用场景。

场景 4.8.1 病虫害识别与精准施药

成熟度 L2—L3 | 价值层次 替代型+增强型。 图像识别技术本身已成熟，制约因素在于终端设备普及与农户使用习惯。

关键指标	传统做法	引入 AI 后	变化幅度
病虫害诊断	依赖农户经验或 农技员到场	拍照即时识别	诊断可及性提升
施药决策	按经验或固定周期施药	按识别结果精准施药	农药用量下降
农技服务覆盖	农技员人数有限， 难以覆盖分散农户	可覆盖所有联网农户	覆盖面数量级提升

公开案例与数据

- 病虫害识别是“农业+大模型”公开案例中最常见的应用方向。农业农村部已发布多批智慧农业典型案例。

落地前提

- 识别模型需按区域主要作物与病害类型定制。通用模型在特定区域的准确率不足。
- 必须配套农技专家的兜底通道。误判导致的减产损失不可逆。

常见坑

- 误判导致错误用药造成药害。这是不可逆的损失，须谨慎设置置信度阈值。

适用与规模判断：规模化种植主体、农业社会化服务组织、地方农技推广部门。

场景 4.8.2 智能农机与无人作业

成熟度 L2—L3 | 价值层次 替代型。 规模化农场的经济性已经成立，小农户受制于设备成本门槛。

关键指标	传统做法	引入 AI 后	变化幅度
作业方式	人工驾驶农机	自动导航作业	劳动强度下降

关键指标	传统做法	引入 AI 后	变化幅度
作业精度	受驾驶员技能影响	厘米级导航，重叠与遗漏减少	用量与效率优化
作业时段	受人力与光照限制	可夜间连续作业	农时窗口扩展
经济性门槛	—	与经营规模强相关	规模决定可行性

公开案例与数据

- 广东等地已建成智慧"无人农场"示范项目，实现耕种管收全程无人化作业。哈尔滨工业大学等机构在农业 AI 领域亦有公开研究成果。

落地前提

- 地块需相对规整且具备导航基准站覆盖。
- 需配套农机社会化服务组织。单个小农户难以承担设备投入。

常见坑

- 在小规模、碎片化地块投入自动化农机。经济性通常不成立。

适用与规模判断：规模化农场、家庭农场、农机专业合作社、农业社会化服务组织。

场景 4.8.3 农技问答与种植决策助手

成熟度 L1—L2 | 价值层次 创造型。 面向小农户的普惠价值最大——把原本只有大型农企才能获得的农艺指导，变成人人可及。但知识准确性要求极高。

关键指标	传统做法	引入 AI 后	变化幅度
农技咨询获取	依赖农技员下乡或经验	随时可问	可及性从无到有
指导颗粒度	区域性统一指导	可按地块、品种、生育期给出建议	颗粒度提升
知识准确性要求	农技员有专业背书	须建立专家审核机制	风险点

公开案例与数据

- 中国农业大模型的数据治理、核心技术与应用挑战已有专门的学术研究，指出数据质量与区域适配是主要瓶颈。

落地前提

- 知识库须由农业科研机构或农技推广部门背书。
- 必须区分"通用知识"与"具体地块建议"。后者需结合本地土壤、气候数据，不可泛泛而谈。

常见坑

- 给出不适配本地条件的建议造成减产。农业损失以年为周期，代价极高。

适用与规模判断：农技推广部门、农业服务平台、涉农金融机构。

关键变量：经营主体的规模化程度。AI 在农业的经济性与经营规模强相关，规模化农场与小农户面临完全不同的成本曲线。

4.9 建筑与房地产

场景 4.9.1 工地安全行为识别

成熟度 L3 | 价值层次 替代型。建筑业 AI 中成熟度最高、监管推动力最强的场景。与场景 3.2.7 技术同源。

关键指标	传统做法	引入 AI 后	变化幅度
违章发现	安全员巡查抽查	摄像头全时自动识别	覆盖率提升
发现时效	事后或巡检时	实时告警	可干预窗口出现
证据留存	依赖人工记录	自动留存	追溯改善

公开案例与数据

- 工地安全行为识别（未戴安全帽、高空作业未系安全带、危险区域闯入）在大型施工企业中已广泛部署，多由监管要求驱动。

落地前提

- 必须明确告警的处置流程与责任人。
- 摄像头需随工程进度移动布设，运维成本需计入。

常见坑

- 只告警不处置。事故后留存的未处置告警记录会成为不利证据。

适用与规模判断：大中型施工企业、建设单位、监理单位。

场景 4.9.2 图纸审查与工程量核算

成熟度 L2 | 价值层次 替代型+增强型。图纸合规校核与工程量计算都是规则明确、工作量大的环节，但受制于图纸标准化程度。

关键指标	传统做法	引入 AI 后	变化幅度
图纸规范校核	人工逐项对照规范	自动校核	校核工作量下降
工程量核算	人工识图算量，耗时长	基于图纸或 BIM 自动提取	算量效率提升
进度识别	人工现场核对	视觉识别施工进度	核对效率提升

关键指标	传统做法	引入 AI 后	变化幅度
效果制约	—	取决于图纸标准化与 BIM 应用程度	前提条件强

公开案例与数据

- 图审与算量 AI 在设计院与造价咨询机构中已有应用，效果与图纸质量强相关。

落地前提

- 图纸需符合制图标准。不规范的图纸（图层混乱、标注缺失）会导致识别失败。
- BIM 模型的完整度决定算量准确性。

常见坑

- 用 AI 算量结果直接作为结算依据。须经造价工程师复核确认。

适用与规模判断：设计院、造价咨询机构、施工企业成本部门。BIM 应用程度高的项目收益更大。

场景 4.9.3 楼宇能耗与设备智能运维

成熟度 L2—L3 | 价值层次 增强型。与场景 4.2.1 同源，价值直接体现为物业运营成本下降。

关键指标	传统做法	引入 AI 后	变化幅度
空调与照明控制	按固定时段运行	按人流与环境智能调节	能耗下降
设备故障	事后维修	预测性维护	停机减少
能耗成本	基准	可直接换算电费节省	可直接入账

公开案例与数据

- 楼宇智能运维在大型商业综合体、写字楼、园区中应用较广，技术逻辑与工业能耗优化一致。

落地前提

- 需要楼宇自控系统（BA）具备数据采集与远程控制能力。老旧楼宇的改造成本需前置评估。

常见坑

- 为节能牺牲舒适度引发租户投诉。需设置舒适度约束边界。

适用与规模判断：大型商业综合体、写字楼、产业园区、酒店的运营方。

4.10 文化娱乐与内容产业

文娱是唯一一个 AI 改变的不只是效率、而是产出物本身形态的行业。因此传统的"提效"指标在此部分失效。

场景 4.10.1 图像与视频内容生成

成熟度 L3—L4 | 价值层次 替代型+创造型。 与场景 3.1.1 同源但应用更广。已成为内容生产的基础工具，改变的是内容供给的成本曲线。

关键指标	传统做法	引入 AI 后	变化幅度
素材获取	拍摄、外采或素材库购买	按需生成	成本结构改变
概念可视化	需美术人员绘制	快速出图	迭代速度提升
分镜与前期	手工绘制，周期长	快速生成参考稿	前期周期压缩
终稿品质	专业团队品质稳定	仍需专业人员打磨	替代前期而非终稿

公开案例与数据

- 内容生成在广告、影视前期、游戏美术、自媒体等领域已广泛应用。公开材料普遍指出其价值集中在**前期与中间稿**环节，终稿仍需专业人员完成。

落地前提

- 须明确训练素材来源的合法性。
- 需建立风格一致性的控制手段。系列内容的视觉统一是主要挑战。

常见坑

- 版权归属不清导致商用风险。生成内容的著作权归属规则在各法域尚不统一，商业使用前须评估。
- 用生成素材冒充实拍。在广告场景中可能构成虚假宣传。

适用与规模判断：广告公司、影视制作、游戏开发、内容创作者、品牌营销部门。

场景 4.10.2 配音、翻译与本地化

成熟度 L3—L4 | 价值层次 替代型。 内容出海成本下降幅度最大的环节。原本每增加一个语种就要增加一套配音与字幕成本，现在这条曲线被压平。

关键指标	传统做法	引入 AI 后	变化幅度
配音	录音棚+配音演员	语音合成	成本大幅下降
字幕翻译	人工翻译+时间轴对齐	自动生成后校审	周期压缩

关键指标	传统做法	引入 AI 后	变化幅度
可覆盖语种	受配音资源限制	模型覆盖语种均可尝试	长尾市场可行
情感表现力	配音演员优势明显	常规内容接近可用，情感戏仍有差距	因内容类型而异

公开案例与数据

- 配音与本地化是内容出海领域 AI 应用最成熟的环节，公开材料普遍将其列为成本下降幅度最大的方向之一。

落地前提

- 声音克隆必须取得声音权利人的书面授权。
- 需建立术语表与人名地名对照表。

常见坑

- 未经授权克隆他人声音。涉及明确的人格权侵权风险。
- 情感类内容直接使用合成配音。观众对情感表现的敏感度高于对发音准确度的敏感度。

适用与规模判断：内容出海企业、影视译制机构、在线教育、有声内容平台。

场景 4.10.3 游戏 NPC 与动态内容

成熟度 L2—L3 | 价值层次 创造型。 改变的是产品体验而非生产效率——NPC 从“有限对话树”变为“可自由对话的角色”，这是原本不可能提供的体验。

关键指标	传统做法	引入 AI 后	变化幅度
NPC 对话	预写对话树，分支有限	可自由对话	体验形态改变
内容量	受策划与文案产能限制	可动态生成	内容量级提升
一致性风险	预写内容可控	**需防止角色崩坏与不当内容**	**新增风险
运行成本	几乎无边际成本	每次对话产生推理成本	成本结构改变

公开案例与数据

- 大模型驱动的 NPC 是游戏行业探索最活跃的方向之一，但受制于推理成本与内容可控性，规模化商用仍在早期。

落地前提

- 必须建立角色设定的强约束与内容过滤。
- 需精算单用户的推理成本。高频对话场景的成本可能超过收入。

常见坑

- **内容失控引发舆情。** 玩家会主动尝试诱导 NPC 产生不当内容，须做对抗性测试。
- **低估推理成本。** 这是本场景商业化的主要障碍。

适用与规模判断：游戏开发商，尤其是角色扮演、模拟经营、社交类游戏。

关键变量：版权与权益规则。 训练数据来源、生成内容的著作权归属、表演者形象权与声音权保护等问题，是该行业规模化商业应用的主要不确定性来源。

4.11 餐饮与生活服务

场景 4.11.1 备货预测与损耗控制

成熟度 L2—L3 | 价值层次 重构型。 餐饮与生鲜零售的核心成本痛点。与场景 3.1.3 同源，但因食材保质期短，损耗成本更高，单位收益更大。

关键指标	传统做法	引入 AI 后	变化幅度
备货决策	店长凭经验估算	基于历史、天气、节假日的预测	依据改变
食材损耗	经验备货导致过剩或短缺	损耗率下降	直接影响毛利
缺货导致的销售损失	热销品早早售罄	预测驱动备货	销售机会保留
店长负担	每日耗时估算订货	系统生成建议单	事务负担下降

公开案例与数据

- 参见场景 3.1.3 多点（Dmall）案例——缺货率从最高约 7% 降至 2% 以内，库存周转从最高约 35 天降至普遍 21 天以内。餐饮场景的技术逻辑一致，但周转周期更短、预测颗粒度要求更细。

落地前提

- **需要按门店、按菜品、按时段的销售数据。** 只有门店总销售额无法支撑备货预测。
- **必须纳入天气与周边活动等外部变量。** 餐饮销量对这些变量高度敏感。

常见坑

- **忽视门店执行偏差。** 店长不按建议订货，收益归零。
- **新店与新品无历史数据。** 需保留人工决策通道。

适用与规模判断：连锁餐饮、生鲜零售、便利店。门店数越多、食材损耗率越高，收益越明显。

场景 4.11.2 评价分析与服务质量管理

成熟度 L3 | 价值层次 增强型。把散落在各平台的顾客评价自动归类分析，从"看评分"变为"看具体问题分布"。

关键指标	传统做法	引入 AI 后	变化幅度
评价处理	人工抽查， 难以系统归纳	全量自动分类与 情感分析	从抽样到全量
问题定位	知道分数低， 不知道为什么低	定位到具体菜品、 服务环节、门店	颗粒度提升
响应时效	定期汇总，滞后	异常实时告警	响应加快
多平台整合	各平台分别查看	统一归集分析	视图统一

公开案例与数据

- 评价分析在连锁餐饮与酒店行业应用广泛，是投入低、见效快的典型场景。

落地前提

- 分类体系需与内部管理动作对应。分类结果要能直接指向责任部门与改进动作。

常见坑

- 只做分析报告不做改进闭环。报告本身不改善服务质量。

适用与规模判断：连锁餐饮、酒店、生活服务连锁品牌。门店数越多，收益越明显。

场景 4.11.3 智能排班与人效优化

成熟度 L2 | 价值层次 增强型。基于客流预测生成排班建议，解决"高峰缺人、低谷闲人"的结构性问题。

关键指标	传统做法	引入 AI 后	变化幅度
排班依据	店长经验	基于分时客流预测	依据改变
排班编制时间	店长每周耗时编排	自动生成建议	事务负担下降
人效	高峰不足、低谷冗余	按需匹配	人效提升
员工满意度	排班随意性引发不满	规则透明可预期	可改善，也可能恶化

公开案例与数据

- 智能排班在连锁餐饮与零售中已有应用，其效果受劳动法规与员工接受度约束。

落地前提

- 必须遵守劳动法规对工时、休息、加班的规定。
- 需纳入员工偏好与技能矩阵。只按客流排班而不考虑员工技能搭配会影响服务质量。

常见坑

- 过度碎片化排班损害员工权益。会推高流失率，长期成本高于人效收益。这是本场景最需警惕的问题。

适用与规模判断：连锁餐饮、零售门店、客服中心等客流波动明显的业态。

4.12 科学研究与专业服务

科研是国务院“人工智能+”六大行动中列于首位的领域（人工智能+科学技术），其价值属于第四层创造型——AI 不是加快了原有的研究流程，而是让某些原本不可行的研究路径变得可行。

场景 4.12.1 文献综述与知识发现

成熟度 L2—L3 | 价值层次 增强型。已成为研究者的日常工具。价值在于扩大可覆盖的文献范围，而非替代阅读。

关键指标	传统做法	引入 AI 后	变化幅度
文献筛选	人工阅读标题摘要筛选	批量筛选与摘要提取	筛选效率提升
跨领域发现	受研究者知识边界限制	可发现跨领域关联	视野扩展
综述初稿	人工整理归纳	快速形成结构化初稿	成稿加快
引用准确性	人工核对	须逐条核对原文	风险点

公开案例与数据

- 文献辅助是科研场景中普及度最高的 AI 应用。学术界普遍要求在论文中披露 AI 工具的使用情况。

落地前提

- 所有引用必须回溯原文核对。
- 须遵守所在期刊与机构对 AI 使用的披露要求。

常见坑

- 引用 AI 编造的文献。已有多起因此导致论文撤稿的公开案例，这是学术场景中最严重的失败方式。
- 用 AI 摘要替代原文阅读做关键判断。

适用与规模判断：高校、科研院所、企业研发部门。

场景 4.12.2 科学计算代理模型加速

成熟度 L2 | 价值层次 创造型。 用神经网络替代部分数值求解，把计算时间压缩数个数量级，从而使原本因算力成本不可行的大规模探索变得可行。

关键指标	传统做法	引入 AI 后	变化幅度
单次仿真计算	数值求解， 以小时至天计	代理模型推理， 以秒至分钟计	数量级加速
参数空间探索	受计算成本限制， 只能试少量方案	可大规模扫描	探索范围扩大
精度	数值方法精度可控	代理模型精度 受训练分布限制	需高保真校验

公开案例与数据

- 参见场景 4.3.1 捷豹路虎案例——空气动力学计算从约 4 小时缩短至约 1 分钟。同类方法在材料、气象、生物、流体等领域均有进展。

落地前提

- 需要足量高保真计算结果作为训练数据。训练数据的生成本身消耗大量算力，是前置投入。
- 必须明确代理模型的适用边界并做外推检验。

常见坑

- 在训练分布之外使用代理模型。预测结果可能严重失真且无警示。

适用与规模判断：有大规模仿真计算需求的科研机构与工业研发部门。

场景 4.12.3 咨询与审计的文档密集型工作

成熟度 L2—L3 | 价值层次 替代型。 专业服务业的直接提效来源。与场景 4.5.3（尽职调查）技术逻辑一致。

关键指标	传统做法	引入 AI 后	变化幅度
资料通读	人工逐份阅读	批量提取要素	通读环节压缩
底稿整理	人工制表汇总	自动生成结构化底稿	制表工作消除
异常识别	抽样检查	全量扫描异常项	从抽样到全量
专业判断	由执业人员负责	仍由执业人员负责	边界明确

公开案例与数据

- 文档密集型工作的 AI 化是四大会计师事务所与管理咨询公司投入最集中的方向，其价值可直接换算为节省的人时与提升的项目承载量。

落地前提

- 须符合执业准则对工作底稿的要求。
- 客户资料的保密与部署方式须匹配。

常见坑

- 用 AI 输出替代执业判断。审计与鉴证意见的责任不可转移。
- 将客户敏感资料上传至不合规的服务。

适用与规模判断：会计师事务所、管理咨询公司、税务师事务所、评估机构。

第五章 七条落地法则

把七十余张场景卡片放在一起比对，会看到一些反复出现的模式。本章提炼七条法则，它们不依赖具体行业，也不依赖具体技术路线。每条法则都附有可直接执行的自检动作。

5.1 法则一：窄而深，胜过宽而浅

这是所有观察中最稳固的一条。国际咨询机构的调研显示，绩效领先的企业平均聚焦的 AI 用例数量明显少于其他企业（约 3.5 个 vs 约 6.1 个），但预期投资回报显著更高（约 2.1 倍）。

本白皮书的场景卡片也印证了这一点：所有达到 L3—L4 成熟度的场景，无一例外都经历了“单点做深—标准化—横向复制”的过程。施耐德电气的视觉质检从单条产线做到 10 个工厂 20 条产线，北京世纪坛医院的病历生成从 3 个专科做到 9 个专科——**都是先把一个场景做到可复制，再谈扩展。**

自检动作：制定 AI 规划时，把场景清单按 SVA-3D 打分排序后，只保留前 20%，其余明确列为“暂缓”并写入文档。不写“暂缓清单”的规划，实际执行中必然会被稀释。

5.2 法则二：把七成资源投给人和流程

国际咨询机构提出的“10—20—70”经验法则——10%的精力用于算法、20%用于数据与技术、70%用于人员、流程与文化——在本白皮书考察的各行业案例中反复得到印证。

江阴银行的案例是一个清晰的例证：其公开披露的成效中，“资料漏采、错采、模糊上传等问题减少 95%”这一项，靠的不是识别模型有多强，而是**把校验环节前移到了客户上传的瞬间**——这是流程设计，不是模型能力。

自检动作：检查项目预算结构。如果模型与算力费用占比超过一半，而流程改造、培训、变更管理的预算接近于零，该项目大概率会以“系统上线了但没人用”告终。

5.3 法则三：先定度量衡，再谈技术方案

本研究院观察到的失败项目中，最普遍的共同点不是技术选型错误，而是**立项时没有定义清楚“成功长什么样”**。

建议在任何 AI 项目立项时，强制回答四个问题：

1. 这个场景当前的基线指标是多少？（例如：漏检率、平均处理时长、一次解决率、单件内容成本）
2. 这个指标的历史波动区间有多大？（如果改善幅度小于自然波动，就无法归因）
3. 改善多少才算成功？谁来确认？
4. 如果改善了，收益归属于哪个部门的预算？（无人认领的收益等于没有收益）

一个具体的反面案例模式

某企业上线 AI 客服，半年后汇报“机器人接待率达到 85%”。管理层追问节省了多少成本，结果发现：客服团队一个人没减，因为剩下 15% 的复杂咨询恰恰最耗时；而人工坐席的平均通话时长因为只处理难题反而上升了。

问题不在于 AI 没用，而在于**从一开始就选错了指标**。“机器人接待率”这个指标可以通过让用户找不到人工入口来轻松刷高，与真实服务质量甚至负相关。如果立项时定的是“一次解决率”与“单次服务综合成本”，整个项目的设计路径会完全不同。

5.4 法则四：区分“AI 提示”与“AI 决定”

本白皮书在六个深度行业中反复看到同一条分界线：AI 提供输入的场景推进顺利，AI 直接产生法律或安全后果的场景推进缓慢。这条线不是技术能力的边界，而是**责任归属的边界**。

表 5-1 两类 AI 输出的设计要求差异

设计要素	AI 提示（辅助输入）	AI 决定（直接生效）
责任主体	使用者（医师、客户经理、编码员）	需明确到机构与具体审批人
留痕要求	记录 AI 建议与人工采纳情况	**全链路留痕，含模型版本、输入数据、输出结果、时间戳**
复核机制	使用者自行判断	须设强制复核或抽检机制
回退设计	停用即可，人工流程仍在	须有明确的回退路径与责任人
可解释性	有帮助但非必需	必需，且须可对外说明
典型场景	影像检出提示、合同风险提示、投研信息聚合	自动授信审批、自动理赔结案、自动行政决定

自检动作：在系统设计阶段就明确标注每个 AI 输出属于“提示”还是“决定”，并为二者设计不同的日志、复核与追溯机制。把两者混在同一套架构里，会导致在合规审查时无法自证清白。

5.5 法则五：数据治理独立立项

在制造、医疗、金融等行业的案例中，数据准备成本普遍占项目总成本的大头，且极易被低估。更重要的是，数据治理的收益是跨场景复用的——同一套数据资产可以支撑多个后续场景。

本白皮书的场景卡片中，“落地前提”一栏出现频率最高的三类要求分别是：**数据打通、主数据治理、知识库版本管理**。这三类工作都具有一个共同特点：单独看它们不产生业务价值，但缺了它们上层场景全部失效。

自检动作：将数据治理从单个 AI 项目中拆出，作为独立的基础设施项目立项和考核。绑定在单个业务项目里的数据治理，会因为该项目的进度压力而被压缩到最低限度。

5.6 法则六：L3 级场景采购，L1—L2 级场景自建

这条法则来自第二章成熟度阶梯的直接推论，但在实践中被违反的频率极高。

已经达到 L3 级的场景，市场上存在成熟产品，价格与交付周期可预期，自建的唯一理由是数据不能出域——而这通常可以通过私有化部署解决。反之，处于 L1—L2 级的场景高度依赖企业自身的业务 know-how，外部供应商难以理解其中的隐性规则，此时自建（或深度共建）才是合理的。

表 5-2 按成熟度的建购决策参考

成熟度	本白皮书中的典型场景	建议决策
L4	商品图生成、详情页文案、反欺诈、推荐排序、口语陪练	直接采购。 按常规 IT 采购流程，重点比价与比服务，不必做技术评估
L3	视觉质检、预测性维护、智能客服、影像辅助检出、无人机巡检、智能节能、编码辅助	优先采购。 自建只在数据合规要求特殊时考虑，且应选私有化部署
L2	智能补货、生产排程、授信报告生成、学情分析、材料预审、网格事件识别	深度共建。 供应商提供技术底座，企业投入业务专家共同定义规则
L1	基层诊疗辅助、应急预案生成、采购寻源智能体、农技决策助手	谨慎试点。 小额投入验证可行性，不做规模化承诺
L0	各类仅有演示的概念	不投入。 保持技术跟踪即可

自检动作：如果贵单位的 AI 团队正在投入大量人力开发一个市场上已有多家成熟供应商的功能，这部分投入的机会成本值得重新评估。

5.7 法则七：为“退出”做设计

这是最容易被忽略的一条。多数 AI 项目在设计时只考虑“如何上线”，不考虑“如果效果不达预期如何回退”。结果是，一旦发现问题，业务流程已经被改造得无法回到原状，只能带着有缺陷的系统继续运行。

自检动作：在上线前明确三件事——回退触发条件（什么指标恶化到什么程度就停用）、回退操作路径（人工流程是否仍然可用）、以及回退的决策人。这三件事写清楚，实际上会显著提高业务部门接受 AI 的意愿，因为它降低了尝试的风险。

七条法则的自检清单

- 1. 聚焦：**本年度 AI 重点场景是否已收敛到 3 个以内，并有明确的“暂缓清单”？
- 2. 投入结构：**流程改造与人员培训的预算，是否达到模型与算力预算的两倍以上？
- 3. 度量衡：**每个在建场景，是否都有书面记录的基线指标与验收标准？
- 4. 责任边界：**系统中每个 AI 输出，是否已明确标注为“提示”或“决定”？
- 5. 数据治理：**数据工作是否已独立立项，而非附属于某个业务项目？
- 6. 建购决策：**自建清单中，是否存在市场上已有成熟产品的功能？
- 7. 退出设计：**每个已上线场景，是否有书面的回退条件与路径？

第六章 这笔账到底怎么算

本白皮书收录的公开案例几乎都只披露收益，不披露投入。这一章补上另一半——完整的账应该包含什么，以及哪些成本最容易被漏掉。

6.1 收益侧：四类节省，只有两类能直接入账

第二章 2.4 节提出的提效账本模型，其核心判断是：AI 带来的节省分散在四种性质完全不同的类别里，而只有其中两类能直接进损益表。

表 6-1 四类节省及其入账方式

节省类型	本白皮书中的典型例证	能否直接入账	计量方式
外采成本节省	商拍摄影与模特费（场景 3.1.1）； 外部律师费（场景 4.5.1）； 外部翻译与配音（场景 4.10.2）	能，且最直接	直接对比采购合同金额。 是最容易被财务认可的收益， 应作为立项材料的主线
产能扩张	客户经理人均日处理量提升 60% （场景 3.3.2）； 数字人直播新增时段（场景 3.1.2）	能，体现在收入侧	单位时间处理量×单位业务毛利。 须证明新增产能确实 转化为了业务量
人力时间释放	医生病历书写从 8 小时降至 25 分钟 （场景 3.4.1）； 教师备课时间压缩（场景 3.5.1）	通常不能 （除非编制真的减少）	只能作为“可再投入产能”记录。 直接折算为工资节省 会被财务质疑
风险成本降低	漏检率降至 0（场景 3.2.1）； 合同条款遗漏风险从 20%降至 2% （场景 4.5.1）	间接，需长周期观测	历史差错造成的年均损失 ×差错率下降幅度。 须有历史损失数据支撑

两个必须记住的换算规则

规则一：人力释放要写清楚“释放去做什么”。“客户经理每天少花 3 小时做材料”本身不是收益，“因而人均日处理业务量提升 60%”才是收益。前者是投入的减少，后者才是产出的增加——财务只认后者。

规则二：不要把单环节改善当作端到端改善。编码效率提升 50%不等于交付周期缩短 50%（场景 3.3.3）；早期药物筛选加速不等于新药上市提前（场景 3.4.6）；仿真计算加速 240 倍不等于研发周期缩短 240 倍（场景 4.3.1）。这是本白皮书中最常出现的论证错误。

6.2 投入侧：六类成本，后三类最容易被漏掉

表 6-2 AI 项目的完整成本构成

成本类别	包含内容	易被低估的程度
1. 软件与模型	模型调用费或授权费、 平台软件费用	低（通常报价清晰）
2. 算力与基础设施	推理算力、训练算力、 存储、网络	中（推理成本随使用量增长， 上线后可能超预算）

成本类别	包含内容	易被低估的程度
3. 系统集成	与既有业务系统的对接、 接口开发、单点登录	中
4. 数据准备	数据采集、清洗、标注、 主数据治理、知识库整理	高。本白皮书多个场景显示 这是最大单项成本
5. 硬件与工程改造	传感器加装、摄像头点位与 角度改造、光源、工位固定、 网络布线	极高。制造与能源场景中 常超过软件成本
6. 持续运维	模型迭代、知识库更新、 规则维护、异常处理、 产品换型后重训	极高。几乎所有项目都 按一次性投入测算， 实际是持续支出

隐性成本清单

以下成本在本白皮书收集的案例中反复出现，但极少出现在立项预算中：

- **业务专家的时间占用。** 缺陷标注需要资深质检员、病历模板定制需要临床医生、合同规则梳理需要资深法务——占用的都是组织中最稀缺的人力。这项成本往往不进预算，但真实存在。
- **审校与复核环节的新增工作量。** AI 产出需要人来核。若审校工时未纳入测算，实际节省会显著低于宣称值。建议在试点期专门记录“各单位产出的审校工时”。
- **变更管理与培训。** 包括流程重新设计、操作手册编写、员工培训、试运行期的双轨并行。这部分对应“10—20—70”法则中的 70%。
- **换型与迭代成本。** 制造业每次产品换型可能需要重新采样训练；政务领域每次政策更新需要知识库同步；金融领域每次监管规则调整需要规则重配。
- **失败试点的沉没成本。** 按本白皮书第一章引用的研究，多数试点不产生可衡量收益。理性的预算应包含一定比例的试错准备。

6.3 一个可操作的测算流程

1. **确定基线（必须先做）。** 用 1—2 个月记录当前状态的真实数据：单件处理时长、差错率、人力投入、外采金额。**没有基线，后面所有计算都无意义。**
2. **按表 6-1 识别收益类型。** 优先寻找有外采支出可替代的收益项，这类收益最容易通过财务评审。
3. **按表 6-2 列全投入项，特别注意后三类。** 数据准备、硬件改造、持续运维三项应至少占预算的一半以上，若不足，说明预算估算偏乐观。
4. **设定观测期与对照方式。** 单环节改善用前后对照即可；涉及多变量的（如营销转化、教学效果）必须设对照组。
5. **计算三种情景。** 乐观情景（达到公开案例披露水平的 70%）、基准情景（40%）、悲观情景（20%）。**用基准情景做决策，用悲观情景做风险评估。** 本白皮书建议对公开案例数据统一打折使用，原因是幸存者偏差。
6. **设定复盘节点。** 至少在 3 个月、6 个月、12 个月做三次复盘，并预先约定“什么情况下停止投入”。

附录 D 提供了可直接填写的测算表模板。

第七章 失败模式与风险治理

本白皮书的案例来自公开渠道，而公开渠道系统性地偏向成功案例。这一章尝试部分弥补这个偏差：先归纳失败模式，再谈风险治理。

7.1 九种反复出现的失败模式

以下九种模式，是从场景卡片的“常见坑”栏目中横向归纳出的高频项。

表 7-1 AI 项目的九种典型失败模式

编号	失败模式	典型表现	本白皮书中的对应场景
F1	指标错配	选了一个可以被“刷”且与真实价值负相关的指标	机器人接待率（3.1.4）、智能应答率（3.6.1）、事件发现量（3.6.4）
F2	局部提效误认为整体提效	用单环节改善幅度论证端到端周期压缩	编码辅助（3.3.3）、药物研发（3.4.6）、仿真加速（4.3.1）
F3	数据准备被低估	预算与工期按“上系统”估算，实际卡在数据	智能补货（3.1.3）、预测性维护（3.2.2）、生产排程（3.2.6）
F4	成本转移而非消除	一端省了，另一端多了，总账未改善	AI 商拍推高退货率（3.1.1）；审校工时抵消生成节省
F5	发现能力超过处置能力	识别系统上线后事件量激增，处置跟不上	网格事件识别（3.6.4）、安全行为识别（3.2.7）
F6	自动化偏见	使用者过度信任 AI 输出，独立判断能力下降	影像辅助（3.4.2）、合同审查（4.5.1）、文献综述（4.12.1）
F7	只上系统不改流程	系统旁挂在旧流程边上，不改变任何环节	智能客服、材料预审、排程等多个场景
F8	前提条件不具备强行上马	数据、标准、制度前提未满足即启动	材料预审需事项标准化（3.6.3）、学情分析需作业数字化（3.5.4）
F9	合规风险后置	上线后才发现个人信息、版权、适当性管理等问题	招聘筛选（4.6.1）、内容生成（4.10.1）、客流分析（3.1.8）

如何使用这张表

建议在 AI 项目的方案评审环节，把这九条作为强制检查项，逐条要求项目组说明“本项目是否存在该风险，如何规避”。

本研究院的经验是：**F1（指标错配）与 F3（数据准备被低估）** 合计解释了大多数项目的失败。如果时间有限，至少确保这两条被严格审查。

7.2 风险图谱：从内容风险到行动风险

随着 AI 形态从对话式助手转向任务型智能体，风险的性质发生了根本变化。助手阶段的风险主要是"说错话"，智能体阶段的风险是"做错事"——后者具有不可逆性和连锁扩散的可能。

表 7-2 行业 AI 应用的风险图谱

风险类别	具体表现	高发场景	主要控制手段
输出准确性	事实性错误、专业判断偏差、引用来源虚构	医疗、法律、投研、公文	检索增强+来源标注；置信度阈值；强制人工复核
行动越权	智能体执行超出授权范围的操作、误触发不可逆动作	金融、政务、能源、运维	最小权限原则；不可逆操作强制二次确认；操作全链路留痕
数据与隐私	敏感数据外泄、训练数据合规瑕疵、跨境传输违规	医疗、金融、政务、人力资源	数据分级分类；私有化或专属部署；去标识化处理
公平性	算法歧视、群体性服务能力差异	招聘、授信、政务服务	算法影响评估；保留人工申诉通道；定期偏差检测
依赖与能力退化	自动化偏见、专业人员判断能力下降、新人成长路径中断	医疗、法律、教育、公文写作	保留独立判断环节；监测采纳率与推翻率；对新人保留手工训练
知识产权	训练素材来源不清、生成内容权属不明、肖像与声音权侵害	内容生成、配音本地化、数字人	要求供应商就素材来源书面承诺；取得形象与声音授权
安全对抗	提示注入、模型窃取、供应链投毒	全行业	输入过滤；沙箱隔离；第三方组件审计
责任归属	出现问题时机构、供应商、经办人之间责任不清	全行业	合同条款前置约定；建立 AI 决策档案；明确审批链路

7.3 国内外研究的共同指向

麦肯锡公司 2026 年 3 月发布的负责任 AI 成熟度调研提供了一组值得关注的数据：近三分之二的受访者将安全与风险问题列为全面扩展智能体应用的首要障碍；74%的企业认为模型输出不准确是高度相关的风险；72%将网络安全视为关键风险。该调研同时显示，各组织的负责任 AI 成熟度平均分从上一年的 2.0 升至 2.3（5 分制），但仅约 30%的组织在战略、治理与智能体控制方面达到 3 级及以上；而明确分配了责任主体的组织，其成熟度评分（2.6）显著高于责任不清晰的组织（1.8）。

中国信息通信研究院在《人工智能产业发展研究报告（2025 年）》中同样将安全治理单列成章，强调需要构建覆盖全生命周期的防护体系。

两份来自不同体系的研究指向同一个结论：**治理能力已经成为 AI 规模化的前置条件，而非事后补充。**

7.4 一个可操作的治理起步方案

对于尚未建立 AI 治理体系的组织，本研究院建议按以下顺序推进，避免一开始就追求完备的制度框架而迟迟无法落地。

1. **第一步：建立 AI 应用台账（1 个月内）。**摸清组织内已经在用的所有 AI 工具与场景，**包括员工自发使用的外部工具**。多数组织在做这一步时会发现，实际使用范围远超管理层认知，这本身就是最大的风险敞口。
2. **第二步：按风险分级（1—2 个月）。**用表 7-2 的风险类别对台账中的每个场景打标，识别出高风险场景（涉及个人敏感信息、产生法律后果、执行不可逆操作）。分级的目的是把有限的管理资源集中到少数高风险场景上。
3. **第三步：为高风险场景建立三项基础机制（2—4 个月）。**一是明确的责任人；二是完整的操作留痕；三是可执行的回退方案。这三项是最低要求，也是出现问题时自证合规的基础。
4. **第四步：制定分级授权规则（4—6 个月）。**明确哪些操作 AI 可自主执行、哪些需人工确认、哪些禁止 AI 参与。这份规则应由业务、法务、技术三方共同签署，而非由技术部门单方面制定。
5. **第五步：建立定期评估机制（持续）。**至少每半年复核一次台账与分级，重点关注新增场景与能力边界变化。

不要把治理做成“审批”

治理体系建设中最常见的失败模式，是把治理简化为增加审批环节。这会带来两个后果：一是业务部门为规避流程而使用未登记的外部工具，风险从可见变为不可见；二是治理部门被大量低风险事项淹没，无力关注真正的高风险场景。

有效的治理设计应当是**分级放行而非一律审批**：低风险场景给出明确的“绿灯清单”，业务部门可自主使用；管理资源集中投向高风险场景的深度介入。

第八章 趋势展望（2026—2030）

本章的判断属于趋势推演而非预测。我们列出五个方向，并为每个方向给出观察指标——读者可以据此自行验证这些判断是否成立。

8.1 智能体渗透进入企业应用主流

研究机构 Gartner 在 2025 年 8 月的预测路线图是：2026 年底约 40% 的企业应用将集成任务型 AI 智能体（2025 年不足 5%）；2027 年约三分之一的场景实现多智能体协作；2028 年智能体生态开始跨应用运作；2029 年至少半数知识工作者掌握相关技能。

对本白皮书场景的影响：目前停留在 L2 的多个场景（智能补货、材料预审、生产排程、进件处理），其瓶颈恰恰是“需要跨系统执行动作”。智能体能力的成熟将直接推动这批场景向 L3 迁移。

观察指标：企业内部 AI 相关岗位描述中“智能体编排”“流程重设计”类职责出现的频率；以及 AI 相关支出从“软件采购”科目向“流程改造”科目的迁移比例。

8.2 价值计量方式从“节省工时”转向“闭环单量”

这是第一章判断三的自然延伸。当 AI 承担完整流程而非片段任务时，“人均节省小时数”这一指标将逐渐失去意义，取而代之的是自动闭环率、人工干预率与单位处理成本。

这一转变对企业是好事——因为第六章指出，“人力时间释放”这类收益通常无法直接入账，而“闭环单量”可以。计量方式的转变将使 AI 项目的财务论证变得容易。

观察指标：行业报告与企业公告中，AI 成效表述方式的变化——从“提升效率 X%”转向“自动处理了 X 万单”。

8.3 行业专用模型与通用模型形成分层

政策层面已出现明确的分层设计。《“人工智能+制造”专项行动实施意见》同时提出“推动 3~5 个通用大模型深度应用”与“形成特色化、全覆盖的行业大模型”；能源领域的实施意见提出到 2027 年推动 5 个以上专业大模型应用；教育领域提出建设教育专用大模型。

这意味着未来的技术格局大概率不是“一个通用模型解决一切”，而是**通用模型提供基础能力、行业模型承载领域知识、企业私有数据完成最后适配**的三层结构。本白皮书中多个场景的“落地前提”栏反复强调“需按本行业/本企业定制”，正是这一结构的体现。

观察指标：各行业协会或链主企业牵头建设的行业数据集与行业模型数量；企业 AI 架构中“行业模型层”是否成为独立的技术选型环节。

8.4 数据资产成为能力差距的主因

当模型能力在行业内趋于同质（可通过采购获得），差异化就只能来自数据。本白皮书在电商、金融、汽车、政务四个领域都观察到同一现象：领先者的优势并非来自更好的算法，而是来自更完整的行为数据、更长的历史积累与更顺畅的数据回流机制。

辽宁 12345 热线案例中的"4 亿余条诉求数据"、上汽享道 Robotaxi 的"400 万公里安全行驶"、多点智能补货所依赖的多年销售历史——**这些都不是可以采购的东西。**

观察指标：企业是否将数据资产纳入正式的资产管理与价值评估体系；行业内是否出现规范化的数据交换与共享机制。

8.5 治理能力从成本项转为竞争力

麦肯锡的调研已观察到，责任更清晰的组织在 AI 价值实现上表现更好。本研究院认为这一相关性在未来数年会进一步强化：在高监管行业，治理体系成熟的机构将获得监管更大的放权空间，从而在应用深度上拉开差距。

这对金融、医疗、政务行业尤其重要。本白皮书反复指出，这三个行业的场景成熟度上限由"敢不敢放"而非"能不能做"决定。率先建立起可审计、可追溯、责任清晰的治理体系的机构，将在监管口径明确后获得先发优势。

观察指标：监管部门是否出现基于机构治理成熟度的差异化监管安排；企业 AI 治理岗位从合规部门向业务部门下沉的程度。

三个需要警惕的反向风险

一是能力同质化带来的投资回报下降。当某类场景全行业普及（如制造业质检、电商内容生成），它带来的将不再是超额收益，而只是维持竞争的必要成本。企业需要持续把投入前移到尚未普及的场景。

二是过度自动化引发的韧性下降。高度依赖 AI 的流程在遭遇模型故障、服务中断或数据异常时，可能因缺乏人工冗余而全面停摆。建议对关键流程保留可用的人工路径并定期演练。

三是技能断层。若初级岗位的工作被大量替代，专业人员的成长路径将被打断，可能在若干年后形成中高级人才断层。这一风险在医疗、法律、审计、公文写作等依赖长期经验积累的领域尤其突出——本白皮书在多张场景卡片中都提示了这一点。

第九章 给不同角色的行动建议

9.1 给企业决策者

1. **做减法而不是加法。**把 AI 场景清单收敛到三个以内，其余明确标注为暂缓。规划文档中没有“暂缓清单”的，通常意味着没有真正做过取舍。
2. **优先选有外采支出可替代的场景。**这类场景的收益最容易通过财务评审（见第六章表 6-1）。电商内容生产、法务外部律师费、翻译配音是典型代表。
3. **检查预算结构而不是技术方案。**你不需要懂模型，但你需要看得懂预算表：如果数据准备、硬件改造、持续运维三项合计不足预算的一半，这个估算大概率偏乐观。
4. **要求书面的度量衡。**在批准任何 AI 项目前，要求提供基线指标、验收标准与收益归属部门，三者缺一不可。
5. **亲自过问退出方案。**问一个问题：“如果三个月后效果不达预期，我们怎么退回去？”这个问题回答不上来的项目，不应该批准。
6. **对公开案例数据统一打折。**建议按公开披露值的 40% 作为基准情景做决策。原因是幸存者偏差——失败案例不会被公开。

9.2 给业务部门负责人

1. **从你最讨厌的重复劳动开始，而不是从最重要的业务开始。**前者容错空间大、员工配合度高，是建立组织信心的最佳起点。
2. **参与场景定义，不要外包给技术部门。**技术部门不了解你的隐性业务规则，由其主导定义的场景往往在细节处失效。本白皮书中大量“落地前提”都属于业务知识，而非技术知识。
3. **提前约定可接受的错误率。**尤其是会给一线增加复核负担的场景（质检、风控、客服）。若误报率过高，系统会被绕过或关闭——这是本白皮书观察到的高频失败方式。
4. **为收益找到归属。**主动认领 AI 带来的指标改善并纳入部门考核，是保证项目持续投入的前提。
5. **警惕“成本转移”。**检查一端的节省是否变成了另一端的增加（生成省了但审校多了、内容省了但退货多了）。

9.3 给技术负责人

1. **先做数据资产，再做模型选型。**在多数行业，数据准备的工作量远大于模型工作量，且收益可跨场景复用。
2. **明确区分“提示”与“决定”两类输出，并为其设计不同架构。**这是未来通过合规审查的基础（见第五章表 5-1）。
3. **避免重复造轮子。**对已达 L3 级的成熟场景，优先评估采购或私有化部署，把自研资源投向真正体现企业 know-how 的 L1—L2 场景。
4. **建设可观测性优先于建设新功能。**没有完整日志与指标体系的 AI 系统，无法证明其价值，也无法定位其问题。
5. **为持续运维预留资源。**模型迭代、知识库更新、规则维护、换型重训——这些是持续支出而非一次性投入。

9.4 给行业组织与政策研究者

1. **推动口径统一。** 当前“采用率”“落地”“规模化”“效率提升”等关键概念在不同报告中定义各异，严重影响跨报告比较与政策评估。行业组织可牵头制定统一的统计口径。
2. **建立失败案例的收集机制。** 现有公开材料存在严重的幸存者偏差。在保护企业隐私的前提下匿名收集失败经验，其公共价值远高于再增加一批成功案例。
3. **关注跨主体数据协同机制。** 供应链、医疗、物流等领域 AI 的下一阶段价值，瓶颈在数据权属与共享规则，这是单个企业无法解决的公共品问题。
4. **重视中小企业的普惠路径。** 公开案例显示中小制造企业以 5 万—30 万元投入即可获得 2—6 个月的回本周期，但前提是有低门槛的标准化产品供给。降低这一门槛的公共基础设施建设具有较高的社会回报。

附录 A 参考信息源清单

本附录按信息源分层列出本白皮书编制过程中使用与查证的公开资料。列出的链接为编制时可访问的公开地址，如遇失效，建议通过发布机构官方网站检索文件名称。

A.1 A 级：政策文件与官方发布

序号	文件名称	发布机构	时间
A1	《关于深入实施“人工智能+”行动的意见》	国务院	2025.08
A2	《关于推进“人工智能+”能源高质量发展的实施意见》 (国能发科技〔2025〕73号)	国家发展改革委、国家能源局	2025.09
A3	《关于促进和规范“人工智能+医疗卫生”应用发展的实施意见》	国家卫生健康委等五部门	2025.11
A4	《“人工智能+制造”专项行动实施意见》 (含五大行业转型指引与企业应用指南)	工业和信息化部等八部门	2026.01
A5	《“人工智能+教育”行动计划》(教科信〔2026〕1号)	教育部等五部门	2026.04
A6	关于征集“人工智能+农业”典型应用场景的通知	农业农村部市场与信息化司	2026.04
A7	《中国智慧教育发展报告（2025—2026）》	教育部（世界数字教育大会发布）	2026.05
A8	《关于加快“人工智能+消费”发展的实施意见》 (商建发〔2026〕89号)	商务部等八部门	2026.06
A9	《卫生健康行业人工智能应用场景参考指引》 (列出84个典型应用场景)	国家卫生健康委等部门	2024.11

A.2 B 级：研究机构与咨询机构报告

序号	报告名称	发布机构	本白皮书引用的要点
B1	《AI Index Report 2026》	斯坦福大学以人为本人工智能研究院（HAI）	企业 AI 采用率约 88%； 生产力收益集中于领先梯队
B2	《State of AI trust in 2026》 (2026.03)	麦肯锡公司	近 2/3 受访者将安全风险列为扩展智能体的首要障碍；74%认为模型不准确为高相关风险；负责任 AI 成熟度均分 2.3，责任清晰组织 2.6 vs 责任不清组织 1.8
B3	《人工智能产业发展研究报告	中国信息通信研究院	2024 年 AI 核心产业规模突破 9000 亿元、

序号	报告名称	发布机构	本白皮书引用的要点
	（2025 年）》		同比增长约 24%；智能原生应用涌现；安全治理单列成章
B4	《政务智能体发展研究报告（2026 年）》	北京大学政府管理学院、中国信息通信研究院	四类政务场景划分；"高频服务先行"梯次路径；三维评估框架；30 项典型经验
B5	《"人工智能+消费"发展研究报告（2026 年）》	中国信通院产业与规划研究所、中国信息消费推进联盟	消费提效期—消费重构期—消费共创期三阶段演进路径
B6	《AI Radar》系列调研	波士顿咨询公司（BCG）	约 25%高管认为已创造显著价值；领先者聚焦 3.5 个用例 vs 其他 6.1 个，ROI 预期约 2.1 倍；"10—20—70"法则
B7	关于任务型 AI 智能体渗透的预测（2025.08）	Gartner	企业应用集成任务型智能体占比 2025 年<5%→2026 年底约 40%；2029 年至少半数知识工作者掌握相关技能
B8	关于生成式 AI 试点成效的研究报告（2025）	麻省理工学院相关研究团队	约 95%的生成式 AI 试点未产生可衡量的损益表影响
B9	《2026 年中国银行业大模型应用跟踪报告》	沙丘智库	银行大模型采用率 39.0%→53.5%；智能体采用率 25.0%→32.3%；无采用计划者仅 2.4%
B10	《2026 中国制造业 AI 场景应用白皮书》（167 个验证案例）	行业专家与企业联合编撰	生产制造 62.3%、研发设计 21.6%、供应链 5.4%；传统机器学习 46.7%、机器视觉 20.4%、AI 智能体约 3%
B11	《中国零售行业 AI 应用产业图谱 2026》	易观分析	AI 从效率工具升级为战略核心；重构人货场与供应链全链路
B12	《2026 中国企业 AI 应用场景报告》（近千份实践案例）	极客邦科技	覆盖金融、零售、能源、制造；落地五大共性：场景适配、智能体四层认知闭环、工程化体系、数据融合、合规安全
B13	《AI 驱动政务热线发展研究报告（2025）》	清华大学数字政府与治理研究院	政务热线 AI 应用的技术路径与治理要求
B14	《2026 数字人电商直播白皮书》	行业机构发布	数字人直播成为电商零售标配能力；AI 数字人市场 2024 年增速 85.3%；布局企业超 114 万家

A.3 C 级：企业公开案例与媒体报道

以下为本白皮书场景卡片中引用的具体案例来源。再次提示：这些数据由披露方自行定义口径，未经独立审计，存在选择性披露与幸存者偏差，不应作为本企业的预算依据。

编号	案例主体与场景	来源与时间	引用的关键数据
C1	AI 内容服务商 (服务安踏、施耐德电气、 比音勒芬、水星家纺、朗姿 等) ——电商内容生产	新浪科技 (2025.04)	详情页撰写 300—500 元/条； 拍摄约 500 元/张；模特费约 2000 元/次； 两天拍摄约 1.6 万元；AI 后百元内生成 数十款；1—2 人替代 30—40 人团队； 成本降幅约 90%
C2	多点 (Dmall) ——智能补 货	腾讯新闻 (2024.04)	缺货率最高 7%→2%以内； 库存周转最高 35 天→普遍 21 天以内； 近 30 人管理近千家门店； 年节约人力成本约 5000 万元
C3	中国光大银行 ——授信调查智能报告	新华网客户端	百页授信报告约 1 周→自动生成 不到 5 分钟；服务客户经理近 2000 名； 覆盖 29 家一级分行
C4	江阴银行——信贷资料智能 核验	新华日报 交汇点 (2026.08)	资料整理时间缩短 70%以上； 进件小时级→分钟级；人均日处理量 提升约 60%；返工率降 80%； 后台减负超 65%；识别准确率超 99%； 漏采错采问题减少 95%； 单户运营成本降低一半以上
C5	上海市第一人民医院 ——病历生成	澎湃新闻	病历书写 5—10 分钟/份→15—20 秒/份； 日均从 8 小时→25 分钟； 意图识别与诊断抽取准确率 90%以上
C6	北京世纪坛医院 ——AI 辅助病历生成	HIT180 医疗信息化	平均 44 秒生成完整病历； 准确率超 90%；看诊时间 6.97 分钟 →约 4 分钟 (降约 40%) ； 覆盖 6 个专科推进至 9 个； 累计生成预问诊病历 6.6 万余份
C7	施耐德电气——AI 视觉质检	InfoQ	漏检率降至 0；过检率稳定 0.5%以内； 人员从"主力"转"候补"； 部署于 10 个工厂、20 条生产线
C8	辽宁省 12345 热线 ——政务大模型	中国日报网 (2026.06)	智能派单准确率 90%以上； 处理效率提升约 25%； 响应时间缩短约 30%； "小宁智能问答"5.7 万次； 数据规模 4 亿余条诉求
C9	法律科技服务商 ——AI 合同审查	新浪科技 (2025.03)	简单合同 30—120 分钟→10—20 分钟； 复杂合同 4—8 小时→20—30 分钟； 案例检索 5—20 分钟→5—20 秒； 条款遗漏风险 20%→2%； 某集团法务 10 人/月→2 人/月； 某金融公司外部律师费 100 万→16 万

编号	案例主体与场景	来源与时间	引用的关键数据
C10	中国电信——AI 智能节能	新浪科技 (2024.10)	年均节电约 8 亿度； 年均节约电费约 5.2 亿元； 覆盖基站、IT 云主机、AB 类机房
C11	货拉拉、快递 100、京东物流、 菜鸟——物流 AI	中新网 (2025.03)	货拉拉准点率 95%、配对率 92%、 大模型问答准确率 90%以上； 快递 100 使乡镇网点人力成本降约 60%； AR 识货体积测量效率提高十倍以上
C12	捷豹路虎、通用汽车、日产——汽车研发 AI	信息化观察网 (2026.04)	捷豹路虎空气动力学计算 4 小时→1 分钟； 通用草图转 3D 从数月→数小时； 日产开发周期目标 60 个月→30 个月
C13	上汽享道、九识智能、吉利、 蔚来、理想、比亚迪、京东——2026 WAIC 汽车智能化	新浪财经 (2026.07)	组合驾驶辅助渗透率约 70%、 NOA 突破 30%；享道 Robotaxi 订单 35 万余次、 行驶超 400 万公里；九识部署周期 一月→一天；吉利星睿智算中心 2.0 算力 23.5EFLOPS；蔚来 NX9031X 出货超 30 万颗； 理想马赫 M100 算力 1280TOPS； 后装面向 3.66 亿辆存量车
C14	AI 医疗规模化落地	新华网 (2026.03)	约 80%县（市、区）建成县域影像等 共享中心；2025 年县域远程影像诊断 超 6800 万人次；肺结节筛查减少医生 30%—50%工作量；影像诊断效率 整体提升约 30%；主动脉夹层判读 15—20 分钟→约 3 分钟
C15	制造业企业 AI 应用 （通威、长虹、海尔、南 钢、 汉得、中小企业）	《2026 中国制 造业 AI 场景应用白皮书》 等公开资料转引	通威降本约 37%；长虹 AI 质检年省千万级、 回本 4—6 个月；海尔年增效超 1.1 亿元、 库存周转提升约 60%；南钢峰谷发电 效益提升约 4.24 倍；汉得产线智能体 单产线年收益 200 万—500 万元； 中小企业投入 5 万—30 万元、回本 2—6 个月
C16	AI 车险定损理赔	南方 plus 等 公开报道	平均结案时间约 2.16 分钟（简易小额案件）； 行业另有"理赔周期 3 天→2 小时" 的公开表述（常规案件）
C17	三大运营商 AI 业务	新华网、新浪财经 (2026.03— 04)	中国联通 AI 收入同比增长超 140%； 中国移动数据算法收入同比增长 12.6%、 推出上百款 AI 产品与数十款垂类智能体

附录 B 场景速查总表

本表汇总本白皮书全部 82 个场景，按行业排列，可用于快速定位与横向比较。"改善幅度"列为公开案例披露值或本研究院基于案例的归纳，仅供参考。

表 B-1 全场景速查表（一）：六大深度行业与获客职能

场景编号	场景名称	成熟度	价值层次	公开披露的改善幅度
3.1.1	商品图与模特图生成	L4	替代型	成本降幅约 90%；人力约 1/20
3.1.2	数字人直播	L3	替代/创造型	开播时长翻倍以上；边际成本趋零
3.1.3	智能补货与销量预测	L2—L3	重构型	缺货率 7%→2%；周转 35 天→21 天
3.1.4	智能客服与售后工单	L3—L4	替代→重构型	响应秒级；质检从抽检到全检
3.1.5	跨境本地化与多语言	L3	替代型	新市场上线周期由月→天
3.1.6	AI 导购与对话式购物	L2—L3	重构型	尚无稳定可比数据
3.1.7	投放素材与智能出价	L3	增强型	素材产能数量级提升
3.1.8	门店客流与陈列识别	L3	增强型	陈列检查从抽检到全检
3.2.1	表面缺陷视觉质检	L3—L4	替代型	漏检率→0；过检率≤0.5%；回本 4—6 月
3.2.2	设备预测性维护	L3	替代/增强型	非计划停机减少（幅度因企业而异）
3.2.3	工艺参数寻优	L2—L3	增强型	公开案例降本约 37%
3.2.4	能耗与峰谷用电优化	L2—L3	增强型	公开案例效益提升约 4.24 倍
3.2.5	工业知识问答与检索	L2—L3	增强型	检索时间大幅压缩
3.2.6	生产排程与产线平衡	L2	重构型	排程编制由小时/天→分钟
3.2.7	安全生产行为识别	L3	替代型	覆盖率从抽检到全时段
3.3.1	授信调查报告生成	L2—L3	替代型	百页报告约 1 周→<5 分钟
3.3.2	信贷资料核验与进件	L2—L3	重构型	进件小时级→分钟级；返工率降 80%
3.3.3	编码辅助与测试生成	L3—L4	替代型	编码环节提效（≠交付周期）
3.3.4	智能运维与根因定位	L3	替代→重构型	告警噪声大幅降低
3.3.5	反欺诈与异常交易识别	L4	替代型	从事后到事中拦截
3.3.6	研报摘要与投研聚合	L3	增强型	信息覆盖面扩大
3.3.7	智能客服与远程银行	L3—L4	替代型	全时段覆盖；全量质检

场景编号	场景名称	成熟度	价值层次	公开披露的改善幅度
3.4.1	门诊病历与预问诊生成	L2—L3	替代型	5—10 分钟→15—20 秒；看诊时长降 40%
3.4.2	医学影像辅助检出	L3—L4	增强型	减少医生 30%—50%工作量
3.4.3	智能导诊与预问诊	L3	重构型	分诊前移至院外
3.4.4	病案首页编码与质控	L3	替代型	质控从抽检到全检
3.4.5	慢病管理与智能随访	L2—L3	重构型	随访覆盖率数量级提升
3.4.6	药物研发辅助	L2	创造型	早期环节加速（≠总周期）
3.5.1	备课与教案、试卷生成	L3	替代型	备课准备时间大幅压缩
3.5.2	主观题批改与作文评价	L2—L3	替代型	批改时长大幅压缩
3.5.3	口语陪练与语言学习	L3—L4	替代/创造型	边际成本趋零；开口量数量级提升
3.5.4	学情分析与教学建议	L2	增强型	定位颗粒度提升至知识点
3.5.5	企业培训与课程生成	L2—L3	替代型	课程开发周期由月→天
3.6.1	政务热线智能受理与派单	L3	替代→重构型	派单准确率>90%；效率+25%；响应-30%
3.6.2	公文起草与校核辅助	L2—L3	替代型	成稿时间压缩；格式校核自动化
3.6.3	办事材料预审与导办	L2—L3	重构型	补正往返次数下降
3.6.4	基层网格事件识别派单	L2—L3	重构型	发现时效大幅提升
3.6.5	政策问答与文件检索	L2—L3	增强型	从"人找政策"到"政策找人"
3.7.1	选题挖掘与对标账号拆解	L3	增强型	分析样本量数量级提升
3.7.2	图文笔记与短视频脚本生成	L3—L4	替代型	100 条短视频 1500 元→300 元（降 80%）
3.7.3	创始人直播与数字分身	L3	替代/创造型	制作周期由小时/天→分钟
3.7.4	账号矩阵与多平台分发	L3	替代型	矩阵渠道流量占比超 60%
3.7.5	评论区与私信即时应答	L2—L3	重构型	留资率+38%；70%成交在私信前 3 分钟
3.7.6	直播获客与切片分发	L2—L3	增强/创造型	单场直播切片产出数量级提升
3.7.7	GEO：生成式引擎优化	L1—L2	创造型	超 41%用户几乎完全转向 AI 搜索
3.7.8	线索培育与私域跟进	L2	重构型	长尾线索生命周期延长

表 B-2 全场景速查表（二）：十二个扫描行业

场景编号	场景名称	成熟度	价值层次	公开披露的改善幅度
4.1.1	智能调度与订单分配	L3	增强型	准点率 95%；配对率 92%
4.1.2	单据识别与自动对账	L3	替代型	网点人力成本降约 60%
4.1.3	仓内机器人与智能分拣	L3—L4	替代型	行走时间消除；峰值弹性改善
4.2.1	智能节能与能耗优化	L3	增强型	年节约电约 8 亿度；省电费约 5.2 亿元
4.2.2	负荷与新能源功率预测	L3	增强型	备用容量需求下降
4.2.3	输变电设备智能巡检	L3	替代型	高危作业大幅减少
4.3.1	整车研发仿真加速	L2—L3	增强型	空气动力学计算 4 小时→1 分钟
4.3.2	智能座舱与车载智能体	L3	创造型	交互范式改变；存量车可后装
4.3.3	Robotaxi 与限定场景无人驾驶	L2—L3	重构型	部署周期一月→一天
4.4.1	网络智能运维与故障自愈	L3	替代→重构型	告警噪声大幅降低
4.4.2	基站与机房智能节能	L3	增强型	年省电费约 5.2 亿元
4.4.3	号码与业务风控反诈	L3	增强型	从事后封停到事中拦截
4.5.1	合同审查与风险提示	L2—L3	增强型	复杂合同 4—8 小时→20—30 分钟
4.5.2	类案检索与法律研究	L2—L3	增强型	案例检索 5—20 分钟→5—20 秒
4.5.3	尽职调查文档批量处理	L2—L3	替代型	通读环节大幅压缩
4.6.1	简历解析与初筛	L3	替代型	批量处理时间大幅压缩
4.6.2	HR 事务智能体与员工问答	L3	替代型	咨询工作量下降；口径统一
4.6.3	面试与情景演练	L2	替代/创造型	约面成本消除；演练频次提升
4.7.1	车险智能定损与理赔	L3	替代→重构型	简易案件平均结案约 2.16 分钟
4.7.2	智能核保	L2—L3	替代型	标准件秒级出结果
4.7.3	代理人展业辅助	L2	增强型	方案制作时间压缩
4.8.1	病虫害识别与精准施药	L2—L3	替代/增强型	农技服务覆盖面数量级提升
4.8.2	智能农机与无人作业	L2—L3	替代型	作业窗口扩展至夜间
4.8.3	农技问答与种植决策	L1—L2	创造型	咨询可及性从无到有
4.9.1	工地安全行为识别	L3	替代型	覆盖率从抽检到全时段
4.9.2	图纸审查与工程量核算	L2	替代/增强型	算量效率提升（依赖 BIM 程度）
4.9.3	楼宇能耗与设备运维	L2—L3	增强型	可直接换算电费节省

场景编号	场景名称	成熟度	价值层次	公开披露的改善幅度
4.10.1	图像与视频内容生成	L3—L4	替代/创造型	前期与中间稿成本大幅下降
4.10.2	配音、翻译与本地化	L3—L4	替代型	出海成本下降幅度最大的环节
4.10.3	游戏 NPC 与动态内容	L2—L3	创造型	体验形态改变；受推理成本制约
4.11.1	备货预测与损耗控制	L2—L3	重构型	参照 3.1.3；生鲜品类收益最大
4.11.2	评价分析与服务质量管理	L3	增强型	从抽样到全量分析
4.11.3	智能排班与人效优化	L2	增强型	排班编制事务负担下降
4.12.1	文献综述与知识发现	L2—L3	增强型	文献筛选效率提升
4.12.2	科学计算代理模型加速	L2	创造型	单次仿真数量级加速
4.12.3	咨询审计文档密集型工作	L2—L3	替代型	从抽样检查到全量扫描

附录 C 场景自评量表

本量表基于第二章 SVA-3D 模型编制，供组织在 AI 场景立项前使用。使用方式：对每个候选场景逐项打分，汇总后按总分排序，优先推进得分最高的场景。

表 C-1 场景价值自评量表（满分 27 分）

维度	评估问题	3 分	2 分	1 分
数据 可及性	所需数据是否已在系统中沉淀？	已结构化沉淀，历史数据充足	存在但分散或为非结构化	尚未系统采集
数据 可及性	获取数据是否需要跨主体协调？	完全在本单位可控范围内	需跨部门协调	需跨法人主体协商
数据 可及性	标注成本是否可接受？	无需标注或可自动标注	需一定人工标注	需大量专家标注
容错 空间	单次错误的后果严重程度？	仅造成返工，无外部影响	影响客户体验或经济损失	涉及人身安全、资金或法律后果
容错 空间	错误是否易被下游发现？	立即可见	需抽检发现	可能长期不被发现
容错 空间	是否存在成熟的复核机制？	已有复核环节可直接沿用	需新建复核机制	难以设置有效复核
价值 可计量 性	是否存在明确的基线指标？	有现成指标且持续统计	指标存在但未系统统计	无可用基线
价值 可计量 性	改善是否可归因于 AI？	可直接归因	需设计对照或分阶段观测	与其他因素高度耦合
价值 可计量 性	收益是否有明确的承接部门？	已有部门愿意纳入考核	需协调确定	无部门认领

C.1 结果解读

总分区间	建议动作
24—27 分	首批必做。 三个维度均处于有利位置，应在 3—6 个月内完成上线，作为组织内部的示范项目。
19—23 分	重点推进，需配套建设。 通常是某一维度存在短板，应先针对该短板设计专项措施（数据治理、复核机制或指标体系），再启动主体建设。
14—18 分	拆解后试点。 不宜整体推进，应从中拆出得分较高的子环节先行验证。
13 分及以下	暂缓。 当前条件下投入产出比不成立，建议列入观察清单，待外部条件（数据、法规、技术）变化后重新评估。

C.2 使用提示

- 本量表的目的不是产生一个精确的分数，而是**迫使团队在立项前把三个维度的问题都问一遍**。九个问题中若有任何一个答不上来，本身就是重要的信号。
- 容错空间维度的三项应由业务与合规部门评分，不宜由技术部门单独判断。
- 对于总分中等但战略意义重大的场景，正确做法不是提高分数，而是**明确记录短板并为其单独立项**（例如把数据治理拆为独立项目）。

附录 D ROI 测算表模板

本模板基于第六章的提效账本模型编制，可直接复制到表格软件中使用。

D.1 第一步：基线记录表

在项目启动前用 1—2 个月记录真实基线。这一步不可跳过——没有基线，后续所有计算都无法验证。

记录项	记录方式	本单位实际值
业务量	该场景每月处理的业务单量	_____
单件处理时长	从开始到完成的平均耗时	_____
投入人力	专职+兼职折算的人力数	_____
差错率	返工、投诉、退货等的比例	_____
相关外采支出	外包、外购服务的年度金额	_____
指标历史波动	过去 12 个月该指标的最大最小值	_____

D.2 第二步：收益测算表

收益类型	计算公式	能否入账	预估金额
外采成本节省	$(\text{原年度外采金额}) \times (\text{可替代比例})$	能，优先列示	_____
产能扩张收益	$(\text{新增处理业务量}) \times (\text{单位业务毛利})$	**能，需证明 确已转化**	_____
人力时间释放	$(\text{节省工时}) \times (\text{人均工时成本})$	**通常不能， 仅作参考**	_____
风险成本降低	$(\text{历史年均损失}) \times (\text{差错率下降幅度})$	**间接， 需长周期观测**	_____
收益合计	前两项为“可入账收益”，后两项为“参考收益”	—	_____

D.3 第三步：投入测算表

成本类别	一次性投入	年度持续投入	备注
软件与模型	_____	_____	注意推理成本随使用量增长
算力与基础设施	_____	_____	—
系统集成	_____	_____	—
数据准备	_____	_____	易低估。含采集、清洗、标注、 主数据治理

成本类别	一次性投入	年度持续投入	备注
硬件与工程改造	_____	_____	易低估。传感器、摄像头点位、光源、网络
持续运维	—	_____	易低估。模型迭代、知识库更新、换型重训
隐性成本	_____	_____	业务专家时间占用、审校新增工时、培训与变更管理
投入合计	_____	_____	后四项合计应≥总预算的 50%，否则估算偏乐观

D.4 第四步：三情景测算

情景	假设	可入账收益	投入	回收期
乐观	达到公开案例披露值的 70%	_____	_____	_____
基准（用于决策）	达到公开案例披露值的 40%	_____	_____	_____
悲观（用于风控）	达到公开案例披露值的 20%	_____	_____	_____

说明：折扣系数的设定依据是幸存者偏差——公开案例均为成功案例，且披露口径通常有利于披露方。本研究院建议以基准情景（40%）作为决策依据。若基准情景下回收期超过 24 个月，建议重新评估该项目。

D.5 第五步：复盘节点约定

节点	检查内容	停止投入的触发条件
第 3 个月	前提条件是否已具备（数据、标准、流程）	数据准备进度低于计划 50%
第 6 个月	试点场景的核心指标是否有可统计的改善	改善幅度小于指标历史波动区间
第 12 个月	是否形成标准化交付物、能否复制	仍需原班人马才能运行，无法复制

附录 E 术语表

术语	本白皮书中的含义
SVA-3D 模型	本研究院提出的场景价值三维评估模型，从数据可及性、容错空间、价值可计量性三个维度判断 AI 场景的优先级。详见第二章 2.1—2.2 节。
成熟度阶梯（L0—L4）	本研究院提出的场景成熟度标度：L0 概念验证、L1 单点试点、L2 单位内推广、L3 跨组织复制、L4 行业基础设施。详见第二章 2.2 节。
价值四层	本研究院对 AI 价值的分类：第一层替代型、第二层增强型、第三层重构型、第四层创造型。详见第二章 2.3 节。
提效账本模型	本研究院提出的收益归类工具，把 AI 带来的节省分为外采成本节省、产能扩张、人力时间释放、风险成本降低四类，其中前两类可直接入账。详见第二章 2.4 节与第六章。
任务型智能体	能够被事件触发、自主规划并执行多步骤操作以完成完整业务任务的 AI 系统，区别于需要人工逐次调用的对话式助手。
检索增强生成（RAG）	在生成回答前先从指定知识库中检索相关材料，据以生成并标注来源的技术路线，常用于对准确性要求较高的专业场景。
自动闭环率	在某一业务流程中，由 AI 系统独立完成全流程且无需人工介入的业务量占比，是衡量智能体成效的核心指标。
人工干预率	需要人工介入修正或接管的业务量占比，与自动闭环率互补。
自动化偏见	人在与自动化系统协作时，过度信任系统输出而降低自身独立判断强度的现象，在医疗、法律、审计等高风险决策场景中尤需警惕。
兜底率	本白皮书针对公共服务场景提出的指标，指智能体无法处理时成功转接人工并最终办结的比例。详见场景 3.6.1。
过检率	质检场景中把合格品误判为不合格品的比例（误报率）。过检率过高会给一线增加复核负担，是质检 AI 被弃用的主要原因。
幸存者偏差	因失败案例极少被公开，导致基于公开案例得出的结论系统性偏乐观的统计现象。本白皮书为此设置了“常见坑”栏目、第七章失败模式与附录 D 的折扣系数。
成本转移	本白皮书归纳的失败模式之一：一个环节的成本下降，但转移到了另一个环节（如生成成本降低但审校成本上升），总账未改善。

关于求智启新人工智能研究院

我们是谁

求智启新人工智能研究院（QIUZHI QIXIN AI RESEARCH INSTITUTE）总部位于杭州浙江大学校友企业总部经济园，由浙江大学管理学院实践教授郭峻峰与企业 AI 落地专家陈灵军联合创办。

研究院汇聚高校教授、企业管理者、产品与项目专家、财务与人力资源顾问、供应链顾问、AI 绘画设计专家、医学教育专家、程序员提效与企业系统改造顾问等多领域实战力量，面向企业家、高管团队、业务部门、职能岗位与产业平台，提供从 AI 认知培训、岗位提效工作坊，到 Agent 数字员工、知识库、GEO 营销增长与企业 AI 转型陪跑的系统化赋能。

截至本白皮书编制时，研究院已服务 50 余家企业，交付 200 余场培训，累计授课学员 3.1 万余人次。研究院编著的《AI 智能体实操指南》由电子工业出版社出版，发行 1.2 万册，并被多所高校采购为教材。

我们的服务体系：四阶段递进

研究院的服务设计与本白皮书第二章的分析框架一脉相承——企业的 AI 能力建设不是一次采购，而是一条从“看得懂”到“用得上”再到“算得清”的路径。

表 F-1 求智启新人工智能研究院服务体系

阶段	服务产品	周期	解决的问题	对应本白皮书
第一阶段 认知扫盲	企业 AI 战略 领导力专题课	2—3 小时	决策层建立统一的 AI 认知与判断标准，避免被演示效果误导	第一章、第二章
第二阶段 单点提效	AI 效能革命 工作坊	1—2 天	选定高价值场景， 让业务团队真正上手用起来	第三章、第四章 场景卡片
第三阶段 整体转型	AI 智能化升级 陪跑营	1—3 个月	从单点走向流程重构， 建立可复制的交付与运维体系	第五章七条法则、 第六章 ROI 测算
第四阶段 长期增长	GEO 生成式 搜索优化	持续	在 AI 搜索时代建立 品牌的被引用能力与获客增量	场景 3.7.7

专项定制产品

- 企业专属 AI 系统与 Agent 定制——面向具体业务流程的智能体设计与交付
- 知识库搭建——把分散在文档与个人经验中的组织知识结构化沉淀
- 零代码工作流——让业务人员自主搭建自动化流程，降低对技术资源的依赖

重点赋能行业与交付案例

研究院重点服务的领域包括大型国企与政府、传统制造与供应链、电商新零售、出海跨境贸易。以下为部分已交付案例，其场景与本白皮书的卡片编号相对应：

表 F-2 部分交付案例

案例方向	交付成效	对应场景卡片
大健康行业知识库	整理周期由约 30 人天压缩至 1 天，沉淀 1500 余条问答	场景 3.2.5 工业知识问答（同类方法）
AI 驱动选品系统	选品周期由 2—3 周压缩至半天	场景 3.1.6 导购与选品辅助
AI 设计赋能	设计效率提升 100%—300%	场景 3.1.1 商品图与模特图生成
自媒体获客闭环	内容生产至线索承接的全链路搭建	第 3.7 节 自媒体获客与创始人 IP
消费者模拟	用 AI 模拟目标客群反馈，前置验证产品与内容方案	场景 3.1.6、3.7.1
法律卷宗工具	文档密集型工作的批量处理	场景 4.5.3 尽职调查文档处理

说明：上述为研究院自有交付案例，成效数据由项目方确认，口径与本白皮书对第三方公开案例的处理原则一致——即由披露方定义口径、未经独立审计，不宜直接外推至其他企业。

联系我们

项目	信息
官方网站	qzqxai.com (可查阅专家团队、服务体系、交付案例与代表客户的完整信息)
地址	杭州市浙江大学校友企业总部经济园 B2 幢 606
合作负责人	陈灵军
电话 / 微信	13148398839
合作方向	企业内训·岗位提效工作坊·AI 转型陪跑营·Agent 与知识库定制·GEO 营销增长

关于本白皮书

本白皮书为研究院行业应用系列研究的组成部分，其素材既来自公开信息的独立分析，也来自研究院在 50 余家企业中的一线交付观察——本书中大量的“落地前提”与“常见坑”，正是从后者中总结出来的。

我们欢迎来自产业一线的批评、补充与案例贡献——**尤其欢迎失败经验的分享**。如第七章所述，公开渠道系统性地偏向成功案例，而失败经验对行业的价值往往更高。相关反馈可通过官网 **qzqxai.com** 或上述联系方式与我们取得联系。

编著：求智启新人工智能研究院

版本：2026年8月 第一版

本白皮书所载信息截至2026年8月，仅供研究参考，不构成任何投资、法律或技术选型建议。文中引用的企业案例数据由披露方自行定义口径、未经独立审计，不应作为本企业的预算依据。