

The logo for ifenxi, featuring the brand name in a bold, blue, sans-serif font. The 'i' has a small white dot, and the 'x' has a white dot. The logo is positioned in the top left corner of the page, which has a white background for the logo itself, set against a larger blue background with a futuristic, glowing digital circuit pattern.

# 应用为王：中国大模型市场 2025-2027价值重构与决胜之路

2025爱分析·大模型应用实践报告

11. 2025

## 报告编委

### 报告指导人

张扬                      爱分析                      联合创始人&首席分析师

### 报告执笔人

李进宝                      爱分析                      高级分析师

### 外部专家（按姓氏拼音排序）

岑润哲                      数势科技                      数智产品总经理

黄文婷                      拓尔思                      成都拓尔思副总经理

## 特别鸣谢（按拼音排序）



# 目录

1. 报告综述	6
2. 市场洞察	15
3. 对话式智能分析市场	23
4. “大模型+政务”市场	37
5. 结语	50
关于爱分析	51
产品服务	52
法律声明	53

CHAPTER

01

# 报告综述

# 1. 报告综述

近年来，全球人工智能产业的焦点已从早期的算法创新和参数规模竞赛，战略性地向**应用深耕与商业化落地**倾斜。中国大模型市场热潮的持续升温，反映了其发展已进入一个由宏观政策主导、内生技术突破支撑、最终由市场价值牵引的成熟阶段。

本报告采用三维分析框架，即**政策主导（确定方向）、技术供给（提供可能）、需求牵引（实现价值）**，对驱动中国大模型市场发展的核心力量进行结构化、前瞻性的深度剖析。只有理解这三股力量的相互作用及其内部结构性变化，才能准确把握未来市场的脉络和战略机遇。

## 1. 国家战略导向与政策体系的系统构建

政策力量是中国大模型市场爆发的最强外部推力。它不仅提供了资金和资源支持，更重要的是通过国家意志确定了行业发展的方向、速度和应用深度，尤其对具有战略地位的央国企和关键领域产生了巨大的“刚性”需求。

### 1.1 顶层设计：从“鼓励创新”到“战略强制”的转变

中国对人工智能的定位已超越单纯的新兴技术领域，被提升至国家战略和基础生产力的高度。在最新的宏观规划中，这一转变得到了明确的确认。“十五五”规划建议已将“全面实施‘人工智能+’行动”定为国家核心战略基调。这意味着人工智能不再是可选项，而是作为一种基础生产力工具和科研范式变革的引领者被定位。规划建议明确指出，要“以人工智能引领科研范式变革”，并“全方位赋能千行百业”。

政策对央国企的直接影响是刚性需求的早期、最大来源。24年初至25年初，国资委推进“人工智能+”的会议，标志着自上而下的战略部署形成。央国企作为国家核心基础设施和关键数据资产的拥

有者，其 AI 化转型是政策驱动型刚性需求的起始和核心。这种顶层设计机制保证了国内大模型生态在商业化初期拥有一个稳定且高标准的市场基础。

## 1.2 AI 发展逻辑的根本性转变与战略优先级

政策的关注点已经从单一的经济产出指标，扩展到了更广泛的社会效益和国家韧性层面。政策文件明确要求加强人工智能与“产业发展、文化建设、民生保障、社会治理”相结合。同时，国务院关于深入实施“人工智能+”行动的意见要求，加快实施“人工智能+”消费提质、科学技术、治理能力三大行动。

这表明 AI 的价值评估体系已经超越了单纯的 GDP 贡献，提升至社会效益和国家治理能力的层面。政府及央国企作为大模型初期最大、最稳定的用户群，通过刚性需求保证了国内大模型生态的基本盘。此外，政策的体系化也体现在风险管理方面。在推动技术加速发展的同时，政策采取了“同步治理”的战略，明确要求“加强人工智能治理，完善相关法律法规、政策制度、应用规范、伦理准则”。这种双轨制策略，特别提到持续推动“智能向善”，构建中国特色人工智能治理框架，旨在保证技术的安全、可控与合规性。对于央国企、金融、医疗等高敏感度行业的应用落地而言，这种合规性基础是推动应用深度普及的必要前提。

## 1.3 政策驱动的广度与深度解析

### 1.3.1 广度拓展：从技术领域到全社会覆盖

未来，政策覆盖面将显著扩大。首先在消费端，加快实施“人工智能+”消费提质行动，旨在拓展服务消费新场景，培育产品消费新业态，直接拉动了 C 端对高性能、高智能服务的需求。其次，政策深入到科研环节，要求“以人工智能引领科研范式变革”。这不仅是应用层面的赋能，更是对基础研究方法的颠覆性投入，预示着 AI 将在科研范式变革中扮演核心角色。

### 1.3.2 深度加码：投入力度与人才生态建设

政策的投入力度已从集中在基础设施（如算力）向更广泛的生态环节延伸，包括人才、标准和安全。

为了解决制约 AI 产业发展的核心瓶颈——高素质人才供给不足，工业和信息化部等机构通过征集“人工智能与数字安全产业人才基地联合建设机构”，旨在推动产业链、人才链、创新链、教育链的协同发展。这种结构性的人才供应链重塑，为行业的长期、高质量发展提供了坚实基础。

同时，政策落地直接催生了新的劳动价值链。例如，人工智能提示词设计员及基于人工智能进行内容创作与模式创新的各类新职业的出现，不仅代表了就业机会的增加，也说明 AI 赋能正在深刻重塑劳动力结构。

## 2. 基础技术突破与高质量供给的持续迭代

技术进步是驱动大模型热潮的核心内生动力。当前的技术供给正在经历一次系统性的升级，涵盖了从底层架构效率、工程化可靠性、部署形态优化到下一代智能形态的全面迭代。这种迭代确保了 AI 应用的商业可行性和泛化能力。

### 2.1 基础架构：效率、规模与范式创新

#### 2.1.1 MoE 架构的市场成熟

稀疏激活与动态选择机制的 MoE（Mixture of Experts）架构，已经成为构建千亿乃至万亿级大模型的主流技术路径。MoE 架构的引入有效地解决了超大规模模型在训练和推理时的计算资源线性增长瓶颈。通过允许模型仅激活部分专家子网络，MoE 极大地降低了超大规模模型在推理时的总拥有成本（TCO），使得“大模型、低成本”成为可能，这是推动大模型商业化落地的关键一环。

#### 2.1.2 非 Transformer 架构的挑战与潜力

在 Transformer 架构占据主导地位的背景下，以状态空间模型（SSM）为代表的新兴技术正在快速崛起。SSM 凭借在长序列处理、训练与推理效率上的显著优势，正成为未来大模型技术突破的关键方向。这种新架构的兴起预示着未来大模型技术将走向多架构并行的生态格局，能够与 Transformer 形成优势互补，为特定任务（如长文本理解、时序数据预测）提供更优的解决方案。

## 2.2 应用工程化路径：从“可用”到“可靠”

### 2.2.1 RAG 成为企业级应用首选

RAG（检索增强生成）结合向量数据库或知识图谱，已成为缓解大模型“幻觉”现象、提升信息准确性与可信赖度的主流解决方案。RAG 允许模型从企业内部或外部的权威知识库中检索实时、准确的信息，显著提升了输出内容的准确性和可解释性。RAG 的工程化成熟度高，正在快速取代高成本、维护难度大的微调技术（Fine-tuning），成为应用厂商构建智能系统、实现知识快速更新的首选路径。这种转变反映了市场已经进入了商业化实用阶段，技术供应商必须在模型性能、成本效益和数据安全之间找到最佳平衡点。RAG 逐渐取代微调，正是市场对“快速迭代”和“成本控制”的明确投票。

### 2.2.2 轻量化技术群推动 AI 能力端侧迁移

量化（Quantization）、蒸馏（Distillation）与剪枝（Pruning）作为大模型轻量化的三大核心支柱技术，共同推动了 AI 能力从高成本的云端向低延迟、高隐私性的端侧高效迁移。这对于满足智能手机、车载系统、物联网等边缘设备的需求至关重要。未来的持续挑战在于如何在压缩率、精度保持与硬件适配之间寻找更优平衡，以确保在资源受限的环境下仍能提供高性能的 AI 服务。

## 2.3 下一代智能的核心形态

### 2.3.1 Agent（智能体）的商业化双路径

Agent 作为大模型落地的核心形态，正沿着“流程引擎”与“自主规划”两条路径并行发展：

- 流程引擎型：强调低门槛、高稳定性，满足企业流程自动化需求，通过结构化的工具调用快速实现商业价值。
- 自主规划型：以高泛化能力探索通用智能的边界，通过复杂的任务拆解和执行，代表了 Agent 的长期发展潜力。

随着 MCP（多智能体协作协议）的普及与大模型能力的提升，Agent 将逐步从“工具型助手”演进为“自主型伙伴”，成为推动社会智能化构建的关键力量。此外，多智能体协作作为 AI 系统组织方式的重要演进方向，预示着 AI 将从“工具”向“组织”跃迁。

### 2.3.2 交互革命与世界模型

OS Agents：OS Agents 作为多模态大模型与操作系统深度融合的产物，正在开启“语义驱动交互”的新纪元。它利用强大的泛化能力和创新的交互潜力，有望成为下一代智能交互的核心形态。

端到端语音大模型：通过一体化建模实现了语音交互的质的飞跃，将人机沟通从“听得懂”推向“说得妙”，成为大模型时代最具情感温度的人机接口。

世界模型 (World Model)：世界模型代表了 AI 系统从“被动响应”向“主动理解”与“预测”的关键跃迁。它通过构建一个能够模拟、理解并预测复杂环境变化的内部系统，为强化学习、机器人控制、自动驾驶等需要高保真环境模拟的领域提供了基础，为构建更通用的人工智能系统奠定了基础。

## 2.4 生态基础设施的韧性与开放性

异构算力整合是构建自主可控 AI 基础设施的关键支撑。通过资源池化与智能调度，异构算力将多元计算资源“化零为整”，不仅提升了大模型的部署效率与系统韧性，也为国内大模型产业的规模化发

展提供了可靠基础。随着国产芯片生态的完善与调度算法的持续优化，异构算力整合将成为大模型产业不可或缺的基础设施能力。

### 3. 市场认知成熟与多维度需求的全面释放

市场需求是技术成果转化为商业价值的最终落脚点。政策驱动和技术成熟共同作用，加速了 G/B/C 三端对大模型的价值认同，并引发了从 IT 预算到战略投资的投入模式转变。

#### 3.1 认知成熟：价值认同驱动战略投入

市场教育的深化使得对大模型的认知不再停留在技术好奇或概念验证阶段。随着技术的可靠性 (RAG) 和效率 (MoE/轻量化) 的提升，市场已经完成了从“技术可行性”到“商业价值”的确认。投入模式随之发生了根本性转变：大模型建设投入从以往的 IT 支出或试点预算，升级为与企业核心业务流程深度捆绑的战略投资。这体现在企业愿意投入资源构建私有化部署、定制 RAG 知识库以及 Agent 系统。

#### 3.2 G/B 端需求的刚性释放与结构性驱动

##### 3.2.1 政策驱动型刚性需求 (政府和央国企)

国资委会议的推动是央国企端需求爆发的关键催化剂。央国企面临国家战略层面的数字化转型压力，必须将 AI 应用内嵌到核心生产和治理流程中。这种政策推进首先作用于央国企，创造了高价值、刚性的“第一批大订单”。

政策机制使得“政府作为第一个大客户”，保障了国内 AI 生态在初期避免陷入纯粹的低价竞争，而是集中资源解决最高标准下的落地可行性问题。这类需求特征是高安全要求 (私有化部署)、强制化 (行业垂直模型)、高合规性。这反过来驱动了国内大模型厂商必须在数据安全、垂直领域精度和私有化部署能力上快速成熟，为后续向中小企业市场扩张奠定了高质量基础。

### 3.2.2 效率驱动型需求 (普通企业)

对于更广泛的普通企业，大模型的价值主要体现在效率提升和成本优化上：

- 知识管理变革：RAG 技术直接满足了企业解决内部海量文档的检索、分析和利用效率问题，实现了企业知识资产的有效激活。
- 流程自动化升级：Agent 流程引擎的成熟推动了 RPA（机器人流程自动化）的代际升级，将自动化范畴从结构化任务扩展到半结构化和非结构化任务，极大地提升了白领工作效率。

## 3.3 C 端消费的场景创新与生态培育

### 3.3.1 AIGC 驱动内容消费

大模型在内容生成（图像、视频、文本）上的能力突破，直接驱动了新一轮的 C 端应用浪潮。特别是在短视频、游戏、设计等领域，AIGC 实现了创作效率和个性化体验的指数级提升。

### 3.3.2 新交互形态的拉动与技术反向推动

OS Agents 和端到端语音大模型通过语义驱动的深度交互，极大地优化了用户体验，使 AI 助手从“功能集合”演变为“主动伙伴”。这种用户对更流畅、更智能体验的追求，刺激了对高性能端侧 AI 硬件的更换需求。

值得注意的是，C 端需求不再是技术的被动接收者，而是通过对体验和成本的严格要求，反向引导了技术栈的优化和迭代方向。例如，用户对智能手机端侧 AI 的需求，直接反向推动了量化、蒸馏等轻量化技术的研发投入，以满足低功耗、高精度运行的要求。这构成了“技术供给”和“市场需求”之间更紧密的闭环反馈机制。

## 3.4 社会与人才的需求重塑

需求的全面爆发正在重塑劳动力市场结构。人工智能产业本身及其赋能的应用场景催生了新的就业机会，例如数据标注员、人工智能提示词设计员及基于人工智能进行内容创作与模式创新的各类新职业。这种人才结构性需求不仅是就业量的增加，更代表了社会对新型 AI 劳动力的迫切需要。此外，随着大模型应用的普及，社会对 AI 伦理、治理和责任创新的需求也同步提升。这种对“智能向善”的治理需求反过来驱动了合规性强的、内建安全机制的大模型产品市场中更受欢迎。

CHAPTER

02

# 市场洞察

## 2. 市场洞察

### 1. 价值迁移——应用层将成为增长的核心引擎

我们的分析基于 2024 年全年及 2025 年上半年的公开招投标数据,这些数据真实反映了市场的既成采购行为与资金流向。在此基础上,我们深度结合了对数十位覆盖模型、算力及应用领域的行业专家的调研访谈,对未来的技术趋势、商业化进程和市场接受度进行了前瞻性测算。

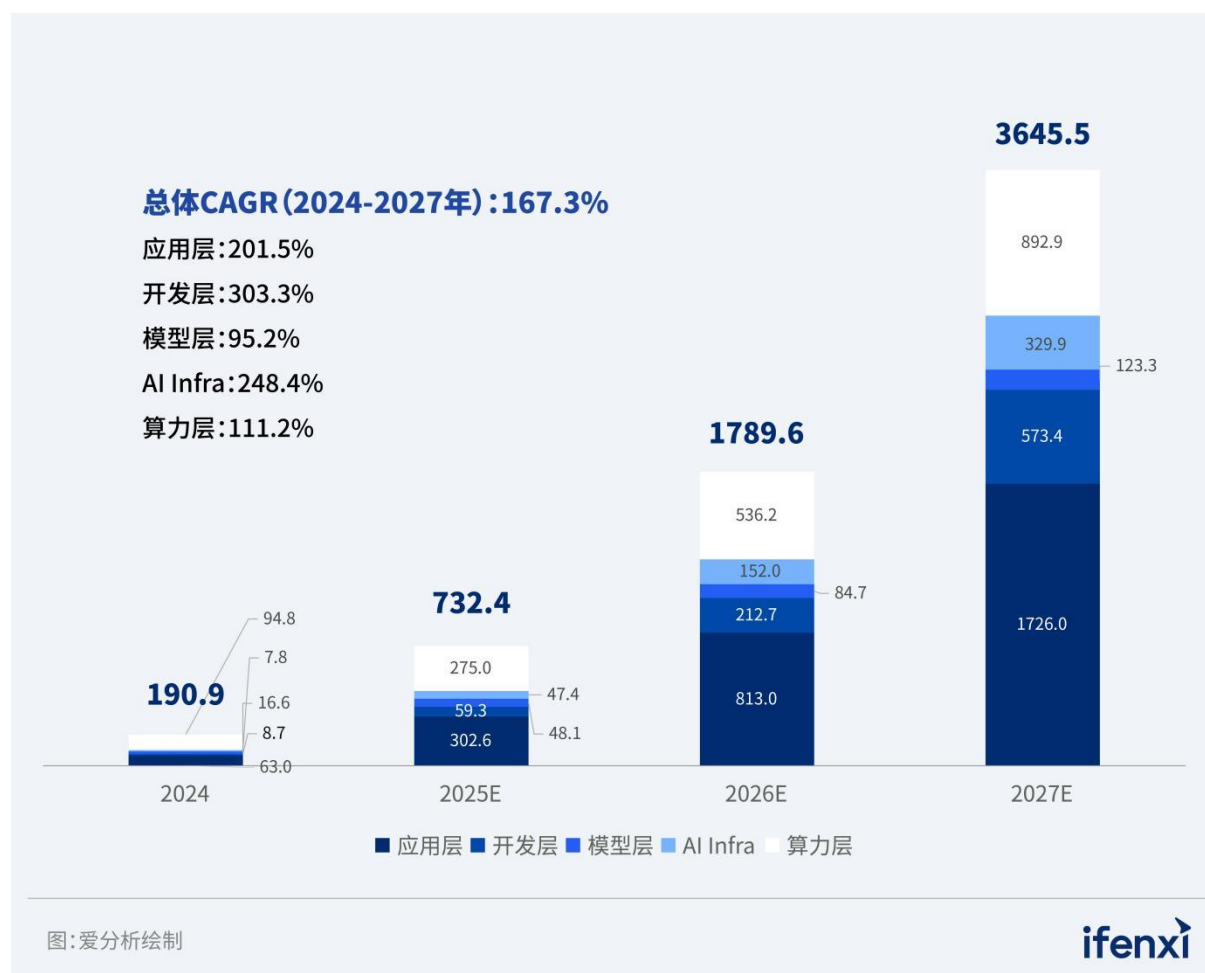


图 1 2024-2027 年中国大模型产业规模 (单位:亿元人民币)

根据我们的测算,中国大模型产业的规模将迎来爆发式增长:

- 2024 年,中国大模型产业规模为 190.9 亿元人民币。

- 从 2024 年到 2027 年，产业的年均复合增长率（CAGR）将高达 **167.3%**。
- 到 2027 年，产业总规模预计将增长至 **3,645.5 亿元人民币**。

然而，仅仅看到总量的爆发式增长是不够的，产业内部的结构分化更能揭示未来的价值核心。我们将产业解构为算力层、模型层、开发层、应用层以及支撑性的 AI Infra，各层级的增长速度存在显著差异。

- 高增长层级 (CAGR: 200%-300%): 应用层、开发层、AI Infra
- 稳健增长层级 (CAGR: 90%-110%): 模型层、算力层

这一数据差异背后，是清晰的价值迁移路径：

- **价值重心正从“基础能力”向“场景落地”高速迁移。**模型层和算力层的增速相对较低，但这并不代表它们不重要。相反，这标志着基础大模型和算力正在加速成为类似于“水”和“电”的新型公共基础设施。随着市场竞争加剧和技术成熟，这一层的市场集中度将提高，价格趋于稳定，其增长将更多由整体用量的提升驱动，而非单价。
- **应用层是增长最核心的引擎，是 Agent 价值的直接体现。**高达 200%-300%的复合增长率预示着，绝大部分的商业价值将在应用层被创造和捕获。企业客户最终购买的不是一个抽象的模型 API，而是一个能够解决具体业务问题、创造真实商业价值的解决方案——例如“AI 核保员”、“AI 工程师”或“AI 法务助理”。应用层的繁荣，正是“劳动力重构”这一宏大叙事的具体落地。
- **开发层与 AI Infra 的爆发是“淘金热中的卖水人”。**要支撑应用层的“寒武纪大爆发”，必须有一个繁荣的开发者生态和完善的配套设施。开发层（如 Agent 开发管理平台）和 AI Infra（如模型推理加速平台）的超高速增长，是应用层需求旺盛的直接投射。它们为开发者降低了门槛、提高了效率，是整个生态能够快速运转和创新的关键齿轮。

## 2. 结构重塑——应用层将超越算力层成为最大市场

大模型产业正经历一场深刻的市场结构变革。一个核心趋势日益清晰：市场重心正从基础设施建设，转向上层的应用与服务。数据预测显示，“算力层”与“应用层”的市场地位将发生历史性的对调。

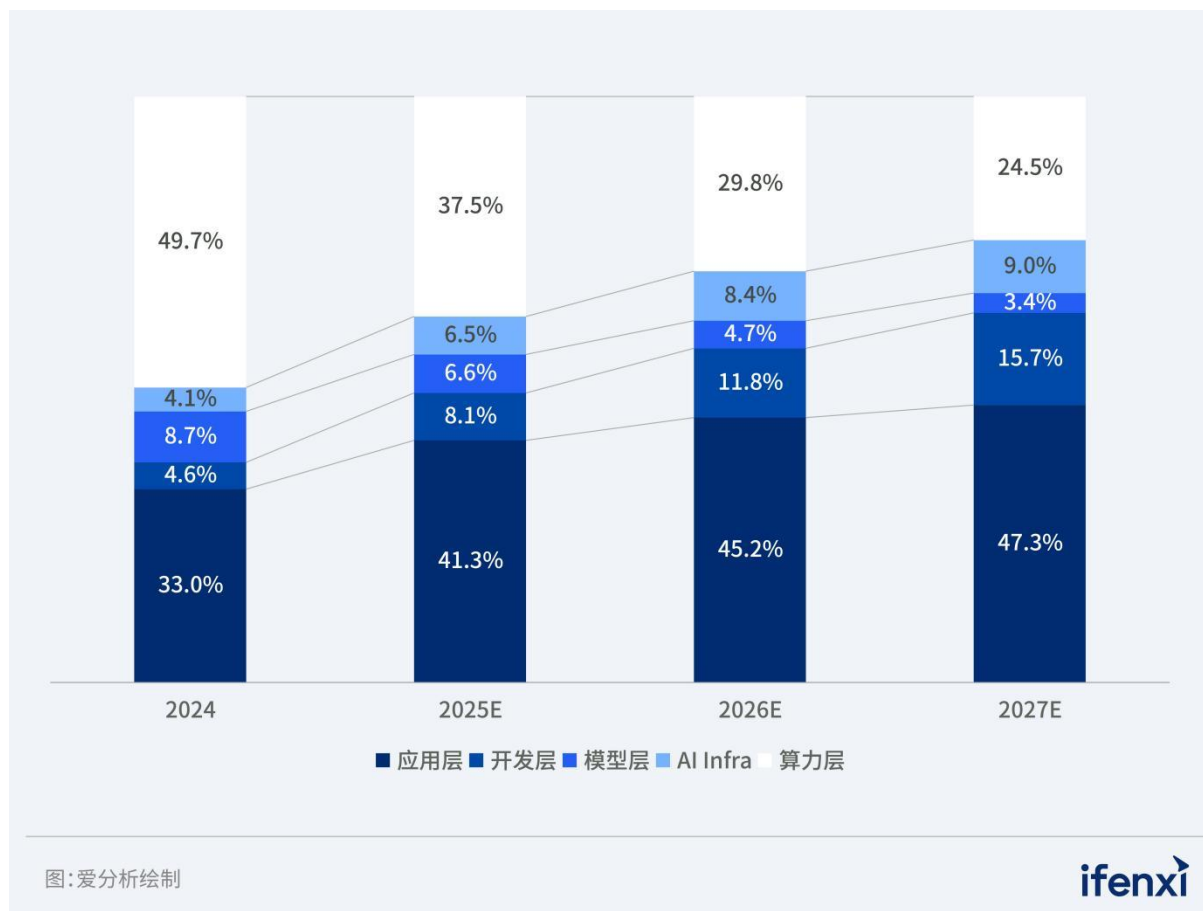


图 2 2024-2027 年中国大模型市场分布情况

### 2.1 2024 年：“基础设施驱动”的初期格局

在 2024 年，市场格局呈现出典型的“基础设施驱动”特征。算力层占据着绝对的主导地位，其市场占比高达 49.7%，几乎占据半壁江山。这反映出当时市场的首要任务是进行大规模算力建设，以满足训练和部署模型的需求。

### 2.2 2025-2027 年：应用层崛起，主导地位易手

这种格局并非一成不变。随着企业算力建设逐步成熟，其市场占比将进入持续下降通道——从 2025 年的 37.5%，最终在 2027 年预计降至 24.5%。这表明市场正从“硬件投入”阶段转向“价值挖掘”阶段。

与此形成鲜明对比的是，应用层正以前所未有的速度崛起。在 2024 年，应用层占比为 33.0%，仍居于算力层之后。但随着技术成熟和行业需求释放，其占比将水涨船高：

- 预计在 2025 年达到 **41.3%**，超越算力层，成为产业中的最大市场，
- 最终在 2027 年达到 **47.3%**。

### 2.3 其他层级的演变路径

除了两大核心板块的此消彼长，其他层级也呈现出清晰的演变路径：

- **开发层**：市场占比将从 4.6%（2024 年）稳步上升至 15.7%（2027 年）。这表明为开发者提供高效工具和平台的“卖水”生意将日益重要。
- **AI Infra**：市场占比从 4.1%（2024 年）增长到 9.0%（2027 年）。这反映出在大规模应用部署背景下，对数据管理、模型部署和运维服务的需求将持续增加。
- **模型层**：市场占比则呈现下降趋势，从 8.7%（2024 年）降至 3.4%（2027 年）。这主要是受**开源模型的冲击，导致闭源模型商业化路径受到根本性挑战。**

### 3. 企业选型策略——大厂（平台型）与初创（专业型）的场景分野

大模型的兴起极大地降低了 AI 应用开发的门槛，但也为企业的产品选型带来了新的挑战。市场上的供应商正清晰地分化为两大阵营：以大型厂商为代表的“平台型”供应商和以初创企业为代表的“专业型”供应商。对于寻求 AI 赋能的企业而言，理解二者的能力边界和优势场景，是制定正确采购决策的关键。

### 3.1 何时选择大型厂商（平台型供应商）

大厂（如互联网巨头、云服务商）的核心优势在于其长期积累的“三驾马车”：雄厚的算力基础设施、海量的多维数据和庞大的用户生态。这使得它们天然占据了产业链的制高点。

企业在以下场景中应优先考虑大型厂商：

- **寻求通用的、与现有办公生态集成的能力。** 大厂具备最强的基础模型研发能力，并已将其深度集成到操作系统、浏览器和办公套件（如 Office Copilot、钉钉、企业微信）等核心生产力工具中。典型场景：企业内部的通用 AI 助理、文档总结、代码生成、多语言翻译、会议纪要。选型理由：这些功能依赖强大的通用模型能力，且与企业现有软件生态绑定。选择大厂可以确保最佳的集成体验、可扩展性和服务稳定性，避免数据孤岛。
- **依赖底层云服务和模型即服务（MaaS）。** 各大云服务商（如阿里云、腾讯云、华为云）提供从 IaaS 到 PaaS 再到 MaaS 的完整技术栈。典型场景：企业希望利用强大的基础模型 API 作为“底座”，自行在上层开发应用；或者需要统一的 AI 开发平台（PaaS）来管理和调度多个模型。选型理由：大厂提供了最具成本效益、弹性和安全合规的底层服务。企业无需自行承担高昂的算力成本和复杂的运维工作，可以按需调用 AI 能力。
- **嵌入现有 C 端或 B 端基础设施的应用。** 对于需要海量用户数据形成飞轮，或已深度嵌入企业流程的场景，大厂具有天然的渠道和数据壁垒。典型场景：电商平台的 AI 导购、社交媒体的 AI 推荐、已有的 CRM/ERP 系统内的 AI 功能升级。选型理由：在这些场景下，AI 能力需要与平台的流量入口和业务数据（如用户行为、供应链数据）紧密结合。大厂凭借其既有渠道和数据闭环，能最快实现 AI 的规模化部署和商业化。

**选型小结：当企业的核心需求是通用能力、生态集成、平台稳定性和基础设施（MaaS/PaaS）时，大型厂商是首选。**

### 3.2 何时选择初创企业（专业型供应商）

初创企业的机会在于避开通用竞争，利用其敏捷性、行业深度和交互创新，在特定领域提供大厂难以匹敌的专业解决方案。

企业在以下场景中应重点考察初创企业：

- **解决极度垂直的专业领域难题（“Know-How”壁垒）。**这是初创企业最主要的突破口。大厂的模型是“通才”，但在需要深厚行业 Know-How 和专业数据的领域，往往不如“专才”。典型场景：①法律科技：特定法系（如海商法）的合同审查、合规 AI；②科研/工业：新药靶点发现、材料基因组 AI、结合 BIM 的建筑工程设计优化；③金融风控：针对特定衍生品交易的实时风险研判。选型理由：这些场景对准确性、专业性和可解释性的要求极高。初创公司通常由行业专家团队主导，拥有大厂难以获取的“小数据”和“专家经验”，能提供真正可用的“专家级”解决方案。
- **追求全新的人机交互范式与 workflow 重塑（“AI-Native”创新）。**大模型带来的不仅是技术升级，更是交互范式的革命。初创公司没有历史产品包袱，能从零开始设计“AI 原生”的产品。典型场景：①Agent（智能体）：能自主完成特定工作流的 AI（如自动化 SaaS 营销、自主交易）；②多模态生产：针对游戏美术、广告创意等特定风格的高效生成工具；③沉浸式应用：利用 LLM 结合 XR 技术，提供创新的教育、培训体验。选型理由：大厂往往受制于现有产品线，倾向于“AI+旧软件”的优化。而初创企业能提供“AI 优先”的产品，彻底重构 workflow，带来极为明显的效率提升。
- **寻求高性价比和敏捷部署的效率工具（SME 市场）。**大量中小型企业（SME）预算有限，无法采购大厂的重型解决方案，它们需要“开箱即用”的轻量化工具。典型场景：电商 SME 的客服自动化、一键商品详情生成、社交媒体内容运营；基于开源模型的高性价比部署方案。选型理

由：初创企业能提供更灵活的 SaaS 订阅模式、更低的部署成本和更快的响应速度。它们专注于解决 SME 的某个高频痛点，性价比极高。

**选型小结：当企业的核心需求是行业深度、流程重塑（效率革命）、AI 原生体验或高性价比时，专业型初创企业是更优的选择。**

CHAPTER

03

# 对话式智能分析市场

### 3. 对话式智能分析市场

对话式智能分析是大模型应用的重要分支之一。企业在规划和落地此类项目时，不应仅关注技术的前沿性，更应聚焦于项目如何成功交付并产生实际业务价值。基于对行业一线厂商的深度调研，我们总结出以下六大关键成功要素，它们共同决定了项目的成败。

#### 1. 价值定位：超越“取数”，聚焦“高价值分析”

项目立项的价值定位是首要的成功要素。

- **避免价值陷阱：**如果产品定位仅仅是“用自然语言替代 BI 工具进行数据查询”，企业很快会发现其价值有限。业务人员在得到一个数字后，本能的反应是“然后呢？”。
- **聚焦高价值场景：**成功的项目应将价值重心从金字塔底层的“数据查询”和“数据统计”，迁移至上层的“数据推理”和“智慧决策”。
- **“报告生成”是杀手级应用：**企业真正的痛点之一是撰写分析报告。一个成功的系统应能实现“目标驱动”，即用户只提出一个目标（如“帮我写一份本月的经营复盘报告”），系统就能自主完成数据提取、多维分析、归因洞察并最终生成一份结构化报告。这种价值提升是将分析师从“以天为单位”的工作量压缩到“以分钟为单位”，其 ROI（投资回报率）远高于单纯的“取数”。

#### 2. 技术选型：采用混合架构，而非单一模型

企业常误认为此类项目只是一个大模型接口。事实上，成功的企业级应用依赖于一套复杂的混合技术架构。

- **建立统一语义层：**在企业现有的数据仓库（DWS）之上，必须构建一个统一的指标语义层。这比直接的“Text-to-SQL”路径更为稳健。大模型通过理解语义层来生成查询，而不是直接编写 SQL，这能从根本上保证指标口径的统一和准确性。

- **采用“推理模型 + 基础模型”的双模策略：**不同任务需要不同模型。简单的查数、制表，使用轻量级基础模型（如 V3 或 72B 模型）即可，响应快、成本低。但要实现复杂的归因分析、报告撰写，则必须依赖更强大的推理模型（如 DeepSeek R1）来执行复杂的思维链（CoT）任务。
- **构建多智能体 (Multi-Agent) 协作架构：**面对“生成一份财务报告”这样的复杂目标，单一 Agent 无法胜任。成功的架构应是多 Agent 协作式的：例如，一个“指标 Agent”负责取数，一个“洞察 Agent”负责归因分析，一个“报告 Agent”负责撰写。这种架构不仅能完成复杂任务，且具有高度解耦和可维护性。
- **具备企业级大数据处理能力：**许多产品在演示时使用本地 Python 脚本，只能处理万级或十万级数据。企业实际业务数据动辄上亿。因此，系统后端必须具备基于分布式计算（如 Spark）的能力，才能满足企业级大数据量的处理和性能要求。

### 3. 信任构建：确保结果“可追溯”与“可干预”

准确性是此类项目最大的挑战和落地痛点。由于大模型存在“幻觉”，用户对 AI 生成的答案（尤其是关键业务数据）天然抱有不信任感。

- **设定合理的准确率预期：**企业必须明确，系统无法达到 100% 的准确率。在项目初期，应设定清晰的验收标准：例如，对于简单的取数，准确率应达到 90% 以上；对于复杂的建模分析，准确率达到 80% 以上被认为是可接受的。
- **过程可追溯：**这是建立信任的唯一途径。杜绝“黑盒”：系统在给出答案时，必须能清晰展示其分析过程。混合生成模式：在生成复杂报告时，成功的做法是“分步走”：首先，大模型利用其

推理能力生成报告的“大纲”；然后，针对大纲中的每一个数据点，系统通过调用语义层（而非大模型编造）来执行可信的数据查询；最后，大模型再对这些“可信数据”进行分析和解读。

- **结果可干预：**“可追溯”不应只是“展示代码”。成功的系统必须允许业务人员（非技术人员）能轻松地对分析过程进行干预和修改。例如，通过可视化的算子流 或点击配置项，用户可以自行修改过滤条件、更换图表类型，并立即看到结果刷新。

#### 4. 场景选择：从“容错场景”切入，逐步赋能“核心决策”

由于准确性无法达到 100%，项目的切入点选择至关重要。

- **优先选择“高容错”场景：**项目启动初期，应选择那些对错误容忍度较高的场景进行试点。例如，在客户营销推广中，AI 分析出 10 个潜在客户，即使用户多发了 1 条错误的短信，其后果也是可接受的，但价值在于提升了效率。
- **定位为“人机协同”的辅助角色：**在准确性要求高的场景（如财务分析、安全生产），系统不应替代人，而应扮演“数据分析师助手”的角色。AI 负责完成 80%的繁重工作（如自动生成算子流或分析报告初稿），但最终的决策和审核必须由人来完成。
- **发掘“大海捞针”型场景：**在某些领域（如情报分析、商机挖掘），AI 的价值在于其强大的信息处理能力。即使 AI 分析出的 10 条线索中有 9 条是错的，但只要有 1 条是对的，就具有了极大的业务价值。

#### 5. 落地前提：客户侧的“数据治理”与“资产积累”

项目的成功不仅依赖于供应商的技术，更依赖于企业自身的“地基”。

- **数据治理成熟度是项目周期的决定因素：** 供应商无法凭空交付价值。如果企业自身的数据仓库（DWS）、数据治理、宽表建设等基础工作尚未完成，项目周期将被无限拉长。反之，如果企业数据基础良好，AI 应用的部署和初始化（如指标配置、模板配置）可在数周内完成。
- **高质量的“行业模板”是关键资产：** AI 分析的质量很大程度上取决于“投喂”给它的范例质量。企业（或供应商）积累的高质量、体系化的分析报告模板和行业黑话（Know-how），是决定 AI 能否生成深度洞察的关键，其重要性甚至不亚于模型技术本身。

## 6. 价值衡量：具备清晰的 ROI，实现“降本增效”

最后，项目必须能被清晰地衡量。成功的项目往往在立项之初就有明确的 ROI（投资回报率）目标。

- **降本：** 最直接的价值体现在“砍预算”。例如，某金融机构原有数百人的数据分析外包团队，每年成本高达数千万元。引入智能分析平台的核心驱动力之一，就是用 AI 替代大部分初级和中级外包分析师的工作，实现显著的成本节约。
- **增效：** 一是将核心分析师从繁琐的“取数”和“报告撰写”中解放出来，投入到更具价值的决策工作中；二是通过降低数据门槛，使一线的业务人员（如分支行行长）也能直接获取数据并进行自助分析，提升整个组织的决策效率和数据参与度。

## 案例 1：城商行借助数势科技 Agent 实现智能分析，降本增效

随着智能化转型的加速，越来越多的银行开始探索大模型技术的应用，以提升运营效率、优化服务质量、优化成本结构。数据分析作为大模型落地的重要场景之一，备受关注。然而，许多银行基于大模型的数据分析项目仅停留在取数阶段，无法实现报告生成这一更深层次的需求，导致大模型在银行数据分析领域的应用价值较为有限。

在这种背景下，某头部城商行与数势科技的合作，为这一难题提供了创新解决方案。通过引入先进的技术和架构，该城商行成功实现了从自然语言取数到深度洞察报告生成的跨越，显著提升了数据处理效率，同时大幅降低了人力成本。本文将详细探讨这一合作项目的背景、解决方案及实施成果，为其他银行提供参考。

### 01.大模型驱动，某城商行开启数据分析降本增效新路径

某城商行的数据管理部拥有一个 500 人左右的团队，其中约一半为外包人员。外包成员主要包括数据分析师、产品经理、工程师等角色，其主要服务内容是利用 BI（商业智能）工具，为分支行领导及业务人员提供数据提取服务，并撰写相关报告。

以核心报告之一——分支行业绩对比及经营考核报告为例。在总行进行经营分析时，需要完成两项主要工作：一是对所有分支行的核心指标进行排名；二是撰写绩效考核报告。这一过程目前主要依靠人工完成。例如，若围绕 10 个指标撰写一份经营分析报告，需先由数据分析师提取相关数据报表，再由业务人员依据该报表制作报告模板，随后数据分析师与业务人员协同手工撰写报告，整个流程通常耗时约两天。

由于效率偏低，虽然数据管理部人数较多，但仍常出现工单积压的情况。并且，该城商行目前每年在 200 多名外包人员上的成本约为 7,000 万元。为优化成本结构，银行管理层期望借助大模型

技术，提升数据处理效率，以及节省相应的人力成本。

## 02.指标语义层+推理模型+Agent，数势科技为城商行降本增效提供新解法

在考察了一些智能分析厂商后，该银行发现目前的智能分析产品大多仍停留在查数这一基础场景上，价值比较有限。然而，银行内部真正高频且亟待解决的痛点是报告生成。例如，员工在向领导汇报工作时，往往需要先制作一份业绩复盘报告，其中不仅要包含数据表格，还要有对数据的解读以及相应的策略建议。只有真正满足报告生成这一核心需求，才能有效赋能团队，实现智能决策。

在此背景下，该银行与数势科技展开接触，并认识到数势科技的独特优势，进而与之展开合作。

具体而言，数势科技的三点独特优势如下：

- **高回答准确率：**回答准确率是银行做数据应用的基础。当前市场上的 ChatBI 类产品多采用 NL2SQL 技术路线，通常准确率在 60%，且存在数据安全风险，而数势科技 SwifAgent 产品则在 2023 年便首创了大模型+指标平台的技术路线，在取数环节采用选 NL2Semantic Layer 的方式，也就是在自然语言和数据底表间构建指标语义层，让大模型去做擅长的自然语言意图理解，让指标语义层去做和底表数据之间的业务逻辑和映射关系，从而解决大模型幻觉问题，将自然语言取数准确率实现 100%。
- **深度需求满足：**在满足了精准问数需求的基础上，数势科技 SwiftAgent 也在 2025 年实现版本升级，通过融入 DeepSeek R1 推理模型和企业知识库，可自动化的沉淀企业知识，降低对于微调的依赖，更好地满足银行进行数据波动归因、预警分析和报告生成等深度需求，为银行提供一站式智能指标平台和智能分析方案。
- **成功案例背书：**据公开资料了解到，2024 年数势科技与中原银行合作智能指标平台建设，项目

以数势科技智能分析助手 SwiftAgent 为核心产品，为中原银行构建统一指标管理平台，确保指标定义的统一性和标准化，通过低代码、自动化的指标生产代替人工开发进而提高指标交付效率，以便更好地挖掘数据中的价值。在此基础上，项目还融合大模型能力构建智能分析平台，实现指标问答、图表生成、报告生成以及归因分析等灵活应用，有效提升了业务分析决策效率，助力银行业务发展。银行内部人员对该项目给予了高度评价。这一成功案例为数势科技的技术实力和产品可落地性提供了有力背书。

### 03.数势科技 SwiftAgent 的产品架构与核心落地场景

SwiftAgent 产品最新版本的核心架构分为三层：数据层、引擎层和应用层。

**数据层**支持多源异构数据接入，提供灵活的数据集成方案，包括存算一体、存算分离和存算外置。通过这些方案，确保不同数据源能够无缝连接并进行高效处理。

**引擎层**是整个产品的智算核心，包含三个关键模块：

- **数据语义模块**：通过精准的指标和标签平台，解决大模型的幻觉问题，确保自然语言与数据之间的精准映射。该模块通过结构化转译消除语义鸿沟，并利用预计算加速引擎将复杂查询响应时间压缩至秒级，确保实时决策需求的满足。
- **智能模型引擎模块**：采用混合智能架构，结合大模型和小模型的动态协同。大模型负责语义理解与逻辑推演，调用行业知识增强的百亿参数模型；小模型则专注于结构化分析，处理时序解析和指标计算等任务。动态路由控制器根据任务自动分配算力，实现推理成本与精度的最优平衡。
- **数据分析技能池**：涵盖从数据提取到归因分析再到报告生成的完整闭环，提供全面的数据分析能力，帮助决策者获取精确的洞察。

应用层是 SwiftAgent 向上为企业提供功能场景的重要支撑，通过 Multi-Agent 架构，将多个数据应用智能体协同工作。在不同业务场景下，智能体可以联合调用，如分析报告场景需要同时调用取数 Agent、分析 Agent 和报告 Agent，实现高效协作，满足业务需求。

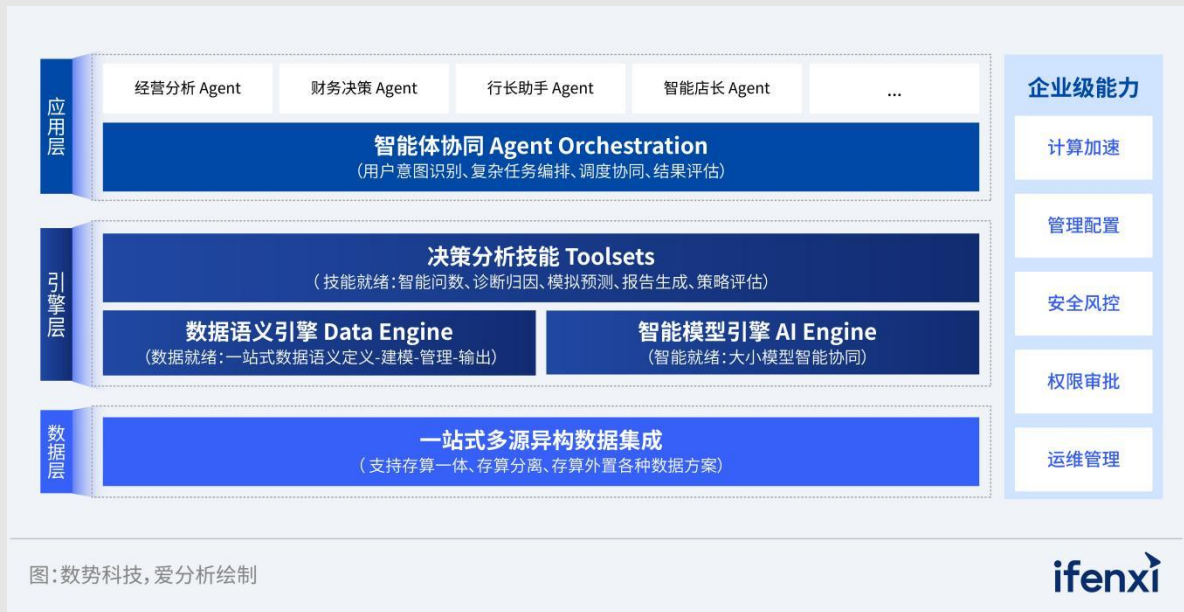


图 3 数势科技 SwiftAgent 产品架构图

基于以上产品架构，在与该银行的合作中，数势科技通过 SwiftAgent 智能分析助手的核心能力，显著提升了企业的数据分析准确性以及洞察决策的效率和质量。

## 1. 低门槛取数与数据洞察

SwiftAgent 利用自然语言交互技术结合 DeepSeek 大模型，使业务人员无需掌握复杂的技术工具，便能通过对话式查询快速获取数据和可视化结果。该功能不仅极大降低了数据获取门槛，还能够将数据结果清晰地呈现给决策者，在该银行的应用，支持业务人员迅速生成包含资产配置优化建议、市场趋势预测等关键报告，帮助金融机构精准研判市场动态，提升决策的专业性和效率。

## 2. 多端适配与决策敏捷性

SwiftAgent 的多端适配能力，确保无论在 PC 端还是移动端，该银行领导也团队都能随时随地获

取数据分析结果，使日常运营决策更加迅速和高效。

### 3. 智能报告生成与深度分析

结合 DeepSeek 大模型的能力，SwiftAgent 在深度思考、归因解读、智能报告生成和决策建议四大维度实现了飞跃。在该项目落地中，SwiftAgent 能够在 5 至 10 分钟内给提供精确、符合银行需求的深度分析报告，报告内容不仅全面，还支持该城商行提供模板定制分析维度，具有高度的可用性和决策支持价值，使得银行的管理层能够迅速调整战略，提高决策效率。



图:数势科技,爱分析绘制



图 4 数势科技 SwiftAgent 生成 AI 报告功能展示

### 04.从 2 天缩短至 10 分钟，数势科技 SwiftAgent 赋能银行报告撰写效率

数势科技 SwiftAgent 报告生成能力是在自然语言精准取数的基础上，结合 DeepSeek 的深度思

考能力和沉淀的金融行业分析模板, 结合自研小模型和多 Agent 架构调用共同实现的。具体而言, 报告生成的过程包括以下几个关键步骤:

- 1. 报告框架生成:** 首先, 基于企业数据库和行业通用知识, 自动化生成报告框架, 作为后续内容填充的结构基础;
- 2. 子问题生成:** 根据指标语义层, SwiftAgent 生成报告中每个子问题, 严格依据存量指标进行构建;
- 3. 数据提取与分析:** 针对每个子问题, 自动化调用相应的 Agent 进行数据取数、分析, 并生成相应的内容;
- 4. 结果填充:** 最终, 生成的分析结果被填充到报告框架的相应位置, 完成报告的形成。

通过这种创新的 AI 报告生成能力, SwiftAgent 能够为用户提供结构化的报告, 如金融领域的资产配置优化建议和市场趋势预测。这种报告不仅帮助用户精准研判市场动态, 也提升了投资决策和风险管控的专业效能。

## 05.六周完成项目落地, 该城商行信用卡部门报告生成效率提升 40 倍

### 1. 项目落地步骤一: 分析报告场景与指标体系梳理 (约 2 周)

在引入 AI 报告生成技术的初期, 关键任务是将抽象的业务需求转化为可量化的分析指标。项目团队首先与业务部门合作, 确定从信用卡用户活跃度和营销策略分析两个场景入手。通过从银行数据中台接入信用卡交易数据、手机银行行为日志以及第三方支付渠道等数据, 团队构建了涵盖 DAU (日活跃用户数)、MAU (月活跃用户数)、消费频次、绑卡交易占比、留存率等指标的活跃度分析体系。



图 5 信用卡用户活跃度分析体系建设方法

同时，在信用卡运营策略效果评估方面，团队与银行共同对活动进行了细致分类，针对促活类、增收类、留存类活动分别设计了不同的评估指标和归因分析模型。

## 2. 项目落地步骤二：分析模板设计与持续优化（约 2 周）

实现报告生成自动化与智能化的关键在于打造可复用的分析模板。项目团队与银行分析师合作，定义了分析模板的整体架构，并为每种分析范式设计了不同的可视化类型和总结侧重点。针对核心信用卡活跃类指标的归因分析，采用了维度归因、因子归因和分析树归因三种范式的融合方案，确保报告接收方能够从多个视角全面了解指标变化的原因。通过与银行分析师的共同努力，项目团队建立了约 20 个分析模板库，按报告类型分类存储，并记录版本迭代历史。这一举措旨在提升

AI 生成报告的采纳率。测试数据显示，采纳率由最初的 30%提升至 80%。

### 3. 项目落地步骤三：培训推广，从试点到规模化的跨越（约 2 周）

试点推广是验证项目价值的关键环节。项目团队与银行合作，选择信用卡中心的用户运营与活动运营团队作为试点对象，覆盖超过 50%的分析师团队，并建立了问题反馈通道，每周收集用户需求。经过 2 周的试点运行，报告生成时间从原来的 2.5 天（纯人工撰写）缩短至 30 分钟（大模型生成 10 分钟+ 人工辅助优化 20 分钟）。

在为信用卡中心快速提效并完成速赢场景后，项目团队制作了培训视频，将报告生成能力推广至对公贷款业务团队，鼓励更多业务团队尝试使用 AI 报告工具。

#### 项目效果

经过一系列实践和探索，该头部城商行信用卡部门在 AI 报告生成方面取得了显著成效。根据业务团队实测反馈，报告生成效率提升了约 40 倍，从 2.5 天缩短至 30 分钟。这一效率提升使业务团队能够将更多精力投入到信用卡生命周期促活策略的制定和优化中，而非重复进行数据搬运工作。



图 6 报告生成用时对比

此外，在项目开展过程中，银行沉淀了 20 多个分析模板，涵盖了用户活跃度分析和营销策略分析等场景。这些模板将原本存在于分析师脑海中的“分析经验”产品化，并让大模型得以吸收，为后续分析思路的“可复制性”奠定了坚实的知识基础。

CHAPTER  
04

# “大模型+政务”市场

## 4. “大模型+政务”市场

大模型在政务领域的应用（如政府门户网站的智能问答），其核心价值在于**激活“沉睡”的政务数据**。

政府在常年运行中积累了海量的网站、政策和服务数据，但这些数据往往未被有效利用，公众访问率低、查找困难。

然而，政务场景对信息的权威性、准确性和安全性要求极高，这使得大模型落地面临极大挑战。基于对行业一线项目的深度分析，我们总结出以下**五大关键成功要素**，它们共同决定了政务 AI 项目的成败。

### 1. 信任与安全：严控“幻觉”是项目的生命线

在政务领域，信息的权威性压倒一切。项目最大的难点和核心痛点，就是解决模型的“幻觉”和敏感问题。

- **严格的内容边界控制 (RAG)**：这是杜绝幻觉的首要技术手段。系统必须限制大模型只在政府提供的、受控的数据库范围内进行回答。对于超出范围的问题（例如非本领域问题），系统应配置为拒绝回答，而不是任由模型从外部循环中编造答案。
- **完备的敏感性知识库**：政务应用必须配备一套积累深厚的敏感词库和问题库。这不仅包括常规的敏感词，还必须针对政务领域的特殊性，如涉及落马官员的政绩评价等问题，进行严格的限制和预设回答。
- **前置的意图识别**：成功的架构通常会在大模型前端部署一个“小模型”。这个小模型不负责生成答案，而是专门用于快速识别用户的意图（例如：是查询政策、查询领导信息，还是查询新闻资讯），然后将问题精准路由到对应的专业知识库（如政策库、领导人库等）。

### 2. 数据基础：统一的数据底座是项目加速器

大模型的效果取决于数据质量。供应商能否成功交付，很大程度上不依赖于模型本身，而依赖于数据基础。

- **统一资源库**：成功的政务项目往往建立在一个预先建成的“统一资源库”之上。政府客户（如省级政府）如果已经将下辖数百个网站的数据统一汇聚到一个标准化的平台进行管理，将极大降低 AI 项目的实施难度。
- **数据的实时同步**：政务信息（如领导任免、政策发布）具有高时效性。AI 所依赖的知识库必须与政府网站的生产系统（内容发布系统）完全打通。当网站内容更新时，知识库必须实现实时同步，否则 AI 将提供过时甚至错误的信息。
- **向量化基础**：如果底层的“统一资源库”在建设时就天然支持向量库，那么数据的接入和实时入库将变得非常简单和高效。

### 3. 领域深度：超越技术，积累政务“Know-How”

政务服务不是一个通用的聊天场景，它需要深度的行业知识。

- **深度的数据加工能力**：仅仅采集数据是远远不够的。成功的关键在于对原始数据（尤其是长文本的政策文件）进行深度的结构化加工和拆解。例如，必须能理解一篇上千字的政策，并将其关键要素（如补贴对象、申领条件、截止日期）提取出来。
- **积累行业通用数据资产**：一个关键的成功要素是，供应商自身是否积累了海量的行业通用数据。如果供应商拥有一个覆盖全国的“政策库”或“政务服务库”，就可以摆脱对单一客户提供数据的依赖。这不仅能保证数据来源的丰富性，还能实现跨区域比较（如比较两市的购车优惠政策），极大提升服务价值。

- **理解公众真实诉求：**必须基于对政府网站长期的用户行为分析（如公众常搜什么、常看什么），来理解公众的真实意图和诉求，并以此为中心优化模型和服务。

#### 4. 架构能力：支持快速复制与迭代

政府体系是分层级的（省、市、县），一个成功的应用必须具备高度的可复制性。

- **中间层管理平台：**在基础模型和上层应用之间，必须有一个强大的“一体化管理平台”。该平台用于管理数据来源、配置数据处理能力、设定数据范围限制以及管理权限。
- **应用的快速复制：**基于该管理平台，为省政府开发的应用可以被快速“克隆”出一个市级版本。实施过程仅需在新应用中配置其对应的数据源和权限，而无需重新开发。

#### 5. 客户因素：关键决策者的认知与推动

- **核心决策者的推动：**项目往往由对新技术有深刻认知、且有强烈意愿解决“数据沉睡”痛点的核心部门（如电子政务处）牵头推动。
- **对新技术的合理容忍度：**政府客户对新技术的接受程度和容忍度至关重要。愿意积极拥抱新技术的客户，会给予项目更多的测试和优化时间，并通过声明（如“AI生成内容，请自行鉴别”）来共同管理风险。

## 案例 2: 智能问答大模型“小贵”上线，助力贵州政务服务升级

贵州省人民政府办公厅为提升政务服务水平，推动数字政府建设，与拓尔思合作开发了智能问答大模型“小贵”。该模型基于拓尔思的拓天政务大模型和大模型一体化管理平台，通过多层次架构设计，实现了对政务数据的深度整合与智能化应用。

“小贵”已正式上线贵州省人民政府官网，提供包括猜你想问、推荐主题、办事主题等功能在内的智能问答服务。它能够一键触达全省各级政府网站及应用的信息和服务，支持 PC 端和移动端，显著提升了搜索的精准度和效率。

项目的实施不仅优化了公众的使用体验，增强了获得感，还巩固了贵州省政府网站集约化改革的成果，提升了资源利用效率。此外，“小贵”的构建模式具备可复制性，为其他地区政务智能化建设提供了参考。



图 7 智能问答大模型“小贵”主界面

## 01.贵州省政府办公厅：以大模型技术突破政务数据价值瓶颈

随着数字政府建设的深入推进，贵州省政府办公厅积极响应数字政府实施工作要求，致力于创新探索云计算、大数据、人工智能、区块链等新型技术在数字政府领域的应用。贵州省政府网站集约化平台作为支撑全省党政机关网站的重要平台，承担着巩固拓展集约化改革试点成果的重任。为实现这一目标，省政府办公厅积极探索利用大模型技术，推动政务大模型在生成、问答、推理等方面的场景应用建设，以满足用户实际业务需求，提升管理和服务水平，打造整体联动、高效惠民的网上政府，进一步优化网站体验。

贵州省作为大数据建设的先行者，建立了全省统一的集约化网站，集中管理全省党政机关网站。然而，随着网站运行，积累了大量政府自有数据，这些数据虽沉淀于网站之中，却未能充分发挥其价值。如何有效利用这些数据，更好地为公众提供服务，成为贵州省政府办公厅亟待解决的痛点问题。以往，政府网站访问量相对较低，公众仅在有需要时才会访问，且有时难以找到相关数据。尽管政府内部沉淀了大量政务服务数据、政策数据等，但未能有效利用，导致数据价值未能充分体现。

在此背景下，贵州省政府办公厅提出建设智能问答大模型的需求，旨在通过大模型赋能贵州省政务智能化应用建设，充分整合现有业务、数据、系统等资源，以基础设施、数据资源、基础能力为支撑，聚焦省政府核心业务，如省政府办公厅的办文、办会、办事等职能。通过对业务职能和需求的深入分析梳理，切实推动信息化、数字化、智能化政府的发展目标，开展业务能力提升与转型，加快推进大模型落地应用，建设覆盖 PC 端、贵政通、中国贵州 APP 等多个应用场景，为贵州省政府数字化建设的发展和 innovation 注入新动力。

## 02.拓尔思：政务数据与智能技术的深度融合

经过综合评估，贵州省政府办公厅选择拓尔思作为合作伙伴，承担智能问答大模型的建设

任务。这一选择主要基于拓尔思在政府行业多年深耕所积累的丰富经验和技術优势，具体体现在以下三个方面：

### **一、强大的数据资源与深度加工能力**

拓尔思在政务领域积累了深厚的数据资源，建立了全国政策库，覆盖国家、省、市以及部分县的几乎所有互联网发布的政策内容。这些数据经过严格的归纳和整理，形成了高质量的数据资源体系。数据的全面性和高质量是大模型效果的关键，拓尔思凭借自身强大的数据采集和深度加工能力，确保了项目实施的高质量。例如，拓尔思能够对政策进行深度分析和比较，为用户提供精准的政策咨询服务。以新能源汽车购买政策为例，拓尔思可以对比不同城市（如成都与重庆）的优惠政策，为用户提供有价值的结论，显著提升用户体验。此外，拓尔思不依赖客户提供的数据，而是通过自身积累的全国数据资源，为项目提供坚实的数据支撑。

### **二、精准把握政府部门业务需求**

拓尔思对政府部门的业务需求有着深刻的理解。拓尔思能够精准拆解政策文件，提取关键要素，并通过技术手段将其转化为可操作的解决方案。拓尔思在政府门户网站建设与运营方面拥有丰富经验，通过用户行为分析，深入了解公众访问政府网站的目的和需求。例如，拓尔思能够分析用户在政府网站上经常搜索的内容，围绕这些常见诉求进行优化，提升用户满意度。凭借大量用户行为数据的积累，拓尔思能够精准把握公众需求，为政府部门提供更具针对性的服务。

### **三、完善的大模型幻觉解决方案**

拓尔思针对大模型可能出现的幻觉问题，提供了一套完善的解决方案。首先，拓尔思通过

限制模型的作答范围，确保其仅在提供的数据库内进行回答，避免模型在外部循环中产生错误信息。此外，拓尔思积累了丰富的敏感词库和敏感问题库，涵盖政治敏感内容等，通过知识库体系对模型进行严格限制，确保信息的准确性和安全性。同时，拓尔思在大模型前端部署了一个大模型，用于识别问题类型和用户意图，精准定位访问的数据库。

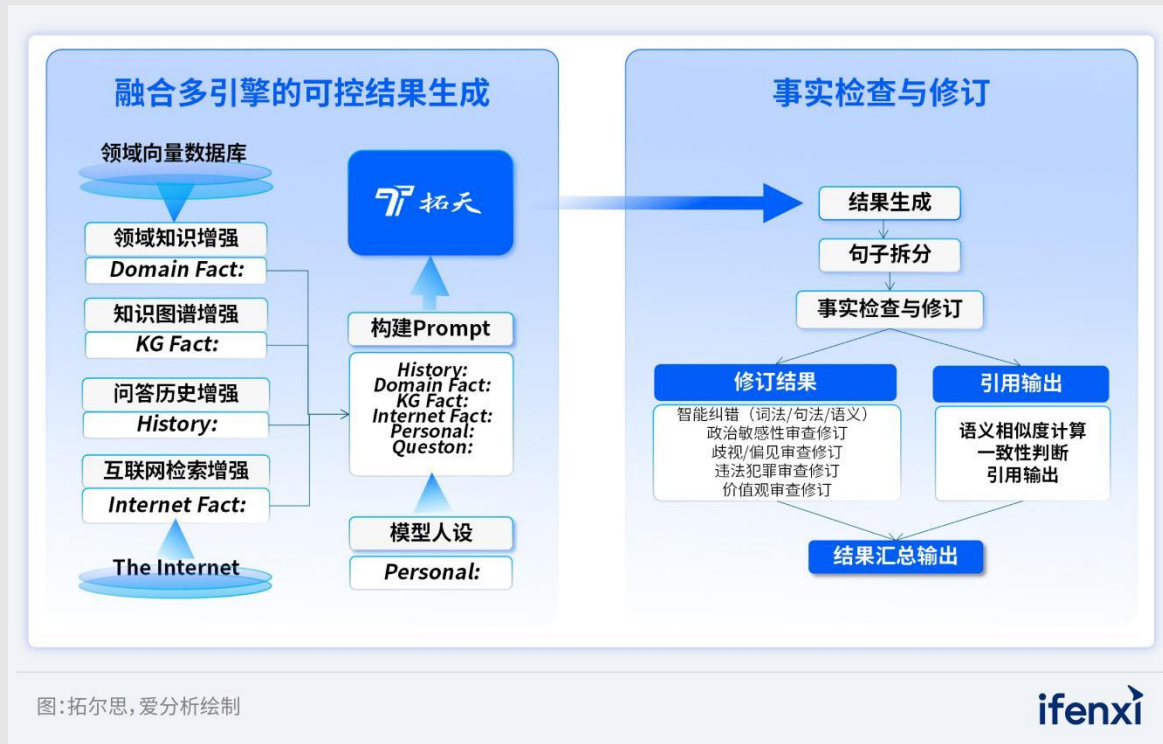


图:拓尔思,爱分析绘制



图 8 拓尔思融合搜索引擎架构

### 03 多层次架构驱动政务智能化升级

为满足贵州省政府办公厅的需求，拓尔思提出了一套完整的解决方案，充分发挥其在政务数据和智能技术方面的优势，实现政务智能化应用的高效建设与管理。

#### 一、多层次的解决方案架构

拓尔思的解决方案由多个层次构成，以确保系统的高效运行和灵活管理。

##### (一) 基础模型：拓天政务大模型

拓尔思训练了基于 DeepSeek 的 32B 基础模型，并在此基础上开发了专门的政务基座模型——拓天政务大模型。这一模型经过政务领域的深度优化，能够更好地理解和处理政务相关的问题和数据。

## （二）大模型管理平台

在基础模型之上，拓尔思搭建了大模型管理平台。该平台的核心功能包括：

- 1.数据接入与处理：**解决政府网站数据的接入问题，确保数据能够按照大模型的要求进行处理和整合。
- 2.幻觉问题的规避：**通过严格的管理和限制，有效杜绝模型可能出现的幻觉回答，确保信息的准确性和可靠性。
- 3.用户权限管理：**实现对用户数据权限的精细管理，保障不同用户之间数据和知识库的安全性和隐私性。
- 4.应用快速搭建：**支持快速搭建新的应用场景，例如为不同层级的政府部门快速部署智能问答应用，提升项目的灵活性和扩展性。

## （三）互联网端服务

在管理平台之上，拓尔思提供面向互联网端的服务，确保用户能够通过多种渠道（如 PC 端、移动端等）便捷地使用智能问答功能。



图 9 拓天大模型一体化平台整体架构

## 二、大模型在具体场景中的应用

以一个具体的场景为例,说明大模型在解决方案中的具体作用。假设用户提出了一个问题:

“当前贵阳市政府的领导班子有哪些?” 以下是整个处理链路的详细说明:

- 1.问题初步处理:** 用户的问题首先被一个大模型接收,该模型负责对问题进行初步分析,识别是否存在敏感内容或关键词。这一环节确保了问题的安全性和合规性。
- 2.意图识别与数据检索:** 如果问题通过初步筛查,大模型会进一步识别问题的意图,并确定需要查询的数据库。例如,对于涉及政府领导班子的问题,系统会优先检索拓尔思专门建立的领导人库。如果领导人库中没有相关信息,系统会自动扩展搜索范围,查询其他相关数据库(如新闻资讯库)。
- 3.数据整合与推理:** 检索到的相关数据会被整合并输入到拓尔思的拓天政务大模型中。大

模型利用其强大的理解能力和推理能力，对数据进行分析和处理，生成准确的回答。

**4.最终回答输出：**经过大模型的处理，系统会生成一个准确、可靠的回答，并反馈给用户。

如果在领导人库中找到了明确信息，系统会直接提供准确答案；如果信息不完整，系统会根据其他数据库中的线索，给出一个基于现有数据的合理推测，并明确标注信息来源和可靠性。

#### 04.智能问答大模型“小贵”的应用与效益

贵州省人民政府门户网站智能问答大模型“小贵”已正式上线运行，面向全省公众提供服务，涵盖个人用户和企业用户。用户可通过进入贵州省人民政府门户网站，点击右下角具有民族特色的卡通形象——小贵，便捷地进入“智能问答大模型”界面。



图 10 贵州省人民政府门户网站的“小贵”入口

“小贵”的主要功能包括：猜你想问、推荐主题、会话框、办事主题（个人办事、企业办

事)、热门问题、历史会话等。



图 11 “小贵”的主要功能

## 一、应用前与应用后的对比

**应用前：**贵州省人民政府网站未提供专门的智能问答服务。公众主要通过网站浏览和智能检索功能查询所需资讯及服务。这种方式在信息查找的效率和精准度上存在一定的局限性，难以满足用户多样化的需求。

**应用后：**智能问答大模型“小贵”实现了在搜索框中一键搜索贵州省人民政府网站、中国贵州 APP 以及全省省直部门、各地州市部门门户网站的全部信息与服务。它全面支持电脑 PC 端与手机移动端搜索服务，显著提升了贵州省人民政府网的搜索服务水平，实现了“随时搜”“随处搜”“随心搜”的能力，并达到了“搜得准”“搜得快”“搜得全”的目标。

## 二、社会效益

贵州省政府网站集约化平台是支撑全省党政机关网站的重要平台，面向全省公众，包括企

业用户和个人用户。智能问答大模型“小贵”以更友好的交互、更主动的服务、更简洁的流程、更直观呈现，回答用户的问题、解决用户的需求。这将进一步提升贵州省政府网站的体验感，增强公众的获得感，推动政府服务向更加高效、便捷、人性化的方向发展。

### 三、经济效益

贵州省政府网站集约化平台是集约化改革试点的重要成果。智能问答大模型对集约化平台的提升，有助于巩固和拓展集约化改革试点成果，确保前期的改革投入能够持续发挥效益。通过优化平台功能和服务，提高资源利用效率，降低运营成本，为政府数字化转型提供有力支撑。

### 四、项目的可复制性与拓展性

贵州省人民政府门户网站智能问答大模型是基于大模型一体化平台构建的，通过知识库、指令库、场景库等的配置，打造出适用于贵州省的智能问答应用。该平台具备高度的灵活性和可配置性，能够根据不同的数据和业务场景，快速构建不同政务业务场景下的智能问答应用。

拓尔思的大模型一体化管理平台为项目的快速搭建和部署提供了有力支持。在该平台上，用户可以轻松完成新建应用的配置，包括连接所需数据库、选择使用数据及数据处理能力，以及设定数据范围等。这一平台的高效性和便捷性，使得类似应用的搭建变得快速且高效，为其他地区的政务智能化建设提供了可借鉴的范例。

CHAPTER

05

结语

## 5. 结语

中国大模型市场正站在一个关键的拐点。回顾过去，政策的强力主导、技术的爆发式突破与需求的初步觉醒，共同完成了市场的基础设施构建和认知教育。展望未来，本报告的三维分析框架（政策-技术-需求）清晰地指向同一个结论：**市场的核心议程已从“建好”大模型，全面转向“用好”大模型。**

我们的研究数据明确显示了这一历史性的价值迁移。到 2027 年，应用层将超越算力层成为市场最大板块，其高达 200%-300% 的年均复合增长率预示着，绝大部分商业价值将在场景落地中被创造和捕获。这不仅是市场份额的易手，更是产业逻辑的根本重构——算力和基础模型正加速成为新型“水电煤”，而真正的价值创造来自于上层的解决方案和 AI 原生应用。

技术的成熟是这场变革的底气。MoE 架构的成熟、RAG 的普及以及轻量化技术的进步，共同推动 AI 从“可用”迈向“可靠”和“可负担”。智能体（Agent）作为下一代智能形态，正从实验室走向商业化，成为重构“数字劳动力”的核心载体。在这个过程中，大型平台厂商与专业初创企业的分野日益清晰，前者提供稳固的生态底座，后者则凭借行业 Know-How 和 AI 原生创新，在垂直领域撕开了价值突破口。

尤为关键的是，中国市场独特的“政策刚需”为这场转型提供了坚实的压舱石。以央国企和政务为代表的 G/B 端需求，率先在金融、政务等高标准、高复杂度场景下锤炼了本土厂商的私有化部署、数据安全和工程化能力。这种自上而下的驱动力，保障了国内生态在商业化初期就具备了攻坚“硬骨头”的能力，为后续向全行业渗透奠定了高质量基础。

总之，中国大模型产业的“基座竞赛”已近尾声，“应用决战”的序幕才刚刚拉开。未来几年的竞争焦点，将不再是参数规模的比拼，而是对业务流程的深度理解、对行业数据的精深治理、以及 AI 原生 workflows 的重塑效率。对于所有入局者而言，只有深刻理解并驾驭从“技术供给”到“价值实现”的惊险一跃，才能在这场智能革命的下半场中赢得最终胜利。

## 关于爱分析

爱分析致力于成为中国领先的数字化市场专业服务平台，为企业用户提供数字化规划与落地全流程服务。

爱分析以咨询为牵引，依托实践经验、专家建议和方法论工具，为企业提供最适配的解决方案和落地服务，帮助企业提高项目成功率、优化投资收益，以及控制风险。

爱分析深耕数字化市场 10 余年，在央国企、金融、消费品、制造等行业积累诸多标杆案例和丰富实践经验。

## 产品服务

咨询

规划咨询

厂商选型咨询

业务落地咨询

研究报告

数字化实践报告

厂商全景报告

市场洞察报告

咨询匹配

需求分析

定向招募专家

workshop 交流

活动

网络研讨会

行业高峰论坛

数字化大会

企业培训

培训主题设定

爱分析专家

外部招募专家

媒体

数字化行业资讯

专家访谈

专家分享实录

## 法律声明

此报告为爱分析制作，报告中文字、图片、表格著作权为爱分析所有，部分文字、图片、表格采集于公开信息，著作权为原著者所有。未经爱分析事先书面明文批准，任何组织和个人不得更改或以任何方式传送、复印或派发此报告的材料、内容及其复印本予任何其它人。

此报告所载资料的来源及观点的出处皆被爱分析认为可靠，但爱分析不能担保其准确性或完整性，报告中的信息或所表达观点不构成投资建议，报告内容仅供参考。爱分析不对因使用此报告的材料而引致的损失而负上任何责任，除非法律法规有明确规定。客户并不能仅依靠此报告而取代行使独立判断。

北京爱分析科技有限公司 2025 版权所有。保留一切权利。

# ifenxi

咨询/合作

微 信: ifenxi888

网 址: [www.ifenxi.com](http://www.ifenxi.com)

北京市朝阳区远洋国际中心E座1906



如欲了解更多爱分析精彩洞见  
请关注我们的微信公众号

©北京爱分析科技有限公司2025版权所有



ifenxi

数字化市场专业服务平台