

规划发展部 · AI应用建设

机场AI大模型平台建设方案

四域全覆盖 · 100%私有化部署 · 信创路线（华为昇腾）

年吞吐量 7000万+

员工 1300人

IT团队 60人

三档投资方案

2027年实施

编制说明：本方案基于机场AI应用需求调研（运行/服务/安全/管理四域）编制，遵循“统一平台、分期实施、先易后难、安全可控”原则。所有报价为行业参考价（2026年），实际以政府采购比价为准。

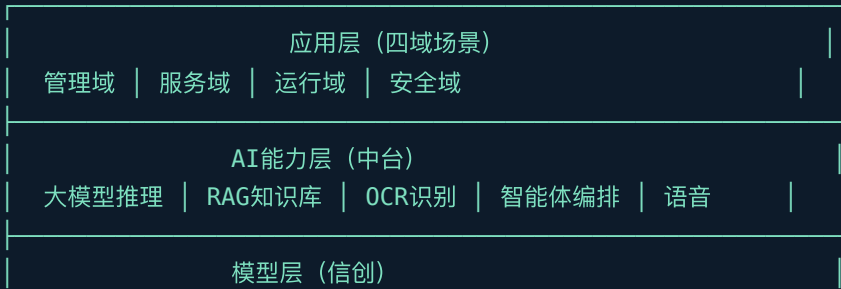
一、建设目标

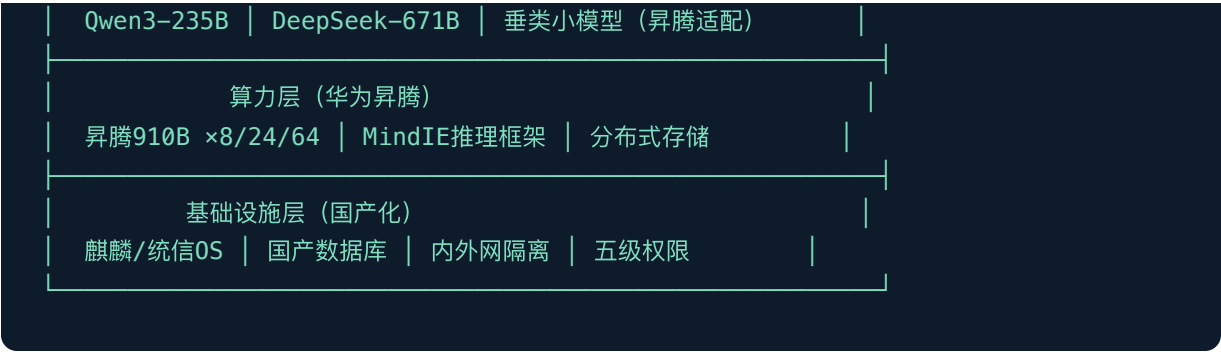
总目标：建设统一的本地化AI大模型平台，让AI进入机场全部工作流程——

- **管理域：**会议纪要、文档撰写、数据汇总、内部咨询（一期先落地）
- **服务域：**智慧客服、非航无纸化结算(OCR)、特殊旅客、商户协同
- **运行域：**航班保障协同、机坪智能巡检、资质校验、特情归集
- **安全域：**隐患识别、应急指挥、安全规章问答、态势分析

约束：① 高度私密，100%本地私有化（不允许混合云）② 信创要求（国产GPU/模型/OS）③ 网络内外网隔离、五级权限管控

二、总体架构





三、技术选型（信创路线）

层级	选型	说明
GPU算力	华为昇腾910B	国产主力AI芯片，单卡约12万，生态成熟
大模型	Qwen3-235B / DeepSeek-671B	国产开源，均有昇腾官方适配版
推理框架	华为 MindIE	昇腾原生，支持KV Cache优化
操作系统	麒麟 / 统信UOS	信创主流，等保合规
数据库	国产数据库（达梦/人大金仓）	信创要求
AI中台	厂商私有化平台 / 自建	统一模型管理、RAG、编排

四、三档投资方案

低配·起步版

配置：8卡昇腾910B | 模型 Qwen3-72B / 235B-AWQ量化 | 场景 3个 | 存储按档配置

单机部署、中等模型、支撑一期3个场景（会议纪要/知识库问答/文档撰写），先易后难验证效果。

硬件

设备	数量	单价(万)	小计(万)
华为昇腾910B GPU卡（含散热）	8张	12.0	96.0
8卡AI服务器整机（CPU/内存/电源/机箱）	1台	85.0	85.0
企业级分布式存储	100TB	0.35	35.0

100G高速交换机 + 下一代防火墙	1套	70.0	70.0
--------------------	----	------	------

软件

项目	数量	单价(万)	小计(万)
AI中台/大模型平台（基础版·私有化）	1套	120.0	120.0
国产操作系统（麒麟/统信）	7套	0.5	3.5
国产数据库	1套	20.0	20.0
场景应用开发	3个	30.0	90.0
现有系统对接集成	3个	15.0	45.0

服务

项目	数量	单价(万)	小计(万)
大模型昇腾适配与调优	1项	50.0	50.0
全员AI应用培训	60人	0.8	48.0

一次性建设投资

662 万元

年运营成本

246 万元/年

5年总拥有成本(TCO)

1895 万元

中配·推荐版  推荐

配置：24卡昇腾910B | 模型 Qwen3-235B全精度 + DeepSeek-671B(AWQ量化) | 场景 8个
| 存储按档配置

24卡集群、双模型（235B全精度+671B量化），覆盖运行/服务/安全/管理四域主要场景。

硬件

设备	数量	单价(万)	小计(万)
华为昇腾910B GPU卡（含散热）	24张	12.0	288.0
8卡AI服务器整机	3台	85.0	255.0
企业级分布式存储	200TB	0.35	70.0
100G高速交换机 + 下一代防火墙	1套	70.0	70.0

软件

项目	数量	单价(万)	小计(万)
AI中台/大模型平台（增强版）	1套	200.0	200.0
国产操作系统	10套	0.5	5.0
国产数据库	1套	20.0	20.0
场景应用开发	8个	30.0	240.0
现有系统对接集成	8个	15.0	120.0

服务

项目	数量	单价(万)	小计(万)
大模型昇腾适配与调优	1项	80.0	80.0
全员AI应用培训	60人	0.8	48.0

一次性建设投资

1396 万元

年运营成本

438 万元/年

5年总拥有成本(TCO)

3589 万元

高配·全域版

配置：64卡昇腾910B | 模型 DeepSeek-V3/R1 671B满血 + Qwen3-235B + 垂类小模型 | 场景 15个 | 存储按档配置

64卡集群、满血671B、四域全场景 + 高并发（支撑全员1300人日常办公）。

硬件

设备	数量	单价(万)	小计(万)
华为昇腾910B GPU卡（含散热）	64张	12.0	768.0
8卡AI服务器整机	8台	85.0	680.0
企业级分布式存储	500TB	0.35	175.0
100G高速交换机 + 下一代防火墙	1套	70.0	70.0

软件

项目	数量	单价(万)	小计(万)
AI中台/大模型平台（企业版）	1套	250.0	250.0
国产操作系统	15套	0.5	7.5
国产数据库	1套	20.0	20.0
场景应用开发	15个	30.0	450.0
现有系统对接集成	15个	15.0	300.0

服务

项目	数量	单价(万)	小计(万)
大模型昇腾适配与调优	1项	120.0	120.0
全员AI应用培训	60人	0.8	48.0

一次性建设投资

2888 万元

年运营成本

833 万元/年

5年总拥有成本(TCO)

7052 万元

五、运营成本分析（以中配为例）

项目	计算方式	年成本(万)	备注
电费	24kW × 8760h × 1.0元/kWh	21.0	GPU能耗约占60%
硬件维保	硬件投资 × 12%/年	82.0	厂商标准维保
AI工程师	4人 × 45万/年	180.0	⚠️ 需新增
软件/平台年费	License + 升级服务	30.0	—
厂商技术支持	服务包	40.0	7×24响应
合计	—	353	约建设投资的25%/年

⚠️ 运营成本要点：

- ① 人力是最大项（180万/年）——你们60人IT团队无AI人才，必须补齐
- ② 电费：GPU满载能耗大，24卡集群约21万/年（若电价为商业电价会更高）
- ③ 维保：按硬件12%/年，5年后硬件可能需更新（AI硬件迭代快）

六、分期实施路线（先易后难）

阶段	配置	落地场景	目标
一期 2027 Q1—Q2	低配(8卡)	会议纪要AI 知识库智能问答	管理域试点 验证效果、培养团队

		文档撰写辅助	
二期 2027 Q3—Q4	扩容至中配(24卡)	智能巡检 客服助手 安全规章问答 数据汇总分析 资质校验	四域扩展 与现有系统深度对接
三期 2028	高配(64卡)	四域全部场景 (运行/服务/安全/管理)	全域覆盖 高并发、全员使用

七、人才补齐方案 ⚠️ 关键

核心风险：IT团队60余人，但没有AI/算法人才——这是项目成败的头号风险。
大模型平台不是"买来就能用"，需要持续调优、场景开发、数据治理。

方式	内容	成本(万/年)
外部招聘	AI算法/大模型应用工程师 2—4人	90—180
内部培养	从现有IT选拔，送培+项目实战	20—30
厂商陪伴	一期由厂商提供"建设+运营"驻场服务	40—80

建议：一期"厂商主导+内部跟学"，二期起逐步自主，三期形成自有AI团队。

八、风险与建议

风险	说明	应对
人才缺口	无AI人才	招聘+培养+厂商陪伴（见七）
数据安全	机场涉密，内外网隔离	100%本地部署，五级权限，审计留痕
模型幻觉	大模型可能"胡说"	RAG知识库约束+人工复核关键输出
信创适配	昇腾生态不如英伟达成熟	选有昇腾适配案例的模型/厂商
投资效益	重资产投入	分期实施，一期验证ROI再扩

硬件迭代	AI芯片3年一换代	预留扩展槽位，避免一次性顶配
------	-----------	----------------

九、三档方案对比与推荐

档位	GPU	建设投资	年运营	5年TCO	适用
低配·起步版	8卡	662万	246万	1,895万	先验证效果
中配·推荐版 ★	24卡	1,396万	438万	3,589万	四域主要场景
高配·全域版	64卡	2,888万	833万	7,052万	全域全覆盖

推荐路径（分期投入，总投入可控）：
一期(低配662万) → 二期扩至中配(+734万) → 三期扩至高配(+1,492万)
若分期，不必一次性投2,888万——按"先易后难"，用一期的662万验证效果，效果好再逐步扩容。

若预算受限（600万内）：一期可做"极简版"—— 8卡 + 仅2个场景（会议纪要+知识库问答）+ 复用现有存储/网络，约500—550万。

💡 补充选项（第四条路）：若预算紧张、或想"先见效果再决策"，可暂不建大平台，先用几万元在 1—2 个部门做"部门级M5方案"试点（详见下章第十章），1—2周见效、数据不出内网，验证价值后再决定是否上大平台。

十、部门级轻量方案（苹果 Mac + 本地大模型）★新增

为什么加这一档？前述三档方案（662/1,396/2,888万）是全机场级的重资产投入，周期长、决策链长。而现实中，很多部门（市场、总办、规划、法务）等不起、也不需要那么大的算力——他们只需要一个"自己部门能用、数据不出门、几万元起步"的AI助手。

本方案给出第四条路：部门级轻量方案——用苹果 Mac（M5 芯片）+ 开源大模型，单台设备、1—2周上线、全程离线运行，作为大平台之前的"练兵场"，也可长期独立运行。

10.1 定位与适用场景

对比项	全机场大平台（前述三档）	部门级轻量方案（本档）
算力	昇腾910B ×8/24/64卡	苹果 M5 单机，统一内存
投资	662万 – 2,888万	约 3–6 万元
上线周期	6–18个月	1–2周
并发能力	全员1300人级	单部门 10–30人
数据安全	机房私有化	单机离线，数据物理不经网络
适用部门	全机场四域	市场 / 总办 / 规划 / 法务 等
与大平台关系	总体规划	前置试点 + 长期补充

核心优势：统一内存架构（UMA）——苹果 M 系列芯片把 CPU/GPU/内存做在同一块芯片上，**内存即显存**，可以直接把 32B–70B 参数的大模型整块加载进内存运行，**不需要像传统 PC那样买昂贵的独立显卡**。这是苹果芯片跑本地大模型的独特优势。

10.2 硬件配置与价格

设备	关键配置	参考价 (万)	说明
Mac Studio（M5 Max/Ultra）	统一内存 64–128GB	3.5–6.0	部门主力，可跑32B–70B量化模型
Mac mini（M5 Pro）	统一内存 32–64GB	1.2–2.0	性价比首选，可跑14B–32B模型
MacBook Pro（M5）	统一内存 24–48GB	2.0–3.0	便携，适合出差/移动办公场景
千兆交换机/网线（复用现有）	—	0	内网共享，无需新增

⚠ 说明：苹果 M5 芯片于 2025 年发布，采用新一代工艺，统一内存与神经网络引擎（AI算力）均有提升。**具体型号配置与价格以苹果官网实时为准**，上表为 2026 年参考区间。也可复用单位已有 Mac 设备先行试验（零采购成本）。

10.3 软件栈（全部开源免费）

软件	用途	费用	备注
Ollama / LM Studio / MLX	本地大模型推理引擎	开源免费	苹果原生优化，MLX为苹果官方框架
Qwen3-32B / DeepSeek-R1-32B 量化版	主力模型	开源免费	中文能力强，需约20—40GB内存
Open WebUI	类ChatGPT网页界面	开源免费	支持多用户、权限、聊天记录
AnythingLLM / Dify	本地知识库(RAG)	开源免费	部门文档导入，私有不出网
OCR组件（PaddleOCR本地版）	扫描件文字提取	开源免费	配合法务/档案场景

总软件成本：0 元。全部为开源项目，无需License，无需联网（模型文件本地加载），符合"数据不出内网"要求。

10.4 搭建方法（六步，1—2周完成）

步骤	操作	耗时
1. 采购/复用设备	采购 Mac Studio/mini（或复用现有 Mac），装好系统	1—3天
2. 装推理引擎	安装 Ollama（一行命令），或 LM Studio（图形界面）	0.5天
3. 拉取模型	ollama pull qwen3:32b（约20GB，需下载一次）	半天
4. 部署交互界面	装 Open WebUI，浏览器访问，类ChatGPT体验	0.5天
5. 建部门知识库	装 AnythingLLM，导入本部门制度/文档/合同范本	1—2天
6. 内网共享	设为部门内网地址，同事浏览器直接访问，配权限	0.5天

搭建后的使用形态（全离线）：

部门同事浏览器 → 内网IP:3000（Open WebUI）

|



10.5 运行费用（极低）

项目	计算方式	年费用	备注
电费	峰值约 150W × 8h/天 × 250天 × 1元/度	约 300 元	满载才150W，日常更低
软件 License	全开源	0 元	—
模型更新	本地下载，无需订阅	0 元	需人工偶尔更新
维护人力	部门内1人兼职（0.1人天/天）	约 0.5 人	无需专职AI工程师
合计	—	约 300 元/年	仅为大平台的 0.01%

对比：大平台年运营成本约 246—833 万元；部门级方案年运行成本约 300 元（电费）。这就是“重平台”与“轻工具”的差别。

10.6 部门级应用实例（真实办公场景）

部门	AI用途	效果提升
市场部	① 竞品动态日报自动汇总 ② 宣传文案/推文初稿起草 ③ 旅客投诉与舆情归类分析	文案类工作提效 50%—70%
总办/办公室	① 会议录音→纪要要点提炼 ② 公文、通知、简报起草 ③ 领导讲话稿框架搭建	纪要整理从2小时→10分钟

规划发展部	① 行业政策文件快速解读 ② 可行性报告/方案初稿 ③ 各类数据表格归纳分析	资料研读效率提升数倍
法务/合规部	① 合同条款风险点初筛 ② 法规条文检索与比对 ③ 法律文书模板生成	⚠️ 本地运行，合同不出内网
财务/采购	① 发票/单据批量OCR录入 ② 报表数据核对异常提示 ③ 采购比价资料整理	数据录入提效 60%+

实例一：市场部——竞品动态日报（原来2人2小时，现15分钟）

过去：每天人工浏览航司/机场公众号、行业网站，摘录竞品动态，整理成日报。

现在：把收集到的网页/文件丢给本地AI，提示"提炼竞品动态要点，分'航线/服务/营销'三类，各3条"→ 直接出稿，人工审核润色。

关键：全程本地，涉密市场信息不出单位。

实例二：总办——会议纪要一键生成（原来2小时，现10分钟）

过去：录音回放逐句听，手工整理，稍不留神漏掉领导要求。

现在：录音转文字（本地Whisper）→ 丢给本地AI："提炼决议事项、责任部门、完成时限，按表格输出"→ 直接生成纪要模板，人工核对即可。

收益：每次例会节省 1.5—2 小时。

实例三：规划发展部——政策文件解读（半天读懂一份文件）

过去：一份部委/行业政策文件动辄数十页，逐页啃，还要查背景。

现在：文件丢给本地AI："用大白话总结要点、列出与我们机场相关的条款、提示需要关注的变化"→ 20分钟出解析，再重点核对。

收益：政策研读从半天缩短到1小时。

实例四：法务——合同风险初筛（保密要求下的最优解）

过去：合同靠人工逐条看，容易疏漏，且绝不能上传到公网AI（泄密）。

现在：本地AI运行，"逐条标出对我方不利的条款、提示缺失的必要条款"→ 初筛后人工复核。

关键：合同文本全程在本机内存中处理，物理上不经任何外部网络——这是云端AI做不到的。

10.7 实施建议（与大平台衔接）

推荐"双轨并行"路径：

1. **现在就能做**：选 1—2 个部门（如规划发展部、总办）试点部门级M5方案， 投资几万元， 1—2周见效果，同时培训人员、积累提示词与场景经验。
2. **同步推进**：按前述三档方案编制大平台预算、走立项流程。
3. **衔接**：大平台上线后，部门级设备转为"边缘节点"或"离线保密终端"， 两者并存——**大平台管全机场通用场景，部门级管高保密/低频次场景。**

好处：① 用小钱先验证AI对机场业务的真实价值，降低大平台立项的决策风险； ② 培养第一批"AI种子用户"，为大平台储备人才； ③ 高保密部门（法务）天然适合离线单机方案。

⚠ 注意事项：① 部门级模型能力弱于大平台（32B vs 235B/671B），适合文档类、问答类任务，不适合复杂推理； ② 单机并发有限（10—30人），不适合全员使用； ③ 需与信息部门报备，纳入统一的安全与资产管理； ④ **价格与型号以苹果官网为准**，本表为2026年参考。

机场AI大模型平台建设方案 | 编制：规划发展部（AI应用调研）

⚠ 价格说明：本方案报价为2026年行业参考价，仅供预算测算，实际以政府采购招标比价为准。GPU/服务器价格波动较大（年降幅约10—15%），2027年实际采购可能更低。

编制日期：2026-10-08