

AI

多模态大模型 赋能 SMB 安防市场多元化跃迁

■ 文 / 毛宏 360 智慧生活集团

作为安防产业结构中分布最广、数量最多的群体，SMB（中小微企业）正迎来前所未有的战略窗口期。据亿欧智库测算，SMB 安防市场在视频监控领域的年规模已达 400 亿元，而在运营管理、场景服务等延展需求带动下，潜在市场空间预计超千亿。尽管项目体量小、部署分散，但通过规模叠加，极易形成高频、高覆盖的市场累积效应，成为安防企业下沉市场、复制扩张的关键战场。

近年来，海康威视、大华股份等头部企业纷纷将城市级方案轻量化、模块化后下沉至 SMB 领域，加速进入连锁商业、小型园区等场景。但在行业整体智能化转型的背景下，SMB 用户的真实需求也暴露出三大痛点：“用不起”“用不懂”“适配难”。多模态大模型的出现，为 SMB 安防市场带来一次技术路径的系统性跃迁。

一、从智能化到可服务化：多模态大模型重构 SMB 安防路径

多模态大模型的兴起，为安防行业带来了从“算法堆叠”到“能力即服务”的根本性变革。以 Transformer 为代表的架构，构建出横跨图像、语音、文本、视频等多源数据的统一语义空间，不仅实现了复杂场景的语义理解，还具备了跨模态、跨任务的泛化与迁移能力。安防 AI 由此完成从“功能模块”向“服务组件”的转型，让智能能力以 SaaS 插件的形式快速部署、轻量调用，推动智能安防迈入“模型即平台”的全新范式。

1. 泛化能力提升：重构碎片化场景的智能适配机制

SMB 用户覆盖商铺、园区、仓库、社区、教育等多元场景，不同企业在点位布局、关注事件、安防目标上千差万别。传统模型需针对每种场景独立训练，开发与交付成本高昂，难以规模化复制。而多模态大模型具备极强的零样本识别与任务迁移能力，通过一次训练可广泛适配多个场景，仅需少量样本即可实现高质量识别。

例如，CLIP 模型通过图文对比学习，构建统一的图文语义嵌入空间，使“以文搜图”“以图搜图”成为现实。用户只需输入“穿蓝色工服的男子走进配电间”，系统即可快速检索相关视频片段，极大提升查找效率。借助 OVD（开放词集检测），安防系统可在无需训练数据的前提下识别并定位从未标注过的目标类别，适配长尾事件识别，如“红色包装箱倾倒”或“陌生儿童进入禁区”等罕见场景。而由 Meta 公司发布的 SAM 模型则实现了“提示即分割”的目标识别方式，用户可通过语音、点击或文字提示快速标注任意目标区域，在安防物品监管、现场物体识别等场景中具有重要价值。

多模态大模型技术赋予了安防系统“即插即用”的泛化能力，为 SMB 用户提供了高精度、低成本、可规模复制的场景智能化方案。

2. 自然语言交互：降低系统使用门槛，释放智能价值

中小企业通常缺乏专业的 IT 与算法运营人员，传统安防系统在功能使用、策略配置方面存在明显门槛。多模态大模型通过引入自然语言交互能力，使 AI 系统从“可调用”走向“可对话”，极大降低了使用门槛。

以 360 视觉云平台为例，依托自研多模态大模型“360 智脑”，用户可通过语音或文本提出自然语言问题。如“昨天有没有陌生人经过后门”“今天早上员工几点打卡”，系统即可通过视频分析与语义推理，返回结构化答案，真正实现“看图说话”。结合 BLIP 模型的图文生成能力，系统还能基于监控画面自动生成具备语义上下文的事件描述，如“上午 10 点，一名男子在货架区域长时间逗留”，告警信息不再是简单截图或时间戳，而是完整、可理解的情境文本。

以多模态大模型为核心的“AI Agent 智能体”，还支持用户通过自然语言配置监控规则、切换摄像头视角、开启告警策略等操作，形成“说一句，系统懂并做”的交互范式，显著增强了安防系统的易用性与普及性。

3. 数据理解与主动服务：安防系统向运营助手进化

在 AI 基础能力不断普及的背景下，SMB 用户对于安防系统的期待正从“守安全”扩展到“助运营”。多模态大模型的语义建模与内容生成能力，使其不仅能“识别风险”，还能理解场景语境、输出建议方案、辅助运营优化。

通过语义聚合与图文摘要能力，系统可将分散的视频告警转化为具备业务洞察力的结构化简报。在 360 视觉云中，可自动生成每日 / 每周“安防运营报告”。如“本周顾客在饮料区驻留时间增长 18%”“出现 4 次异常徘徊行为”，帮助管理者直观掌握安全与运营动态；结合用户画像、门店类型、历史数据，系统还可实现智能推荐与策略优化；提示“工作日午间门店后门开启频率高，建议提升巡检频次”，或“促销区人流密度高于阈值，建议加强布控策略”，让安防系统从“反应机制”升级为“决

策协助”。

通过以语义理解为核心的服务能力升级，安防不再只是“看录像”的工具，而是参与企业运营决策的智能助手。

二、SMB 安防新商业模式与多模态技术带来的新业态

随着多模态大模型在安防领域的深入应用，SMB 安防的商业模式正经历从“硬件销售”向“智能服务”转型，从“项目交付型”向“平台运营型”演进，逐步催生出以 SaaS 订阅、云边协同、数据增值为特征的新业态。

1. 从硬件销售到智能服务：订阅制与云化运营的兴起

传统 SMB 安防项目普遍采取“硬件采购+本地安装”的交付方式，设备周期长、能力固定、服务链条短，用户体验与系统价值难以长期保持。在互联网化的加持下，安防 AI 能力正在实现“云端推理+持续迭代”，推动安防产品转向“能力即服务”的智能化升级。

以 360 视觉云为代表的 SaaS 平台，正在引导 SMB 用户进入“订阅+模块化”的新消费模式。用户不再需要一次性投入昂贵设备，而是可以低门槛接入基础视频服务，并按需订阅 AI 能力模块，如“AI 智能巡店”“视频告警改写”“图文事件搜索”“语音问答”等。服务可按月、按年计费，支持灵活启用与终止，不仅降低初始成本，也增强了功能更新的即时性与使用弹性。通过这种“订阅即交付”的模式，打破以往硬件绑定业务的局限，将 AI 能力商品化、插件化，为安防平台创造了新的长期收益曲线。在国外成熟市场，运营服务收入已占整体安防营收的 35% 以上，而中国当前该比例尚不足 10%，增长空间巨大。

2. 普惠 AI+ 边云协同：让智能能力触达“最后一公里”

在传统安防架构中，智能识别与语义分析能力主要集中于后端系统，依赖大量数据回传和中心化计算，造成部署复杂、响应迟缓，SMB 用户常因资源受限、技术门槛高而望而却步。而多模态大模型的引入，结合“边云协同”架构，让 AI 能力以服务化形式精准触达终端设备，实现了“低门槛部署+高智能响应”的融合范式。

360 视觉云平台率先构建了基于多模态大模型中台+前端 AI 容器化推理框架的能力架构。一旦前端设备触发特定事件（如人员驻留异常、非营业时段入侵等），系统即将关键图像帧、时空信息等结构化数据上传云端，由多模态大模型进行进一步的语义理解与复杂推理。借助图文融合、语言生成等能力，平台可自动完成事件摘要、行为判定、风险归因等高阶分析，为用户提供高度语义化、可理解的反馈。

例如，在某连锁饮品品牌门店中，用户可基于视觉云平台为摄像头配置“顾客驻留”检测技能，用于识别商品区域的关注热度。当系统检测到驻留时长异常时，会自动上传内容，通过多模态模型生成语义报告“顾客于 11:20 在促销区域驻留超 3 分钟，可能存在选购犹豫”。若进入促销期，还可自动切换为“员工离岗检测”模块，实现“同一设备，不同时间段，不同任务”的柔性智能调度。

3. 数据增值：从防损系统走向经营大脑

安防系统正从“看监控”演进为“读数据”。多模态大模型具备“跨模态理解+内容生成+语义推荐”的能力，不仅能识别安全风险，更能提供高价值的运营洞察。通过图像、文字、语音等多模态数据融合，平台可为商户生成如“顾客画像分析”“驻

留路径热力图”“员工在岗效率报表”等运营数据，并结合事件摘要能力，定期推送可视化报告，辅助企业做出陈列优化、人员排班等经营决策。

例如 360 视觉云在某仓储物流企业的应用中，通过“区域驻留+语义分析”功能发现入库区夜间高频进入情况，平台自动生成异常报告，协助管理者调整监控策略，并将风险时间段重点布控，实现从“被动告警”到“主动运营”的价值跃迁。

三、现实挑战：多模态大模型在 SMB 安防市场落地的关键难点

尽管多模态大模型为 SMB 安防市场带来了前所未有的技术跃迁与服务创新路径，但其在真实环境中的大规模落地仍面临多重现实挑战。从模型部署到生态协作，从用户认知到数据合规，每一个环节都考验着厂商的综合能力。

1. 模型部署成本高，轻量化能力尚未普及

当前多模态大模型大多基于 Transformer 等复杂结构，虽然具备优越的语义理解能力，但实际部署到 SMB 场景时仍存在较高门槛：对前端设备算力要求高，难以在成本敏感型市场普及；云端推理虽然能力强，但调用频次与数据传输会带来较高运营成本；推理速度、带宽占用、延迟管理仍需精细设计，尤其在多设备并发场景下对系统架构提出挑战。

2. 数据安全要求提升，责任闭环如何落实机制

随着安防系统日益智能化，摄像头所采集的信息正从视频片段扩展到人脸、动作、语音、语义等高敏感数据。2025 年 4 月起施行的《公共安全视频图像信息系统管理条例》，进一步提高了对视频数据采集、使用、存储、传输的合规要求，SMB 市场

的“数据责任”问题被正式推上台面。

对于设备厂商与平台服务商而言，必须在系统架构中引入全链路加密、权限分级、日志审计等合规机制，同时向 SMB 用户提供可视化、低门槛的隐私管理工具，确保“谁用谁负责”的合规逻辑落实落细，但在实际落地中也存在诸多难点。

3. 多模态标注体系缺失，制约模型泛化能力提升

多模态大模型的性能依赖于大量跨模态训练数据（图像、文本、语音等），但在 SMB 安防场景中，采集数据往往呈现出分布零散、类型复杂、标签不标准等典型“长尾”特征：不同场景缺少统一语义标准，如“徘徊”“打架”“离岗”行为在各行业定义不同；视频标签主要集中在通用类别，缺乏贴合业务语义的“软标签”；小企业缺少数据整理能力，也缺乏意愿参与数据回流与优化。

如何构建覆盖多行业、多场景、多模态的标注规范体系，形成企业、平台、模型之间的“闭环数据流”，是打破智能化瓶颈的关键。

四、发展与展望：走向 AI 原生的智能安防新范式

未来 3 到 5 年，SMB 安防市场将加速迈向“AI 原生架构”，以端侧 AI 协同云端多模态大模型为核心，推动从“设备驱动”向“智能服务”转型。通过动态 AI 加载订阅与平台化交付机制，企业可按需加载 AI 技能，实现低成本、高适配的服务部署。结合千人千面的智能通知与语义理解能力，安防系统将从被动记录进化为主动洞察，并通过建立“视觉大模型+SaaS 平台+技能生态”的新优势体系，抢占全球智能安防创新高地。🔥