

面向行业的“云+AI” 应用研究报告

—AI行业云平台助力行业大模型应用落地

2026年1月

前 言

行业大模型正成为推动产业智能化升级的关键突破口。当前，人工智能正从技术探索迈向产业落地的关键阶段，随着“人工智能+”行动的深入推进，AI不再仅是实验室中的算法模型，而是驱动千行百业智能化升级的核心引擎。这一过程的实现不仅要求AI具备强大的通用能力，更强调其在具体行业场景中的专业性、可靠性和可落地性。由此，能够深度融合行业知识、理解业务逻辑、支撑关键决策的行业大模型正是在这一需求驱动下，逐渐演变为支撑行业智能化升级的关键引擎。

云计算正从资源供给平台，跃升为AI创新的核心引擎。行业大模型的发展并非孤立的技术演进，而是高度依赖于底层基础设施与生态体系的协同支撑。在此过程中云平台不仅提供弹性可扩展的算力资源，支撑大模型训练与推理的高成本、高复杂度需求；更重要的是，行业云平台天然汇聚了长期沉淀的业务数据、流程规范、知识图谱与合规框架，是行业大模型的专业化提供了不可或缺的土壤。同时，云原生架构、多租户安全隔离、数据主权保障以及与ERP、MES、CRM等企业级系统的集成能力，使得AI能力能够真正嵌入实际业务流，实现从“演示效果”到“生产可用”的跨越。

“云+AI”的融合，正在催生一种全新的产业智能范式：以AI行业云平台为底座，行业大模型作为智能引擎，通过平台化方式实现技术、数据、应用与商业价值的闭环。这种模式不仅解决了AI落地中常见的碎片化、孤岛化问题，更通过持续的数据回流、模型迭代与场景反馈，构建起可进化、可运营的智能系统，行业智能化的价值正通过“云+AI”的融合效应得以规模化释放。

本报告聚焦“云+AI”深度融合的发展趋势，系统性梳理如何以 **AI** 行业云平台作为支撑行业大模型可持续发展的核心载体。通过分析行业云在高质量数据供给、可信运行环境、生态协同机制和商业化路径等方面的独特优势，深入探讨其如何赋能行业大模型实现从技术研发到业务价值转化的全生命周期管理。同时，结合气象、制造、医疗、能源等领域的典型实践，展望“行业云+AI”协同演进的技术路径与产业图景，旨在为各行业推进智能化转型提供兼具前瞻性与实操性的参考框架。

目 录

一、AI 行业云平台是 AI 在行业侧应用的最佳载体与价值放大器	1
(一) 数据基础	1
(二) 安全可信	2
(三) 应用生态	3
(四) 能力封装	4
(五) 开发支撑	4
二、云+AI 典型应用实践	5
(一) 智能审方助手赋能药店零售审核	5
(二) 基于大数据边缘云平台的多模态充电桩智能规划系统	11
(三) 5G+云+AI 安全在钢铁行业应用	15
(四) AI 赋能气象 MaaS 服务创新应用	23
(五) 电网计量问答与设备运维智能体	29
(六) “本安数斯”能源生产安全卫士	36
(七) 基于现货市场机制智能云	44
(八) 东莞市人工智能大模型中心	49

一、AI 行业云平台是 AI 在行业侧应用的最佳载体与价值放大器

行业大模型的规模化落地，离不开与行业场景深度融合的基础设施支撑。在通用大模型技术日趋成熟的背景下，如何让 AI 真正落地到行业一线，解决实际业务问题，已成为产业智能化的核心命题。单纯的模型能力提升已不足以应对复杂的行业需求，需依托一个能够连接数据、知识、应用与用户的综合性平台。行业云平台正是在这一需求驱动下，逐渐演变为支撑行业大模型发展的关键基础设施。

AI 行业云平台是实现“云+AI”赋能行业的关键载体。行业云是由行业龙头企业或区域政府单位统一建设，聚焦领先的实践经验和解决方案，以解决垂直行业或特定区域内需求的云平台，为行业内企业、集团内子单位、区域内企业提供 IaaS、PaaS、SaaS 等云上产品和服务。AI 行业云平台是在行业云平台基础上，新增对算力、数据集、模型、开发工具链等人工智能关键要素的深度支持能力，打造面向 AI 应用的综合性平台。由于 AI 行业云平台集成了行业专属的数据资源、业务系统、安全规范与服务生态，相较于通用云平台，AI 行业云平台更强调对行业特性的深度理解与支持，具备更强的专业性、合规性与集成能力。正是这些特性，使其成为行业大模型落地的理想载体。

（一）数据基础：AI 行业云平台提供高质量、可治理的训练“燃料”

行业云平台的核心优势在于其沉淀了高价值、低噪声的行业特性数据。数据是大模型训练的“燃料”，行业大模型所需的并非泛化的互联网语料，而是高度专业化、结构化、场景化的行业数据。AI 行业云平台建设、运营方通常是行业内头部企业或地方政府出资的专业公司，长期服务于特定行业客户，积累了大量真实、动态、闭环的业务

数据，如金融交易记录、工业设备运行日志、医疗电子病历、政务审批流程等。这些数据具有高价值、低噪声、强关联等特点，是训练具备行业理解能力模型的稀缺资源。

AI 行业云平台内嵌了大量由行业专家、标准规范和历史经验凝练而成的结构化知识。例如设备维修手册、监管合规条款、药品相互作用图谱、政策法规库等。这些知识以规则引擎、知识图谱、标签体系或向量数据库等形式存在，可提升其在专业任务中的准确性、可解释性与合规性。例如，当大模型回答“某类贷款是否符合监管要求”时，若能实时检索并引用最新的银保监会文件条款，其输出将远比仅靠语言模式推测的结果更具权威性和实用性。

AI 行业云平台构建的闭环数据资产体系，是行业大模型实现深度业务赋能的根本基础。行业云平台所提供的是一套完整、闭环、可治理的业务数据资产体系——它既包含“发生了什么”的事实记录，也包含“为什么这么做”“应该怎么做”的专业判断，构成了行业大模型区别于通用模型的核心竞争力，也是其实现深度业务赋能的根本基础。

（二）安全合规：AI 行业云平台构筑合规可控的运行环境

AI 行业云平台的安全机制确保用户在享受大模型能力的同时，满足数据不出域等核心合规需求。在金融、医疗、政务等对安全要求极高的行业，数据隐私与系统可控性是 AI 应用的前提。行业云平台通常遵循严格的行业监管标准，具备数据不出域、本地化部署、审计追溯等能力，为大模型的运行提供了可信环境。

通过私有化或混合部署架构，AI 行业云平台为高敏行业提供符合监管要求的安全运行环境。通过构建“私有化大模型”或“混合部署”模式，企业可在保障核心数据安全的同时，享受大模型带来的智能化服务。例如，某大型金融机构采用私有化部署方式，在内部数据中心

运行专属的大模型，确保客户敏感信息不外泄，同时利用行业云提供的丰富 API 接口实现智能客服、风险评估等功能，满足银保监会的数据安全要求。这种“安全优先”的架构设计，是通用云服务难以替代的核心优势。

AI 行业云平台支持细粒度权限控制与全链路审计，保障业务安全。 AI 行业云平台还支持细粒度的权限管理和动态访问控制策略，确保不同角色只能访问与其职责相关的数据和功能。结合加密技术与日志审计机制，可以全方位保障数据及应用在整个生命周期内的安全性与合规性。

（三）应用生态：AI 行业云平台打通 AI 能力与业务系统的“最后一公里”

AI 行业云平台依托现有 SaaS 和业务系统生态，可有效嵌入业务流程。 大模型的价值最终体现在业务场景的赋能上。行业云平台往往已集成大量 SaaS 应用与业务系统（如 ERP），形成了成熟的行业应用生态。通过开放 API、SDK 或低代码工具，大模型的能力可以快速嵌入现有 workflow，实现大模型应用落地的无缝集成。

AI 行业云平台通过嵌入式 AI 模式，显著降低智能化应用的使用门槛。 例如，在客服系统中接入智能应答引擎，不仅能够自动解答常见问题，还能根据用户行为推荐个性化解决方案；在审批流程中嵌入合规审查助手，自动识别潜在风险并提供改进建议；在生产调度中引入预测性维护建议，提前预警设备故障，减少非计划停机时间。这种“嵌入式 AI”模式大幅降低了 AI 应用的使用门槛，加速了智能化在组织内的普及。

AI 行业云平台的统一架构促进跨单位、跨部门的 AI 能力共享与协同，释放规模化效应。 AI 行业云平台通过统一的身份认证体系与

微服务架构，实现了跨系统、跨部门的高效协同。无论是集团内子单位还是区域内企业，都能基于同一套基础设施共享 AI 能力，形成规模化效应。

（四）能力封装：AI 行业云平台的运营体系促进 AI 能力产品化

AI 行业云平台将大模型能力产品化，形成可持续的商业模式。

行业大模型的可持续发展离不开清晰的商业模式。行业云平台通常已具备成熟的用户体系、计费机制与服务体系，可将大模型能力封装为 AI 产品，按需订阅、按量计费，实现商业化闭环。

AI 行业云平台通过用户反馈与使用数据驱动模型持续迭代，形成“使用—优化”正向循环。以某省级工业 AI 行业云平台为例，该平台将预测性维护大模型作为 SaaS 服务向中小企业开放，企业可根据实际需求灵活选择订阅周期和服务套餐，极大降低了技术门槛与成本负担。同时，基于平台的用户反馈与使用数据，还可以持续优化模型性能，形成“使用—反馈—迭代”的正向循环，推动模型在真实场景中不断进化，真正实现“可持续运营”。

AI 行业云平台的多租户运营体系兼顾安全隔离与资源共享，加速产业级智能应用推广。多租户运营体系还支持不同规模、不同行业的用户在同一平台上独立运作，既保证了数据隔离与安全，又促进了资源共享与经验交流。这种模式不仅提高了资源利用率，也加快了新技术、新应用的推广速度，助力整个产业链的数字化转型。

（五）开发支撑：AI 行业云平台降低智能应用开发门槛

AI 行业云平台预置行业基座模型与场景模板，避免重复开发，提升 AI 落地效率。行业大模型要实现快速迭代与规模化落地，离不开高效、专业且贴合业务场景的开发支撑体系。AI 行业云平台在服务大量客户的过程中，逐步沉淀并集成了覆盖数据处理、模型训练、

评估优化到部署运维的 AI 开发工具链，显著降低行业用户的 AI 应用门槛。同时，平台通常预置了经过验证的行业大模型基座和典型场景模板，用户可在此基础上快速定制专属模型，避免重复造轮子。

AI 行业云平台实现开发、数据、安全与应用的天然打通，提供开箱即用的 AI 工程化能力。开发工具与平台内的业务数据、安全策略和 SaaS 应用天然打通，开发者可在不离开平台的情况下完成“数据—模型—应用”的端到端闭环。例如，某 AI 行业云平台推出 AI 开发套件能力，支持开发者通过配置化方式在 5 分钟内完成一个业务型 AI Agent 的搭建——用户只需选择大模型引擎、上传内部知识库、开启联网或文件处理等扩展能力，即可一键发布至小程序或 H5 应用。

AI 行业云平台为行业大模型落地所提供的不仅是算力资源，更是一套面向行业用户的、开箱即用的 AI 工程化能力，有效解决了“有数据不会训、有模型难上线、有应用不可维”等现实痛点，为行业大模型的持续演进和广泛应用提供了坚实的技术底座。下一章将深入探讨基于行业云的“行业云+AI”典型应用场景与实践案例。

二、云+AI 在各行业的典型应用实践

（一）智能审方助手赋能药店零售审核

1、案例概述

国药数字科技（北京）有限公司（国药数科）作为中国医药集团有限公司（国药集团）的全资子公司积极落实集团工作部署，研发“AI 审方智能体”，以四川省为试点省份，充分调研 DTP 药房购药审核及结算业务流程，结合四川省医保局对单行支付药品和高值药品的管理政策、国家医保局智能监管“两库”规则，采用大模型提取用药规则及医保报销规则，构建审方知识库，同时一线药师发挥专业经验，积

极通过真实业务数据反复校验优化模型及智能体效果，为项目提供精准“业务大脑”；技术团队聚焦一线真实痛点，突破复杂医疗资料快速准确识别、复杂知识库自动化构建、复杂业务规则智能审核等技术难题，并综合利用 AI 辅助业务流程编排、机器人流程自动化等先进技术，打造“辅助审方智能体”，实现患者资料“一键归档”、合规要求“一键审核”、业务需求“一键开单”。

2、需求分析

随着国家医保政策的调整，“双通道”药品面临国家医保的严格监管。特药医保业务经办流程繁琐，本身相关知识复杂，需审核的资料多且专业。同时专业药房业务量级大，巨大的处方量给药师带来沉重的审方负担，审方效率低且易出错，影响患者用药安全，并且超量超适应症使用未有效拦截，会造成医保基金滥用、损失。因此，需依托算法模型精准识别用药风险，为 DTP 药房实现一键智能化审方功能和一键开单功能，提升一线药师的审核效率，并确保药师审方的质控水平持续、稳定在较高水平。

系统以云+AI 技术为核心，深度融合医保政策与药品审核规则，实现了从“人工经验审方”向“智能审方”的转型。主要满足了三个方面的需求：

1) 对于药房：系统推动行业从人工审核向智能审核转型，其“规则库构建-智能解析-一键开单”全流程方案，从传统人工走向智能，节省的不只是时间，更是患者的焦虑和药师的专业价值，效率的倍增，极大缓解了药师工作压力，降低了因人工疏漏导致的用药与报销风险，让药师得以专注于更复杂的临床用药指导。合规的不止是满足医保基金监管要求，有效防范超量、超适应症等医保违规行为，更重要的是帮助患者在合规前提下更好地享受医保政策带来的红利，助力医保基

金的合理使用与监管闭环。另外，数字化档案管理模式解决了药房特药资料存储难、调阅慢的行业共性问题。

2) 对监管机构：系统实现“合理用药+医保合规”双重质控，构建了智能的“防火墙”，实现了从“被动处罚”到“主动防控”的转变，为精细化监管提供了工具，推动行业提升审方标准化水平，助力行业应对医保飞检常态化趋势。

3) 对行业：为同类药房提供可复用的数字化路径，推动药房的服务数字化升级，提升了药房运营效率与服务标准化水平。成功探索了“业务+技术”深度融合的数字化转型路径，打造了鲜活可复制的样板，为进一步丰富人工智能在医药零售领域应用范围、探索智慧药店模式，继续在守护医保基金安全、提升“三医”协同效能上发挥示范作用。

3、项目实施情况

系统于 2025 年 8 月 18 日上线，系统和算力均部署在国药云上，所有数据在国药自有 IT 基础设施及算力上进行传递和计算，建设严格的数据隔离及保护机制，确保数据安全。

系统在国控四川 6 家试点 SPS+药房落地，支持 508 种特药审核，审方效率提升 70%，准确率超过 90%，极大地提升了一线药师的审方效率，实现合理用药+医保合规双重质控，推动审核从“事后核查”向“事前预警”转型，满足医保基金监管要求，帮助患者在合规前提下更好地享受医保政策带来的红利。

AI 辅助审方智能体的构建与运行，依赖于国药云平台的强大支撑。平台不仅为 AI 应用提供了坚实的基础设施保障，还在数据管理、安全保障以及运维支持等多个层面，构建了完整的技术服务体系，确保 AI 应用能够高效、稳定、安全地运行，从而真正发挥其在医药领

域的价值。



图 1 国药云平台

基础设施支持方面，国药云平台基于先进的云计算架构，为 AI 辅助审方系统提供了强大而灵活的计算资源、存储能力和网络性能。在 AI 应用中，尤其是在模型训练、推理和实时审方过程中，往往需要处理海量的数据和复杂的计算任务，这对底层硬件资源提出了极高的要求。国药云平台通过高效的计算资源调度和分配机制，能够灵活应对这些高负载场景，保障 AI 系统的高性能运行。此外，平台采用高可用架构设计和云资源池化管理，不仅提升了资源利用率，也极大地增强了系统的稳定性与可靠性。为了保障全国范围内用户的顺畅接入，国药云还采用了 SD-WAN 技术，并结合多运营商接入方案，实现了跨地域、跨网络的高速、低延迟访问，确保各地药房和医疗机构都能高效使用该 AI 系统。

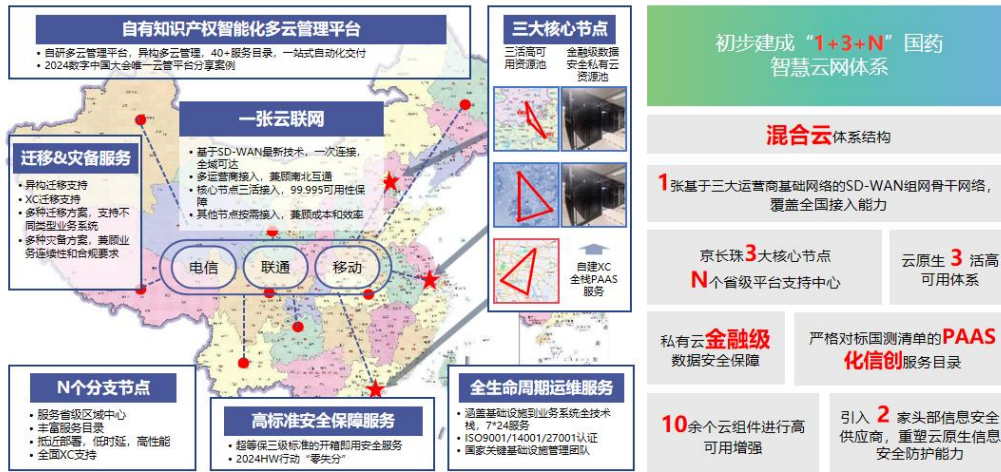


图 2 1+3+N 国药智慧云网体系

弹性与扩展性方面，AI 应用常需要处理大量数据，云平台的弹性伸缩特性能够根据计算需求动态调整资源，保障 AI 应用在负载高峰期仍能稳定运行，确保服务的高可用性。

云平台与应用场景结合方面，结合传统 OCR 与大模型的多模态信息处理能力，实现对非标准化资料的智能识别与语义理解；引入更多样化的专业模型，混合模型交替识别，关键信息互相校验的方案；为提升准确率审方要求，在识别后增加语义精确识别环节，显著提升知识命中准确率和结果准确率；并基于 AI 等综合手段动态规划审方业务流程，减少不必要的审方环节，提升并行度。算力总体消耗较少，并且总算力未增加的情况下，通过系统工程化方法，将系统并发能力提升一倍，信息识别准确率提升约 50%。

数据管理与安全方面，AI 应用依赖于大量数据进行训练和持续优化，云平台提供了高效的存储解决方案和数据保护机制。国药云平台通过统一的安全管理体系、多层次的数据加密技术，以及严格的访问控制策略，保障数据在云端的安全与合规。云平台支持数据的快速备份、灾难恢复及高效的故障处理，确保 AI 应用数据的安全性和业务连续性。



图 3 集团网络安全&数据分级保护体系

服务与运维方面，云平台为 AI 应用提供了完整的生命周期管理，从应用的开发、测试、部署到运维。国药云通过智能运维平台，结合监控、告警和自动化管理，保障 AI 应用的持续稳定运行。

4、场景效益

AI 辅助审方智能体不仅带来了效率的提升，更从合规、人力、体验多维度优化产业生态，是人工智能与行业结合的典型案例。

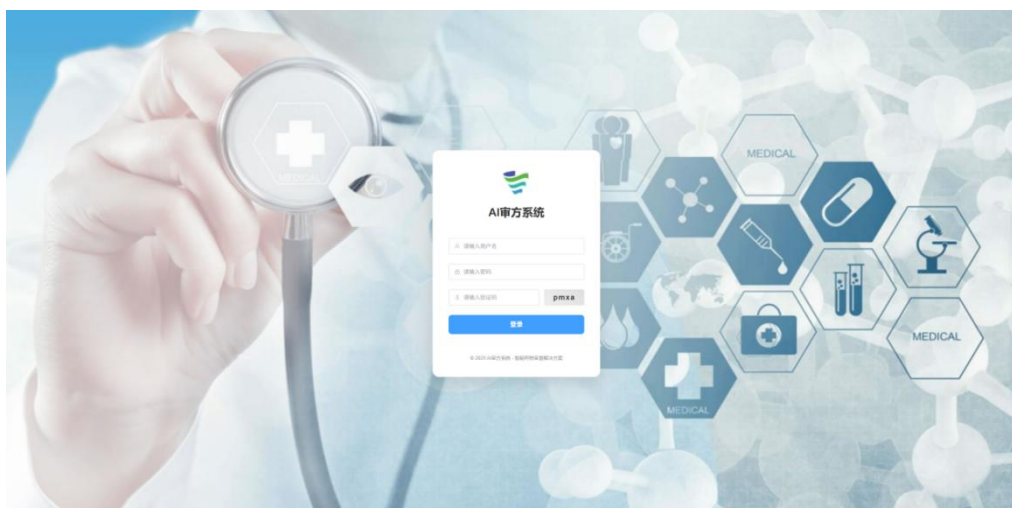


图 4 AI 辅助审方智能体

1) 效率提升，聚焦专业服务

单处方全流程（审方+开单）时间从 10 分钟压缩至 3 分钟，智能审方环节仅需 40-50 秒，效率提升超 70%；系统稳定支撑日均超 100 单审核，累计完成 2700 余单，适配高业务量级需求。

截止目前完成 508 个品种的审核规则入库，说明书及医保规则录

入合计 1422 条，特药录入完成率达到 98%，通过大量全面的规则录入，大幅减轻药师重复录入、规则核对的工作，让药师聚焦用药指导等核心专业服务，人力效能显著释放。

2) 合规质控，精准拦截降低风险

辅助审方结合药师人工复核后，问题处方有效拦截准确率超过 90%，AI 准确率持续提升，更加稳定的质控水平，更加多样化的审核视角，实现合理用药与医保合规双重质控。通过智能规则库精准匹配医保政策与药品说明书，避免人工审核疏漏导致的医保拒付、违规处罚等风险，满足医保飞检常态化监管要求，为药房合规运营筑牢防线。在 6 家药房的近 2 个月的试点过程中，已协助药师完成 2000 多笔订单的审核，成功拦截了 17 笔不合规报销，涉及医保基金 5 万元。

3) 运营优化与体验升级

数字化档案管理实现患者资料安全存储与快速调取，解决特药资料存储难、调阅慢的行业痛点，目前已归档电子档案数 7000+；“一键审方+自动填单”简化操作流程，降低药师专业门槛与学习成本。对患者而言，审方周期

(二) 基于大数据边缘云平台的多模态充电桩智能规划系统

1、案例概述

为破解行业难题，中国移动通信集团广西有限公司为广西电网有限责任公司、南方电网广西电动汽车服务有限公司打造“基于大数据边缘云平台的多模态充电桩智能规划系统”，构建“云边协同、云智一体”的解决方案：以大数据边缘云平台为统一基座，实现“人-车-桩-网”多源数据实时汇聚、治理，为上层 AI 智能体提供高质量的数据供给。同时，采用“大模型+AI 智能体”双引擎架构，将复杂的时空计算与业务分析能力封装为智能服务，用户通过自然语言交互，即可一站

式获得从需求预测、站址推荐到运营优化的全链路决策支持。

2、需求分析

随着新能源汽车产业迅猛发展，充电需求呈现爆发式增长。国家能源局数据显示，2025年国庆假期，全国高速公路新能源汽车充电量创历史新高，达到12287.32万千瓦时，日均充电量1535.91万千瓦时，较去年同期增长45.73%。然而，当前充电桩布局普遍存在供需空间错配、数据要素割裂、潮汐需求响应滞后等问题。传统选址依赖人工经验与静态数据，难以实现多源数据的实时融合与智能分析，导致资源闲置与用户“找桩难、排队久”现象并存，制约了充电网络效率与服务体验。

当前充电设施行业面临三大核心痛点，对智能化升级提出迫切需求：

1) 规划维度单一，科学依据不足

充电设施选址决策高度依赖土地可用性、电力接入条件等静态物理因素，以及有限的局部历史充电数据，缺乏对动态人流、车流、区域功能演变等多维度时空规律的融合分析，导致布局与真实需求脱节，“盲目建设”与“充电荒漠”现象共存。

2) 数据价值沉睡，协同壁垒高筑

影响选址决策的关键数据（如电网实时负荷、车辆行驶轨迹、商业热力、人口流动等）分属电力、交通、通信、规划等多个系统，标准不一、难以互通，形成牢固的“数据烟囱”。数据要素无法流动与聚合，其潜在的乘数效应无法释放，制约了整体决策水平的提升。

3) 运营手段僵化，难以动态适应

充电需求具有显著的潮汐性、时段性等特征。传统运营缺乏基于实时数据感知和预测的智能调度能力，无法实现价格弹性调节、资源

柔性分配与用户引导，导致设施利用率波动巨大，运营商收益不稳，用户体验不佳。

因此，亟需一个能够打通数据壁垒、承载智能算法、并支持敏捷迭代的一体化平台。本项目旨在构建一个以行业云平台为数字底座、以多模态 AI 为核心引擎的新型解决方案，实现对充电设施“规-建-运”全生命周期的数据驱动与智能决策，推动行业从“经验驱动”向“智能驱动”的范式变革。

3、项目实施情况

本项目构建以大数据边缘云平台为核心基座，贯穿“数据-模型-能力-应用”全链路架构，通过云边协同机制，将底层海量数据与上层 AI 智能体深度融合，形成从数据感知到智能决策的完整闭环。架构图如下：

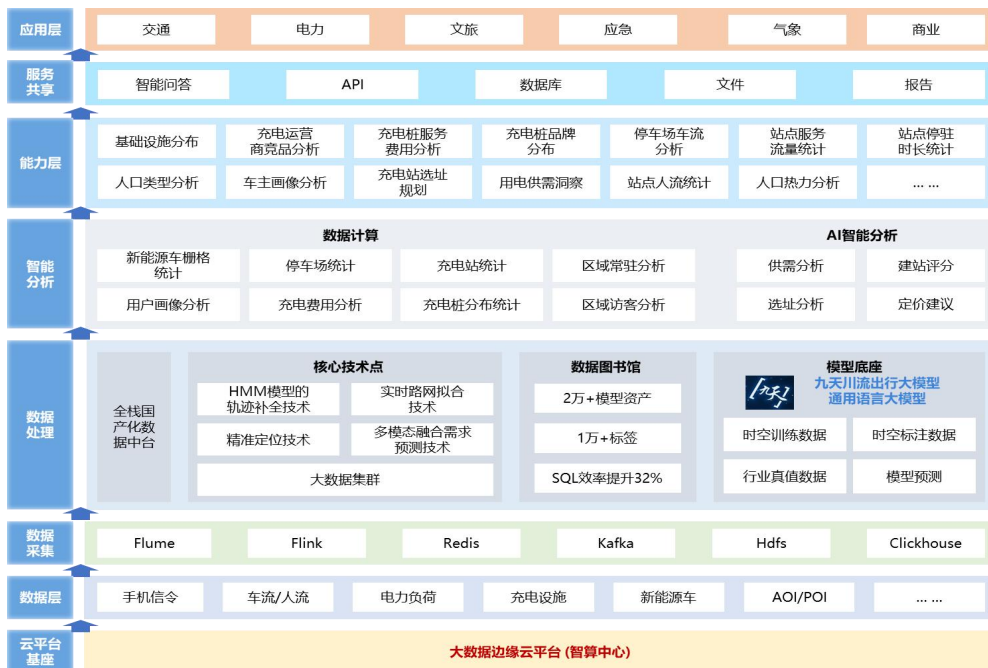


图 5 项目架构图

本项目中，行业云平台与 AI 应用的深度融合主要体现在以下三个方面，云平台为 AI 应用提供了坚实底座和核心支撑：

1) 云平台构建了 AI 应用的统一数据基座

项目基于“N+X”资源池构建了智算中心，通过分布式架构实现对运营商信令、电网负荷、车辆轨迹、地理信息等多源异构数据的规模化接入、治理与融合。云平台采用 Flume+Kafka+Flink 的实时处理架构，对海量数据进行清洗、对齐和标准化，形成高质量、标准化的数据供给，彻底解决了传统模式下数据孤岛、数据质量不高的难题，为上层 AI 模型的精准训练和推理提供了可靠的数据保障。

2) 云平台提供了 AI 全生命周期管理能力

平台采用“集中训练、分布推理”的云边协同架构。在云端，利用强大的算力资源完成 15 亿参数出行大模型的集中训练与迭代优化；在边缘侧，通过部署轻量化推理引擎，实现低延迟的智能决策服务。云平台的统一管控与弹性调度机制，可根据业务负载动态分配计算、存储资源，保障了 AI 应用的高性能、高可用和弹性扩展，实现了 AI 模型从开发、训练到部署、运维的全流程高效管理。

3) 云平台实现了 AI 能力的标准化服务化赋能

平台将复杂的 AI 算法模型（如轨迹补全、需求预测、选址评分等）封装为标准的 API 服务接口，并通过构建自然语言交互界面，极大地降低了 AI 技术的使用门槛。无论是充电设施的选址分析，还是应急管理的预警推送，均可通过统一的服务门户按需调用 AI 能力。这种服务化的赋能模式，使得智能决策能力能够快速、低成本地赋能于电力、应急、文旅等多个行业场景，实现了“一基座多场景”的乘数效应。

4、场景效益

本项目通过 AI 与云平台的深度融合，在成本、效率、价值和资产盘活等维度取得了显著成效：

1) 成本控制：通过 AI 智能决策替代传统人工选址与数据分析，

每年可节省 3 人选址团队人力成本约 30 万元，并削减外部商业数据服务费用约 50 万元，实现年化综合成本节约 80 万元。

2) 效率提升：依托出行大模型，实现了选址策略的自动化生成，将传统需要 30 天以上的复杂选址流程缩短至 3 天内完成，效率提升超过 90%。试点充电站平均利用率提升 18%，用户排队等候时长下降 23%。

3) 价值创造：通过实时数据监测与智能调度，优化了充电桩运营策略，提升了运营商收益。同时，将沉淀的数据要素复用于应急、文旅等领域，开辟了新的收入增长点，累计带动经济效益 1961.98 万元。

4) 资产盘活：基于统一的数据底座和能力平台，项目沉淀的高价值数据资源在交通管理、应急预警、文旅洞察等场景中被重复调用，跨行业数据复用率高达 80%，支撑了项目的快速规模化应用。

(三) 5G+云+AI 安全在钢铁行业应用

1、案例概述

由中国移动通信集团湖南有限公司（以下简称“湖南移动”）联合湘钢、华为、启明星辰等合作伙伴，共同打造了面向钢铁行业的“5G+云+AI 安全”工业互联网安全运营体系，主要面向钢铁行业智能制造转型中的新型安全挑战，针对终端接入风险、传输安全威胁、事件响应滞后等痛点。项目深度融合 5G 传输加密、MEC 边缘安全防护、工业协议深度解析等技术，构建了集智能监测、预警、处置、溯源于一体的闭环安全防护系统。

2、需求分析

当前，钢铁行业在数字化转型过程中面临严峻的新型安全挑战。传统安全方案难以适应工业互联网环境，主要表现为三大痛点：一是

工业协议与设备高度复杂，涉及 PLC、天车等专用设备及多种工业协议，传统规则引擎覆盖率不足 40%，无法深度解析异构流量；二是威胁响应效率低下，平均事件处置时间超过 10 分钟，误报率高达 30%，高度依赖人工研判；三是企业安全原子能力缺失，尤其广大中小企业缺乏可弹性部署、持续运营的安全基础设施。

针对上述痛点，项目团队开展了系统性需求分析，重点聚焦四大典型场景：MEC/CRM 等办公网面临虚拟化边界模糊、数据泄露与非法控制风险；无人天车/远程 VR 等 5G 应用存在数据窃取、恶意操控及网络覆盖稳定性问题；工业控制网络需应对外部攻击、内部漏洞及供应链威胁；企业普遍缺乏全天候安全运营能力，导致运营目标不清、流程不完善、风险不可控。基于此，项目规划并实施了“云-边-端”协同、AI 驱动的安全运营体系。核心路径涵盖终端接入认证、MEC 边缘安全防护、传输加密、数据安全防护及态势感知系统建设，并将云端原子能力（态势感知、僵尸蠕虫拦截等）下沉至工业现场。

3、项目实施情况

5G+云+AI 安全在钢铁行业应用平台为湘潭钢铁提供工业资产管理、漏洞风险、安全事件、威胁风险、数据识别、数据泄露监测、AI+安全大模型等能力，服务企业的工业智能制造、工业供应链安全、工业+AI 大模型等场景。



图 6 安全运营平台

1) 打造“安全管家”产品

由信安部牵头，协同信息化支撑中心、网维中心、信息技术中心，

建立网络安全赋能工作机制，打造标准化安全产品，2024年4月已完成。

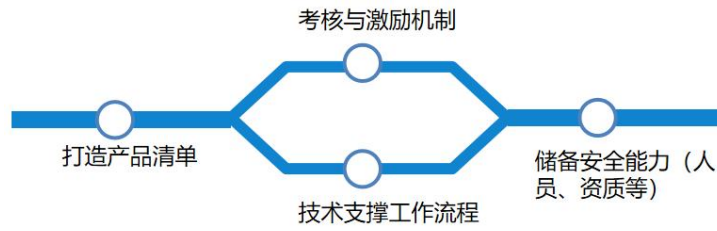


图 7 网络安全赋能工作机制

2) 制定资费和套餐

由政企部牵头，协同云能，明确目标客户，制定资费价格，打通计收流程，2024年4月已完成。

套餐	建议套餐价格	使用群体	服务内容	交付内容
安全运营托管 (基础级)	49000元/年	客群分析：基础安全服务，有安全需求，业务系统不大复杂不高的党政机关、医院、中小学、企业等全行业单位均可。	提供200Mbps流量或10个资产内一年的安全托管检测与响应服务，服务内容包括日志分析、每季度一次的漏洞扫描及远程加固指导、告警远程封堵服务。通过微信服务群或电话方式进行安全威胁监测通告。	《安全运营报告》、《漏洞扫描报告》、《漏洞复合报告》。
安全运营托管 (专家级)	66000元/年	客群分析：专家安全服务，客户面更广，有安全需求，业务系统有一定规模，预算充足的党政、医疗、教育、国企等行业单位均可。	提供500Mbps流量或20个资产内一年的安全托管检测与响应服务，服务内容包括日志分析、每月一次的漏洞扫描及远程加固指导、告警远程封堵服务、每半年一次风险评估服务、包含一次应急响应服务。通过微信服务群或电话方式进行安全威胁监测通告。	《安全运营报告》、《漏洞扫描报告》、《漏洞复合报告》、《风险评估报告》、《安全加固建议》、《应急处置结果报告》。
安全运营托管 (保姆级)	97125元/年	客群分析：保姆级安全服务，针对顶层客户，有安全需求且对业务数据本地化有要求，业务系统有足够规模，预算充足的省市级党政机关、医疗、教育、国企等行业单位均可。	提供1Gbps流量或50个资产内一年的安全托管检测与响应服务，服务内容包括日志分析、每月一次的漏洞扫描及远程加固指导、告警远程封堵服务、每月一次现场安全巡检并协助风险加固、每半年一次风险评估服务、应急响应服务。通过微信服务群或电话方式进行安全威胁监测通告。	《安全运营报告》、《现场巡检报告》、《漏洞扫描报告》、《漏洞复合报告》、《风险评估报告》、《安全加固建议》、《应急处置结果报告》。

图 8 套餐与资费

3) 运营平台打造

由信安部牵头，协同启明星辰、云能打造安全运营托管服务能力底座，2024年6月已完成。

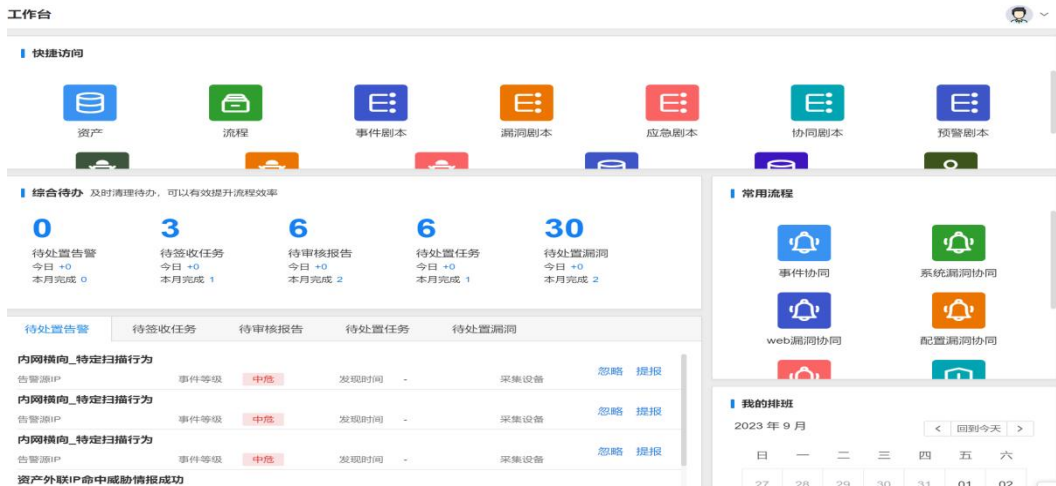


图 9 运营平台

4) 做好安全运营保障

由信安部牵头，协同信息化支撑中心、网维中心、信息技术中心，组织开展网络安全运营服务产品交付，组织省级专家做好二线售前、售中、售后支撑服务。



图 10 态势中心

5) 系统核心功能

① 资产管理。依托自动化的互联网资产探测平台，基于域名或 IP 地址信息采用系统探测、应用探测、端口服务探测、WEB 探测等多种技术对暴露于互联网的资产进行主动发现，获取其服务的主机、系统、应用、框架、中间件、数据库等信息，通过资产数据采集器与数据合并关联工具整理后，系统性的可视化生成详细的资产信息列表，包含域名、IP、端口，中间件(如: weblogic、nginx、Apachelog4j 等)、应用(如: OA、邮件、财务等系统、网络设备、视频监控服务或者非

WEB 类的 ES、redis、mongodb、ldap 等)、框架(Springboot、strust2、Thinkphp、indexCMS、JEECMS...)等信息，自动全面的帮助构建暴露于互联网侧的资产与风险台帐。

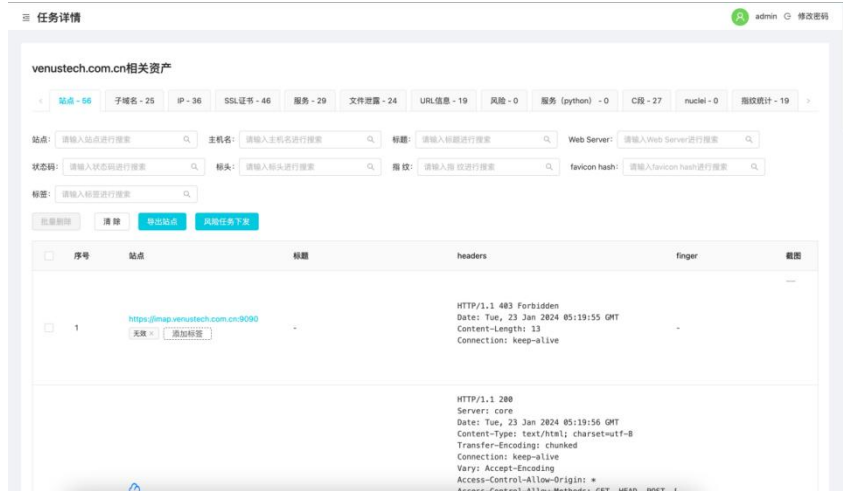


图 11 资产管理概览



图 12 资产管理详情

② 漏洞风险。通过专业的漏洞扫描工具，对应用系统进行非破坏性的漏洞检测，识别操作系统、数据库、中间件上可能存在的安全漏洞，发现系统面临的安全威胁，并输出详细的漏洞检测报告，报告中包括对应的漏洞修复建议，并在漏洞修复完成后进行复测，验证漏洞是否修复成功。

③ 安全事件。监测安全运营范围内发生发现的安全事件，并提供相应的处置、跟踪等必要的响应措施，确保安全事件的完整闭环。

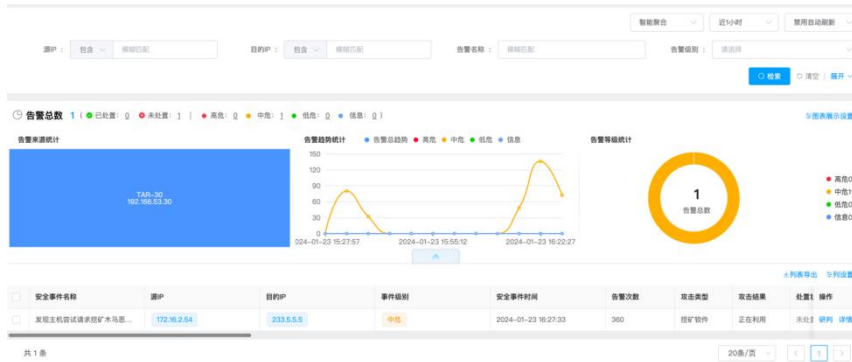


图 13 安全事件

对采集到的安全数据进行安全分析，挖掘安全潜在风险，协助客户快速发现与定位安全事件，了解安全威胁与影响。



图 14 风险研判

④ 威胁风险。将威胁情报信息（如恶意 ip 地址、恶意 URL、恶意域名等）与安全事件特定的信息（如源地址、目的地址、请求域名、载荷内容等）进行实时关联分析实现威胁情报信息获取到威胁情报分析利用，为湘潭钢铁企业提供安全管理提供纵深防御能力。

⑤ 数据识别。对数据资产有全盘的掌握，针对业务系统复杂、数据资产分散、内容不详、风险不清、无法检测等问题，基于资产测绘引擎通过主动+被动相结合进行多维数据资产属性测绘，明确管理对象，梳理数据资产目录，对数据进行分级分类管理，同时对数据的

流动进行监管测绘，从而完善数据通信双方信息及构建数据资产业务、应用和风险的上下文关联信息，摸清数据资产家底，为数据安全监管提供基础资料。

⑥ 数据泄露监测。基于流量、日志、数据包、文件等进行分析，结合数据安全治理场景的监测分析引擎，挖掘敏感数据、重要数据跨境等安全事件。系统支持数据安全分析，内置数据安全知识库，包含个人普通信息、个人敏感信息、用户相关数据；单位/企业自身相关数据、研发域数据、生产域数据、基础属性类数据、征信、气象、医疗卫生、地理、摄像头等方向。同时，支持 MYSQL/ORACL/SQLSERVER/HIVE/PostgreSQL 等操作审计、异常连接、爆破分析等数据库风险的监控。

针对明文分析引擎通过对明文信息中敏感数据、医疗影像/生物特征类数据等内容的识别，对境外代理访问行为、敏感信息批量/高频传输等行为的监测，从而防止明文数据泄露，保障数据安全。

⑦ AI+安全大模型。AI+安全大模型通过引入人工智能技术，实现对网络安全风险的智能化识别、分析和响应。

安全大模型基于海量工业领域安全数据和先进的算法模型，对网络流量、系统日志、应用行为等进行实时监测和分析，快速识别出潜在的安全威胁。同时，通过自动化响应机制，能够在发现威胁后迅速采取隔离、阻断、告警等措施，有效遏制威胁扩散。

4、场景效益

① 生产效率显著提升。建成 5G+MEC 尊享网络：2.6G+4.9G 混合组网，双 MEC 部署，实现数据不出园区，端到端业务时延从 26ms 降至 9.9ms，单服务器转发能力达到 15Gbps，满足低时延、大带宽和安全性需求。通过 AI 安全大模型的智能研判与自动化响应，安全事

件平均处置时间从超过 10 分钟缩短至 3 分钟以内，响应效率提升 300%，极大减少了因安全威胁可能导致的生产中断时间。同时，平台实现了对 428 台工业设备和 21 种专用协议的实时深度监测，将传统安全方案的覆盖盲区降低了 80%，累计成功处置异常事件 9263 次，有效保障了生产系统的连续、稳定运行。

② 运营成本大幅降低。项目为企业带来了直接且可量化的经济效益，年均节省运维成本达 450 万元。这主要得益于云化部署模式降低了 70% 的初始硬件投资，以及 AI 自动化运营减少了对 12 名高技能安全人力的依赖，人力成本节约约 350 万元。此外，该方案已创造直接订单合同超 900 万元，展现出良好的市场回报与商业可持续性。

③ 安全防护水平实现质变。智能化应用使安全运营的精准度与广度发生了根本性改变。平台通过 AI 算法对海量告警进行智能降噪，将误报率从 30% 以上降低至 10% 以下，有效释放了运维精力。同时，其对 150 多种工业协议的深度解析能力，将协议覆盖与理解率从不足 40% 提升至 95% 以上，实现了对 APT 攻击、勒索软件等新型威胁的精准识别和有效防御。

④ 系统可扩展性与标准化程度高。该方案已形成可快速复制的标准化产品与服务。新项目的部署周期从传统的数月缩短至数周，效率提升 80% 以上。其“云-边-端”协同架构与模块化设计，已成功适配并推广至三一重工、港口、电力等 10 余个跨行业项目，防护设备超 1 万台，验证了解决方案强大的行业普适性与规模化推广能力。

⑤ 社会与行业效益显著。项目超越了企业自身的经济账，产生

了广泛的积极影响。它通过构筑可靠的安全底座，保障了钢铁等关键基础工业的生产连续性，间接守护了产业链与民生稳定。同时，该模式以运营商云端服务的普惠方式，显著降低了广大工业企业，尤其是中小企业获得高水平安全防护的技术门槛与成本，为整个工业领域的数字化转型提供了安全护航，起到了重要的行业示范与引领作用。

（四） AI 赋能气象 MaaS 服务创新应用

1、案例概述

联通（上海）产业互联网有限公司联合上海市气象局、上海市经信委、上海人工智能实验室以及上海联通等企事业单位共同打造集人工智能气象应用技术创新、产业孵化以及国际合作于一体的综合性平台-气象 AI-MaaS 平台。该平台以“强基赋能·生态共创”为核心，构建了“强基础设施+大能力平台+高质量运营”的创新体系，通过融合算网基础设施、畅通气象语料资源、整合大模型基础组件，打通了人工智能产业上下游，实现“人工智能+气象”服务一站式开发应用，赋能产业生态共建，加速“人工智能+气象”服务于各行业场景应用。

2、需求分析

全球气候变化背景下，重点产业对上海天气预报服务能力提出新的需求，上海市高温热浪、超强台风等极端天气气候事件总体呈上升趋势，气温显著增高。近 150 年来，上海平均气温增温速率为 0.18°C/10 年。同时，这 150 年间，上海暴雨日数和雨量显著增多，降水极端性增强。这些包括暴雨、台风、高温在内的极端天气气候事件，对能源、航运、航空等行业的安全运行带来更大的挑战，对精细化天气预报服务业务能力提出了更高的需求。随着深度学习技术和大规模人工智能模型的快速发展，人工智能在各个行业中的应用需求

迅速增加，尤其是在数据量庞大、预测复杂的场景中展示了显著优势。

气象预测作为多变量、高复杂度的领域，传统预测模型在处理海量气象数据时存在精度瓶颈和数据处理性能的限制。相比之下，基于深度学习的大规模气象模型，通过对历史气象数据的训练和分析，可以识别出复杂的气象模式，实现对未来天气的高精度预测，为提供满足能源、航运、航空等行业智能化调度和精准预测需求的预报与服务提供了新的技术手段。基于以上需求背景，产生以下两方面总体业务需求：

一是构建人工智能赋能的气象核心预报能力体系。包括构建多模态气象数据集以支撑人工智能气象预报模型训练需求；构建人工智能预报算法库以提升无缝隙网络预报精准度；研制新一代端到端人工智能预报模型体系以构建人工智能预报新技术。

二是构建人工智能与行业融合的气象智能服务新生态。包括开展大语言模型垂类应用，研制生成式气象服务大模型，构建打造面向产业气象服务需求的个性化智能体、能够学习理解和编制业务流程的业务助手和助手配置训练平台；通过隐私计算和区块链技术实现气象与行业大数据融合，完成从气象要素预报到行业应用的延伸，充分发挥气象数据的趋利避害效益。

3、项目实施情况

本案例以融合架构计算系统为平台，以数据为资源，能够以强大算力驱动人工智能模型来对数据进行深度加工，源源不断产生各种智慧计算服务，并通过网络以算力服务的形式向气象人工智能服务需求进行供应。基于联通云双引擎底座打造 AI 算力平台，将云的技术积累转化为 AI 算力平台的优势。主要由 AI 硬件基础设施、AI 软件基础设施构成，并由联通云进行统一运维管理和安全防护。

底层搭建人工智能框架、训推一体的公共服务算力基础设施，以联通云为能力底座融合高性能计算集群与边缘计算节点，通过智能算力调度平台实现资源动态分配。

中层利用联通数据管理、MaaS 两大中台能力打造气象大模型底座，整合卫星、雷达、数值模式等多源数据，构建“雨师”、“扶摇”等高精度预报模型；

上层利用人工智能应用服务平台，结合气象行业专项场景，开发行业化应用模块，实现专业气象数据价值最大化，赋能气象行业场景应用，打造算力在气象行业应用新标杆。



图 15 人工智能+气象技术架构

技术路径以下三个层级：

1) 气象 MaaS 底层支撑体系

作为 AI 赋能气象服务的底层支撑，整合万卡规模智算集群、超高速互联网络与 AI 原生存储架构，构建“算力 - 网络 - 存储”三位一体的基础保障体系。

核心依托联通临港智算集群，以高性能裸金属与随云部署资源为核心，实现 80% 的并行计算加速比，通过统一基座与专属隔离设计，为气象大模型训推提供高规格智算供给；网络层面采用 IB / RoCEv

2 方案构建算内高速无损网络，1.6Tbps 集群内带宽与多平面隔离设计破解通信瓶颈，算间通过专线与 5G 连接保障跨域数据传输；存储采用 AI 原生热存储与 NAS 温冷存储结合的架构，搭配二级缓存加速机制，实现海量多源数据毫秒级响应、安全存取与高效检索，为 AI 模型训练与推理筑牢硬件根基。

2) 气象 AI 算力智能调度体系

围绕气象 AI 应用的动态需求，构建全域算力的智能协同调度能力，实现算力资源的最优配置与价值最大化。基于联通“星罗”调度平台，构建“通算 + 智算 + 超算”融合调度能力，实现多元异构资源的高效协同与智能分配。平台统一纳管 CPU / GPU / NPU 等异构芯片，支持千卡至万卡规模集群管理，通过多租户隔离与故障快速定位保障运行稳定性；针对气象 AI 训推差异化需求，提供时延优先、算力优先等多策略调度，搭配数据加速、计算框架加速等增效能力，显著提升算力利用率（优化约 30%）；创新“中训边推”模式，依托 200 + 骨干云池与边缘一体机资源，实现模型定点推送与一键多点分发，结合多层次数据管理与跨域预加载功能，打通算存协同链路，为气象 AI 应用提供弹性、高效、可扩展的算力调度支撑。



图 16 算力调度平台功能特点

3) 气象 MaaS 核心能力底座平台

作为连接底层算力与上层服务的核心枢纽，该平台以“数据 - 平台 - 模型”三位一体架构赋能气象 AI 创新。首先平台聚焦能源、航空等重点领域，整合卫星、雷达、数值预报等多源异构数据，构建高质量、标准化的人工智能气象数据集，奠定模型训练的数据基础；其次搭建双核心平台——MaaS 平台提供一站式大模型开发应用工具链，结合智能体任务编排降低开发门槛，数通平台基于可信数据空间机制，实现数据可信管控与价值共创；最后研发分层级模型体系，既部署通用大模型，又针对性开发“雨师”“风鸟”等气象专业模型，提升灾害性天气精准预报能力，同时结合行业需求打造垂直大模型，形成“数据夯实基础、平台降低门槛、模型赋能场景”的完整支撑链，为上层行业应用提供核心 AI 能力输出。

本案例基于开展“人工智能+气象”应用攻关研究，创新气象大模型应用场景建设，依托算力、数据、MaaS 三大中台能力，在高性能智算集群及算法平台支撑下，结合上海联通云网算安数用一体化能力，通过算力网络将算力、数据、算法、应用进行全要素连接，输出“雨师”、“扶摇”等气象算法大模型，助力气象场景应用，业务流程如下：

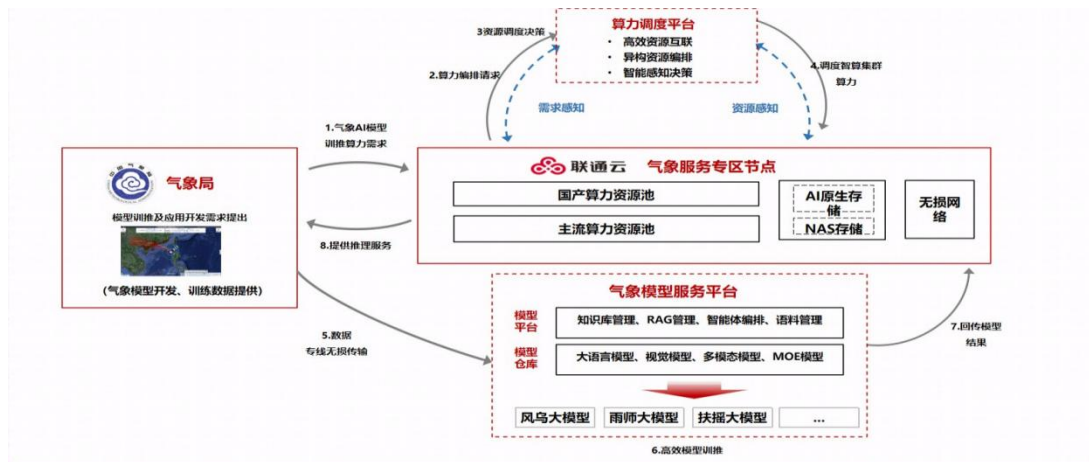


图 17 气象创新应用业务流程图

1) 气象局用户向联通云气象专区发送气象 AI 模型训练算力需求

2)气象专区接收训练需求后，向算力调度服务平台发送训练算力编排请求。

3)算力调度平台接收需求，根据现网的资源情况进行决策，调度气象服务专区算力进行训练

4)气象专区接到调度指令后，分配训练计算资源，向气象局算力节点发送传送训练数据请求

5)气象局节点响应请求，通过专用接口和高速互联专线将训练所需要的数据传送给云上气象服务专区

6)气象专区接收数据完成，基于气象服务平台，和气象数据集，开展高效模型开发及训练推理任务，孵化气象行业大模型

7)模型训练完成，气象专区将训练后的模型回传给气象局节点

8)气象局节点基于模型结果进行推理服务，使用推理结果开展气象创新应用。

经过上述流程后，可成功满足气象大模型训练的全栈服务需求，训练时长缩短数倍，效率得到了极大提升。基于气象模型服务平台，可实现行业数据库的快速构建，多种气象行业智能体打造，智能化服务行业大模型孵化。

4、场景效益

1) 精准服务驱动行业效率跃升

项目通过算力网络实时融合风云卫星、雷达、自动站等多源数据，在强对流预测精度与预报更新频次上实现双重升级。在强对流预报方面，“扶摇”“雨师”实现观测资料驱动的端到端推理，推理时间仅5至10秒，支持10分钟级高频更新。2025年13次局地强对流过程中，强对流预通报平均提前4小时20分钟，暴雨预警准确率达89.7%，暴雨蓝色预警时间提前47.7分钟。

2) 智慧气象筑牢城市全域治理屏障

城市治理的智慧网络中，精细化管理气象先知系统如同神经中枢，已实现市、区、街镇三级服务全覆盖。在进博会、F1中国赛、ATP1000网球大师赛等重大活动中，本项目孵化“运动气象指数”，实现分钟级监测保障到位。在航空航运领域，“上天”、“入海”双线发力，国内试飞气象保障自主可控，低空飞行气象服务拓展创新，远洋气象保障立体网络逐步完善。

3) 嵌入式服务赋能行业纵深

项目以“通用气象 AI 能力 + 行业专用场景”为核心，在与申能集团的合作中开创“气象智能芯片”模式，将传统气象数据产品升级为可嵌入式 AI 组件，标志着气象 AI 技术实现了从通用工具到核心组件的飞跃，全面落地“通专融合”的核心理念。“气象智能工具箱”直接嵌入集团能源保供 AI 智能体，深度适配新能源功率预测、燃气负荷调度等能源行业专用业务场景，为关键业务决策提供实时、精准的气象智能支持。

（五）电网计量问答与设备运维智能体

1、案例概述

广东电网有限责任公司是国有大型骨干企业中国南方电网有限责任公司的全资子公司，在传统电力计量系统智能化升级面临数据价值难释放等挑战，而通用大模型因行业知识适配差等问题无法直接应用的背景下，构建电力计量领域专属知识库与行业大模型，研发问答与运维智能体，实现复杂口语化问题精准应答及运维模式从被动响应到主动预警的转变，提升故障定位准确率并减少人工复核工作量。实现方式是以“大模型+业务智能体”为核心架构，依托国产化垂直大模

型构建智能底座。先通过图像文字识别增强技术提取多格式文档信息，结合语义向量转化与检索引擎构建多模态计量知识库；再经有监督微调、偏好对齐等策略优化模型，提升领域适配性。基于自然语言交互技术开发数据库智能问答功能，结合故障树分析实现设备诊断与运维决策。采用全国产化技术栈，通过分层解耦架构集成现有业务系统，实现“知识-决策-行动”完整闭环。

2、需求分析

传统电力计量系统作为电网“感知神经”与“数据基石”，智能化升级深陷多重困境。每日 10TB 级电能量数据超 70%未有效挖掘，数据价值难以释放；专业知识散落于纸质文档、独立系统等多载体，复用率不足 5%，新员工需数月才能掌握业务逻辑；故障排查依赖人工经验，诊断准确率仅 75%，平均处理时长超 8 小时；多子系统缺乏统一整合机制，单次任务响应时间超 60 秒。同时通用大模型存在行业知识“水土不服”、数据隐私保护难、国产算力适配不足等问题，无法直接服务电力计量场景，成为行业智能化转型的核心瓶颈。

基于行业痛点，项目聚焦数据利用、知识管理、故障诊断及系统协同四大核心需求。核心目标是构建电力计量垂直大模型与知识库，实现复杂查询精确率超 95%、知识复用率提升至 98%以上，突破多模态业务推理技术，将推理延迟降至 15ms 以内；开发计量问答智能体，支持口语化指令输入，单次任务响应时间不超 60 秒，推动专家知识高效传承；打造“AI 指挥官”运维智能体，将故障识别准确率提升至 90%以上，人工复核工作量降低 85%；建立实时故障预测与健康管理机制，提升故障处置效率 6 倍以上，增强电网运行可靠性。

3、项目实施情况

“电网计量问答与运维智能体”项目项目依托广东计量云平台的

底层支撑，破解了电力计量领域数据多源异构与知识碎片化难题，推动业务模式向智能化转型，为能源行业智能化升级提供了可复制的实践范式：

1) 广东计量云平台与本项目 AI 应用的核心结合

广东计量云平台与电网计量问答与运维智能体融合形成“数据-知识-模型-应用”协同体系，核心体现在五大关键领域。

多模态计量知识库构建融合。项目借助广东计量云平台的多源数据整合能力，通过 OCR 增强技术与自定义 QALoader 工具，从平台汇聚的多格式业务文档中精准提取文本、公式及表格内容；依托 Embedding 技术将文本语义转化为高维向量，结合 Proxima 向量检索引擎构建向量数据库，整合计量专业文本、研判规则与业务思维导图，形成系统化知识资产，为 AI 应用筑牢基础。

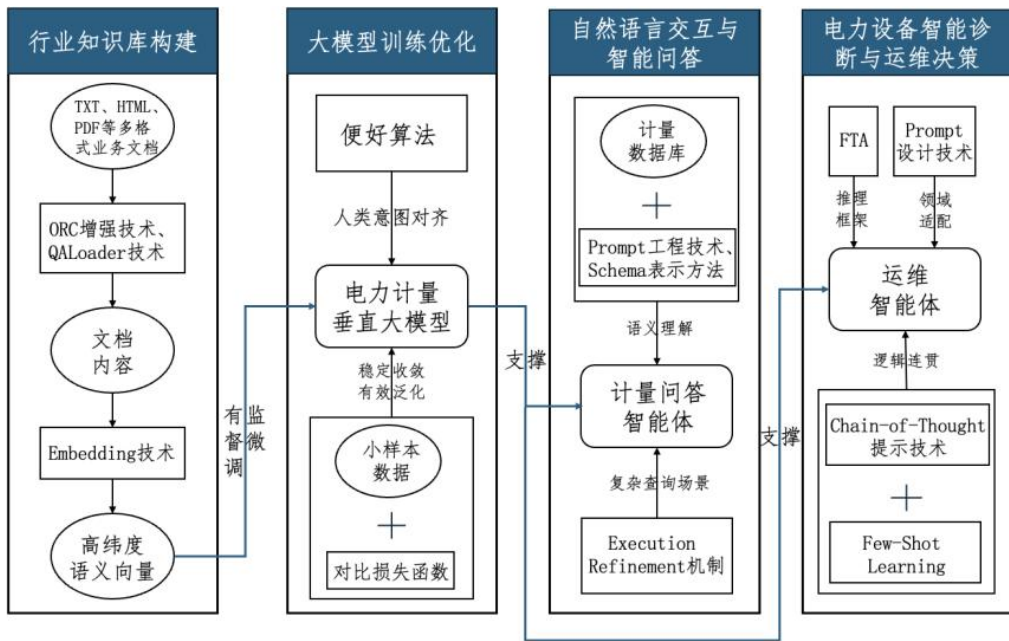


图 18 技术路线

电力计量垂直大模型训练优化与算力适配融合。依托平台国产化算力资源，项目完成基座大模型的有监督微调、对比损失函数优化、直接偏好算法对齐等全流程训练；平台适配海光、阿里云国产芯片及推理框架，通过优化计算架构与推理流程，在保障模型训练与推理性

能的同时，有效降低部署运维成本，确保大模型在计量领域的适应性与准确性。

计量数据库自然语言交互智能问答融合。基于平台整合的 12 万个分库分表异构数据源，项目开发 NL2SQL 智能问答系统；通过语义增强 Prompt 工程、半结构化 Schema 表示及双向 Schema Linking 机制提升模型对数据库语义的理解能力，实现高精度表列召回；结合 Execution Refinement 机制修正 SQL 语法与逻辑错误，大幅降低业务人员使用门槛，提升数据查询效率。



图 19 电网计量自动化系统

电力设备智能诊断与运维决策融合。依托平台打通的 SCADA、DMS、工单管理等业务系统数据，项目构建基于故障树分析（FTA）的智能诊断系统；通过建立多层次故障树、融合时序故障概率模型实现设备状态动态评估；结合 Few-Shot Learning 与 Chain-of-Thought 提示技术，整合历史工单与典型故障案例，提升模型故障根因推理能力，为预测性维护决策提供有力支撑。



图 20 计量设备 AI 指挥官

国产化技术适配与业务系统集成融合。广东计量云平台作为核心枢纽，实现 AI 模型与电力现有业务系统无缝对接，打通数据壁垒；平台采用国产化技术栈，集成自研 OP-VLLM 推理加速框架与阿里云 PPU 芯片，保障 AI 应用自主可控，形成完整业务闭环。

2) 广东计量云平台对本项目 AI 应用的支撑逻辑

广东计量云平台以分层解耦、云原生化的架构设计，为 AI 应用提供全链路支撑，构建起“数据层-知识层-模型层-应用层”的四级支撑体系：

在数据层，平台依托广东电网计量自动化 3.0 系统，整合档案、采集、工单等异构数据源，通过统一语义层建模管理，解决了 AI 应用的数据来源分散问题，为模型训练、知识构建提供高质量数据输入。

在知识层，平台搭建多模态知识库与向量数据库双引擎，通过 Proxima 向量引擎与 Flink 实时计算框架，支撑 PB 级知识库的构建与毫秒级检索，实现知识的存储、索引与实时更新，为 AI 应用提供高效知识检索与调用能力。

在模型层，平台集成模型训练、微调、压缩和推理服务，通过自

研 OP-VLLM 框架对外提供高并发、低延迟的模型 API 服务，适配国产化算力环境，满足 AI 应用的高性能计算需求，保障模型稳定运行。

在应用层，平台构建智能问答和 AI 指挥官两大核心应用入口，提供自然语言交互、可视化报告和智能决策看板等一站式服务，实现 AI 技术与业务场景的深度融合，推动 AI 能力向业务价值转化。

综上，项目通过广东计量云平台的分层支撑与 AI 技术的深度渗透，实现了电力计量业务的智能化升级。平台的国产化架构与数据整合能力，为 AI 应用提供了安全可靠的运行环境与数据支撑；而 AI 技术则赋予平台智能分析、决策优化的核心能力，两者协同构建了电力计量领域智能化转型的全新模式。

4、场景效益

1) 首创电力计量垂直大模型技术，基于 PB 级别数据库，构建计量结构化与多模态知识库，利用七百二十亿级参数语义网络，研发“大语言模型+参数蒸馏”智能体模型框架，复杂查询精准率达 95.2%、知识复用率提升至 98.5%，形成计量领域可自进化的智能基座。

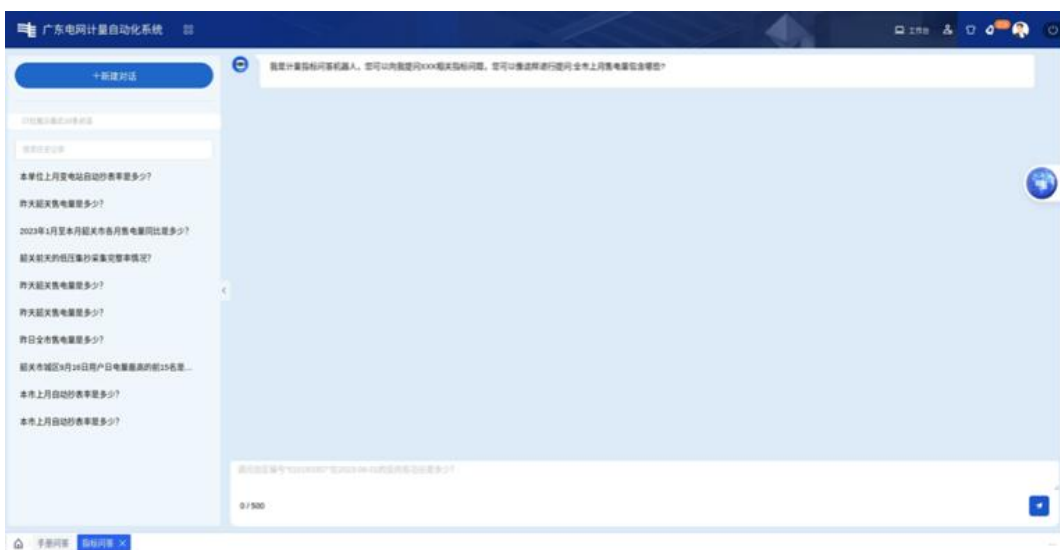


图 21 电力计量智能问答

2) 电力计量多模态业务推理加速技术，提出了低秩微调与知识

蒸馏技术，提升了大模型推理效率，推理延迟降至 15ms 以下，支持电力计量的高并发多模态业务。

3) AI 计量智能问数推理技术，设计了多模态动态知识路由引擎，实现自然语言指令与系统操作的精准映射，单次任务时长压缩至 60 秒，指令匹配准确率 95%以上。

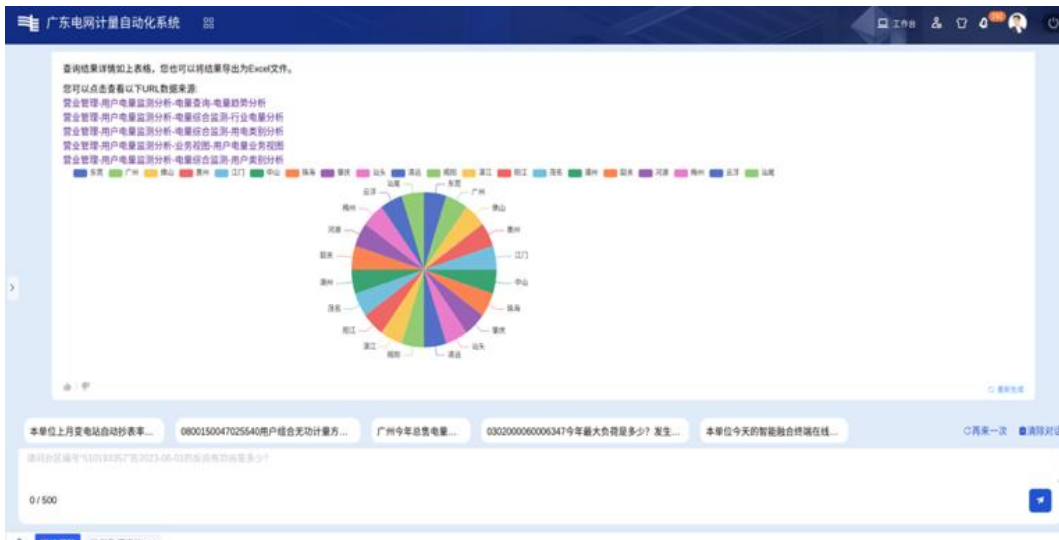


图 22 电力计量智能问数

4) 打计量设备 AI 指挥官，对电力计量运维进行智能决策，通过多智能体协同与强化学习相结合，优化了设备故障诊断与巡检流程，提升任务处理效率 12 倍，标注成本降低 80%，可在 1 小时内完成广东电网 5000 万只电表故障诊断，全面智能生成抢修策略。



图 23 计量设备 AI 运维

5) 合作与推广，该技术应用已推广至广西、海南等电网公司，支撑变电、配电、营销等领域知识库升级，设备远程“问诊”效率提升40%，巡检故障诊断准确率提升至91.2%。与浙江正泰电器股份有限公司合作，将“计量 AI 大模型”融入设备运维系统，实现故障类型自动判断、处置方案生成及诊断报告输出，终端运维检修智能化水平提升30%，为新能源设备管理提供决策支持。

（六）“本安数斯”能源生产安全卫士

1、案例概述

杭州华电能源工程有限公司研发的“本安数斯”系统，打造覆盖生产与建设全链路的安全多模态大模型。以多模态感知技术为核心，融合图像、视频、文本、传感器等多源数据，创新采用“大小模型”融合互补与“云-边”协同部署架构，云端承担模型训练与全局优化，边缘端实现实时响应与本地化分析，构建“知识服务+个性化学习+智能预警”一体化平台，形成“数据采集-智能识别-人工复核-整改追踪”全要素闭环管控流程。

2、需求分析

1) 传统场景面临以下三类痛点：

① 安全检查痛点：传统人工巡检在输卸煤、汽轮机运转层等复杂场景中效率低下、漏检率高；现有视频监控智能化不足，无法主动识别隐患；储能、氢能等新兴领域缺乏成熟风险识别标准与技术体系，安全防控处于被动状态。

② 人力资源痛点：安全巡查覆盖设备运行、人员操作、环境监测等多个维度，任务繁重导致工作人员压力较大；行业安全管控人才梯队建设不完善，新老员工技能水平差异显著，部分员工缺乏系统安全知识与应急处置能力，安全管理执行效果参差不齐。

③ 技术应用痛点：现有安全管理系统依赖单一识别技术，在高粉尘、强光/弱光交替等复杂工况下易出现误报、漏报，难以实时精准预警；安全相关法规标准、事故案例等文档数量庞大，工作人员无法快速获取所需知识，影响隐患处置与技能提升。

2) 衍生出的五类需求：

基于行业痛点，发电企业迫切需要具备“全域感知、智能预警、知识赋能、闭环管控”四大核心能力的一体化解决方案：

① 多源数据融合需求：打破数据孤岛实现跨模态协同分析；

② 精准高效预警需求，在复杂场景下保障高识别准确率与低误报率，实现毫秒级响应；

③ 知识服务需求，提供便捷检索与个性化学习功能，快速提升员工安全技能；

④ 全流程管控需求，构建从隐患识别到整改追踪的完整闭环，确保管理可审计、可追溯；

⑤ 满足灵活部署需求，适配不同发电场景，兼顾实时性与数据安全性，契合国家安全生产智能化建设强制要求。

3、项目实施情况

“本安数斯”依托“云-边”协同架构，实现复杂场景高效部署与毫秒级响应，降低人工成本与安全风险，提升设备可靠性与巡检质效，同时提供可规模化复制的本质安全解决方案，助力“人工智能+”与能源行业安全领域深度融合。

1) 云平台提供算力支撑，保障 AI 模型高效运行

① 云端算力集群搭建：以华为昇腾 910B 为核心算力支撑，在云平台上构建高性能分布式算力集群，专门用于多模态大模型的训练与微调。集群通过 RoCE 高速计算网络实现节点间低延迟通信，搭配

大容量高速存储网络，满足 1081 项法规标准、海量监控视频、传感器数据等多源异构数据的并行处理与模型训练需求，为大模型高效迭代提供坚实基础。

② 算力动态调度机制：采用 K8s 容器编排技术与 CANN 计算架构，构建灵活高效的算力调度体系。根据模型训练任务优先级、数据处理量及边缘端推理请求压力，动态分配算力资源，实现算力负载均衡；针对大模型训练资源消耗大、持续时间长的特点，采用弹性伸缩策略，训练高峰期自动扩容算力，空闲期释放冗余资源，大幅提升算力利用效率。

③ 跨层级算力协同支撑：云端算力集群承担多模态基座模型的训练、优化与版本管理，处理复杂的跨模态数据融合推理与全局策略优化任务；边缘端部署华为 300I Duo、寒武纪 MLU370 等信创推理设备，运行经蒸馏压缩的 7B 轻量化模型，负责本地实时数据处理与快速响应。云端与边缘端通过算力协同，既满足大模型高精度训练的算力需求，又保障边缘端实时预警的低延迟要求，形成“云端强算力支撑+边缘端高实时响应”的算力互补格局。

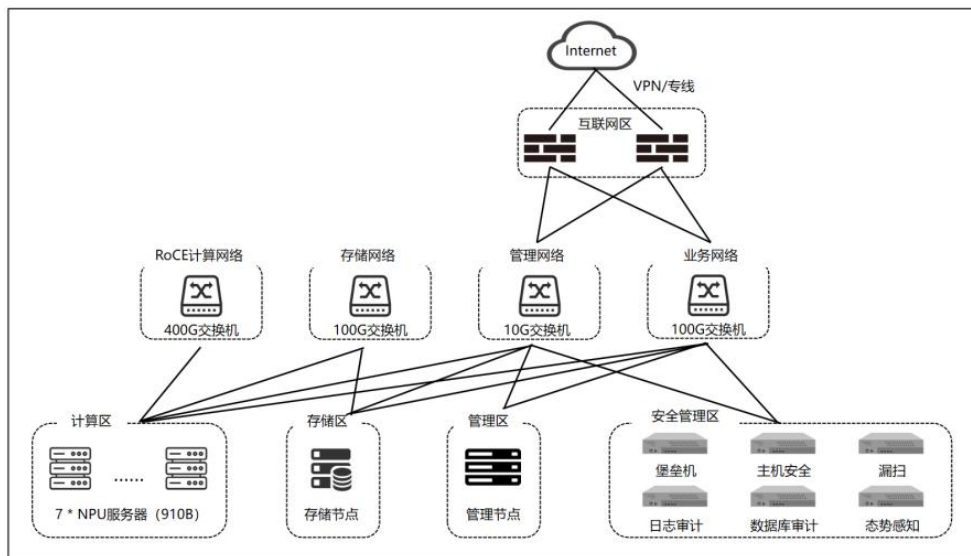


图 24 项目架构图

2) 云平台构建数据中枢，赋能 AI 多源数据融合

① 多源数据整合：云平台构建多维度数据接入体系，通过业务网络、管理网络及物联网终端，全面汇聚发电行业多源异构数据。具体包括 1081 项法规标准、事故案例、操作手册等文本数据；固定监控视频、无人机巡检影像、移动端上传图片等图像视频数据；物联网传感器采集的温度、震动、压力等设备运行数据，以及设备参数、隐患台账等结构化数据，实现“人、机、环”全要素数据的一站式整合，达成行业知识全覆盖。

② 标准化数据处理体系：云平台部署专业数据处理工具链，首先通过数据清洗工具去除噪声数据、补全缺失值，确保数据质量；随后利用 NLP 工具完成文本数据的分词、实体识别、知识抽取与向量化处理，通过计算机视觉标注工具对图像视频数据中的违规行为、设备缺陷、环境异常等要素进行精准标注；最后通过数据融合技术将不同类型、不同格式的数据标准化处理，构建结构化、语义化的多模态关联数据库，为 AI 模型训练提供高质量数据支撑。

③ 动态知识图谱构建：基于 NLP 技术深度挖掘文本数据中的实体（如“违规操作”“设备缺陷”“处置方案”）与实体间关系（如“隐患-关联法规”“缺陷-处置流程”），结合多模态数据关联分析结果，构建覆盖发电行业安全管理全领域的动态知识图谱。该知识图谱能够实现“隐患识别-风险评估-处置推荐”的智能链条生成，为 AI 模型的跨模态推理提供丰富的知识支撑，提升模型决策的科学性与准确性。

3) “云-边”协同架构支撑 AI 模型灵活部署与迭代优化

① 模型差异化部署策略：云端部署完整版多模态基座模型 Qwen2.5-VL 和 DeepSeek MoE，负责模型的整体训练、架构优化与版本迭代，同时承担全局数据的深度分析与复杂场景的泛化推理任务；边缘端部署经蒸馏压缩的 7B 轻量化模型，该模型保留核心识别与推

理能力，体积小、响应快，能够适配边缘计算终端的硬件资源，实现视频流本地化实时分析与毫秒级预警响应，无需依赖外网连接，满足现场实时管控需求。

② 大小模型协同运行机制：创新采用“小模型抢时效、大模型提质量”的协同模式，在输卸煤等关键场景中，小模型快速识别皮带撕裂、给煤机卡堵等设备故障并立即触发启停指令，防止故障扩大；大模型接收小模型上传的数据后，进行多维度深度分析，挖掘故障根源并优化模型参数，反馈至边缘端实现模型迭代。同时，小模型处理单一模态局部数据时易受干扰产生误报，大模型依托全流程历史数据与跨模态融合分析，过滤无效预警，降低误报率，二者协同实现“高效响应+精准判断”的双重目标，达成 7×24 小时无人值守与故障自愈。

③ 模型全生命周期管理：云平台建立统一的模型管理系统，负责模型版本控制、训练任务调度、性能评估与迭代优化。云端通过持续接收边缘端上传的实际应用数据与反馈结果，结合安全奖励模型+PPO 强化学习算法，对模型进行持续微调，提升模型在复杂场景下的识别准确率与鲁棒性；优化后的模型通过“云-边”协同通道快速部署至边缘端，实现“采集-分析-优化-再部署”的闭环迭代，确保模型性能持续适配行业场景变化与新需求。

④ 多场景自适应适配：云平台统一管理多场景预警规则与任务策略，根据输卸煤、汽轮机运转层、检修作业等不同发电场景的安全管理需求，灵活调整模型参数与识别重点；边缘端模型通过接收云端下发的场景化配置，快速适配高粉尘、强光/弱光交替等不同场景的环境特点与管控要求，实现一套模型体系满足多场景安全管控需求，大幅提升系统的通用性与部署效率。

4) 整合应用入口，推动 AI 功能落地与闭环管理

① 多终端交互体系构建：整合“移动端 APP+电厂监控终端+Web 端知识平台”三大应用入口，形成全场景覆盖的交互体系。移动端 APP 支持一线工作人员实时上传现场图片、查询安全知识、接收预警通知与跟踪整改进度；电厂监控终端用于集中展示全域安全态势、实时预警信息与设备运行状态，方便管理人员全局把控；Web 端知识平台提供安全文档检索、个性化学习课程、数据分析报告等功能，满足员工学习与管理层决策需求，支持 2.2 万次问题同时开展。

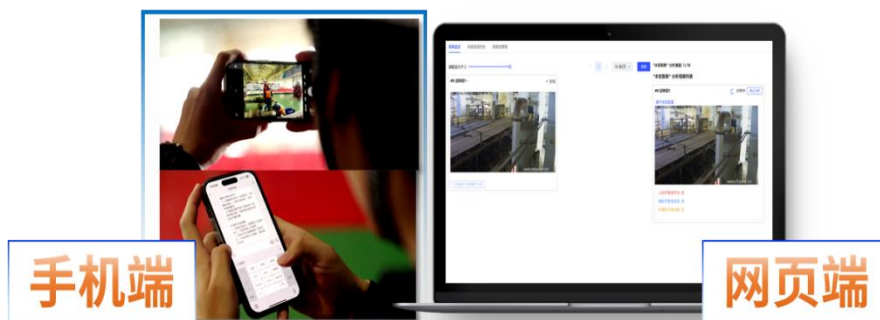


图 25 应用入口

② 核心 AI 功能模块化整合：将 AI 核心功能拆解为三大模块，实现模块化部署与协同运行。安全知识问答模块基于 NLP 子模型与知识图谱，支持语音、文本双输入，快速检索安全规范、操作流程、应急处置等 1081 项安全管理内容，响应时间控制在秒级；个性化学习模块通过分析用户行为数据、答题记录与岗位技能要求，为新老员工推送定制化学习内容，支撑多层次人才培养；风险隐患预警分析模块整合边缘端实时分析结果与云端数据，生成隐患分析报告、处置建议与趋势分析图表，精准识别人员、设备、环境三类安全风险。

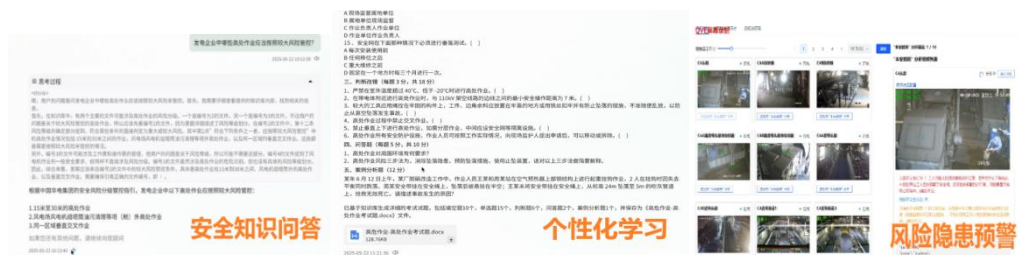


图 26 AI 功能模块

③ “模型-制度-人工”深度融合闭环: 将 AI 模型输出与企业安全管理制度深度绑定, 构建“自动识别-人工复核-整改追踪-效果评估-模型优化”的全流程闭环管理体系。AI 模型自动识别隐患后, 通过多终端推送预警信息; 人工复核确认后, 生成整改任务并分配责任人, 跟踪整改进度; 整改完成后, 系统对整改效果进行评估, 相关数据反馈至云端用于模型优化; 同时, 所有流程数据均有留存, 形成可审计、可追溯的安全管理台账, 确保安全管理工作透明化、规范化。

④ 跨部门协同支撑: 打破企业内部部门数据壁垒, 实现安全管理、生产运营、人力资源等多部门数据共享与协同工作。安全管理部门通过平台获取实时预警信息与整改数据, 优化安全管控策略; 生产运营部门基于设备隐患数据优化运维计划; 人力资源部门根据员工学习数据与技能评估结果制定培训计划, 通过 AI 技术支撑, 实现跨部门协同联动, 提升企业整体安全管理效率。

4、场景效益

“本安数斯”在能源领域应用成效突出, 助力企业提升安全管理水平、实现降本增效, 市场潜力广阔, 可助力更多的能源企业提升安全智能监测与风险防控水平。

1)应用规模

截至 2025 年 9 月 5 日, “本安数斯”云服务平台注册用户超 2.3 万人, 累计回答安全措施 8 万余条, 排查隐患 11 万余条, 日均访问量 1500+, 生成隐患报告超 4 万条。长远来看, 华电旗下有着众多电厂, 国内其他发电集团以及各类能源生产和使用企业也都是潜在用户, 具备持续扩大的潜力。

2)应用成效

① 永安电厂

在永安电厂输卸煤场景中，对生产现场的人、机、环要素安全隐患或问题 5-20s 内实现主动预警，纵向上较模型强化训练之前的平均时间（25-30s）至少提升 20%，横向上对输卸煤系统原预设的 13 种场景平均统计 9.97s，较原来 13 个串行单一模型，整体运行效率提升 67.06%。



图 27 永安电厂应用

② 十里泉电厂

在十里泉电厂汽轮机运转层场景中，“本安数斯”突破了传统计算机视觉 CV 小模型的性能瓶颈，显著提升了对关键安全细节的识别精度，以人员防护识别为例，不仅可准确判断安全帽是否佩戴，还能细致识别提醒安全帽下颚带是否规范系住；在安全带检测方面，实现了对安全带挂钩固定点的精确区分与识别。



图 28 十里泉电厂应用

3) 经济效益

降低企业运维成本：在永安发电输卸煤场景测试中，分析响应时间缩短至 20 秒以内，大幅减少人工巡检成本与运维压力。

减少事故损失：提前识别风险并预警，降低设备损坏和停产损失，帮助企业避免数百万甚至上亿元的潜在经济损失。

（七）基于现货市场机制智能云

1、案例概述

国家能源集团山东电力有限公司（以下简称“山东公司”）打造“山东电力多市场智能辅助决策云中枢”，构建覆盖发售两侧的智能辅助决策体系，通过 AI 技术实现经营主体参与多级市场策略的自动生成，推动运营决策从经验依赖向数据与模型驱动转型。

项目通过云平台与智能决策深度融合，集约化支撑 AI 全生命周期。通过浪潮平台提供从算力调度、模型训练、智能体编排到应用发布的端到端支撑，实现 AI 资源弹性供给与成本可控，确保智能体在高并发、低延时场景下的稳定运行；依托云上统一数据湖，汇聚发电侧（机组状态、机组成本、燃料库存）、售电侧（用户负荷、合约信息）及外部市场（电价、政策、气象）等多维数据，通过标准化治理与实时更新机制，为智能体提供一致、可靠、鲜活的数据集支撑；借助浪潮智能体开发平台的可视化编排能力，业务逻辑可快速转化为智能体工作流，大幅缩短“算法—应用”转化周期，单个智能体开发时长缩短 50%。

2、需求分析

2025 年是国家电力市场化改革的关键攻坚期，政策要求 2025 年底前基本实现电力现货市场全覆盖，全面开展连续结算运行。山东省作为改革先行试点省份，于 2025 年 8 月推出《新能源机制电价竞价实施细则》，推动电力市场多元化发展，而山东公司作为山东电网的

主力发电企业，8家火电电厂24台火电机组深度参与电力现货市场。在现货市场加速成熟、新能源装机占比突破49%的背景下，传统运营模式面临严峻的系统性挑战：一是火电转型压力显著，在“双碳”目标下，亟需统筹火电基荷保障与新能源消纳，传统运营模式难以支撑多市场协同决策；二是新能源波动性加剧，风光发电占比已提升至35%，但预测精度不足和储能协同机制缺失制约高效消纳；三是售电市场竞争日趋激烈，全省售电公司数百家，企业亟需构建贯穿“发-售-服”全链条的智能辅助决策体系，以提升市场响应能力与风险管控水平。

本项目对于传统场景痛点精准识别，深入山东公司24台火电机组运行一线及1000余家签约客户服务体系，系统梳理出三大核心挑战：一是现货市场分钟级价格波动下，人工制定策略时间较长（平均耗时4小时），难以匹配市场节奏；二是机组组合、燃料成本、电网约束等多维动态参数耦合复杂，传统经验决策难以兼顾经济性与合规性；三是市场感知与客户运营数据割裂，气象、生产、用电行为等关键因子未有效融合，制约策略前瞻性。

3、项目实施情况

1)项目整体实施情况

① 需求分析与全局规划

针对上述需求分析的痛点，山东公司率先于2025年2月底启动人工智能项目规划，截至11月底已部分实施体系化的解决方案：构建协同进化的多智能体系统，在已上线的“智慧交易助理”（电力市场分析）与“客户价值寻优”（电力精准营销和客户精细化服务）系统基础上，构建面向山东电力交易市场特点的电力生产、运行、营销的多智能体，全面感知市场动态、智能形成决策指导，实现多感知、多市场、多目标的及时决策一体化系统；方案所研发的“智慧交易员”智能

体，构建“规则-数据-模型”三位一体技术体系，基于数据基座，采用深度强化学习与博弈论模型，实现发电侧精准博弈。



图 29 现货交易员首页



图 30 基础数据接入



图 31 约束条件

② 政策导向深度契合：严格对标《关于全面加快电力现货市场建设工作的通知》（2025 年 4 月）及《推进“人工智能+”能源高质量发展的实施意见》，确立三级穿透式架构，以数据贯通驱动源端机组优化、荷端客户服务与市场规则适配的协同演进，打造可复制、可推广的省级电力市场智能决策样板。

2)建设进度与应用成效

① 云智融合底座建成：基于浪潮企业级云平台，构建高安全、

高可用的私有化云环境，集成 GPU 算力池、链式调用大-小模型与智能体开发框架，完成电力行业大模型的领域微调，并成功封装部署“智慧交易员”“客户价值寻优”两大核心智能体。

② 智能体应用初显实效：智慧交易员智能体已在国能山东部分火电机组投入试运行，日前运营策略生成时间由 4 小时压缩至 30 分钟内，支持多机组“一机一策”动态推演，并实现交易结果自动复盘与策略迭代；

客户价值寻优智能体完成对 1000+ 签约客户的五维价值评估（经济价值、风险控制、市场响应能力、潜力、绿色贡献度），为销售团队提供精准套餐优化建议与高价值商机预警（如企业扩产、搬迁等），线索转化效率显著提升。

3) 客户服务模式革新

项目重新构建售电业务价值链，实现从“潜在客户挖掘—签约客户分层—套餐动态优化—流失主动干预”的全流程智能闭环。通过融合工商注册、用电行为、舆情监测等多源数据，项目建立客户五维雷达图，实现差异化服务策略；同时，项目基于场景模拟引擎，自动生成推荐套餐方案，告别依赖人力覆盖的粗放式营销，较人工营销方法，该项目的实施使得客户续约率与满意度稳步提升。

4) 云平台与 AI 决策深度融合

① 集约化支撑 AI 全生命周期

浪潮云底座提供从算力调度、模型训练、智能体编排到应用发布的端到端支撑，实现 AI 资源弹性供给与成本可控，确保智能体在高并发、低延时场景下的稳定运行。

② 高质量数据治理筑牢决策根基

依托云上统一数据湖，汇聚发电侧（机组状态、机组成本、燃料

库存）、售电侧（用户负荷、合约信息）及外部市场（电价、政策、气象）等多维数据，通过标准化治理与实时更新机制，为智能体提供一致、可信、鲜活的数据集支撑。

③ 智能体开发平台加速价值落地

借助云帆智能体平台的可视化编排能力，业务逻辑可快速转化为智能体 workflow，能够以链式调用的方法串联起大模型与小模型，既发挥大模型泛化与语义优势，又用小模型弥补速度与确定性短板，大幅缩短“算法—应用”转化周期，单个智能体开发时长缩短 50%。

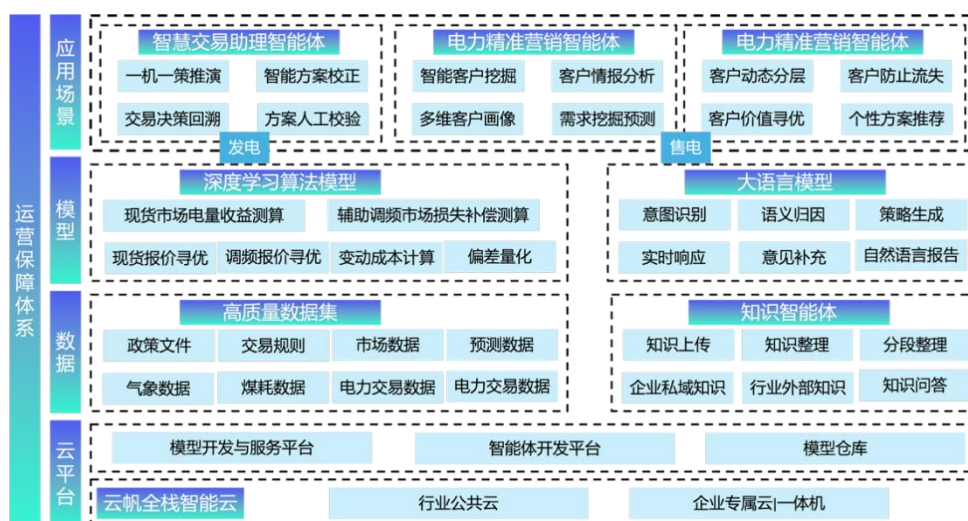


图 32 整体方案架构图

本项目已形成可量化、可验证、可推广的电力市场智能决策实践范式，为全国发电企业参与现货市场、提升经营效益提供了坚实支撑。

4、场景效益

本项目通过 AI 大模型赋能发电、售电业务实现提质、增效、降险的实质性突破，成效可测、可验、可推广：

一是决策效率实现数量级跃升。发电侧日前运营策略生成时间由原 4 小时压缩至 30 分钟内，效率提升 8 倍以上；售电套餐设计与评估周期从人工 5-7 天缩短至 2 小时内，客户需求响应速度提升超 10 倍，显著增强市场适应能力。

机组报价方案																		导出方案	收起
#6		#5		#1		#1		#2		#2		#3		#4					
序号	出力 (MW)	报价 (元/MWh)	序号	出力 (MW)	报价 (元/MWh)	序号	出力 (MW)	报价 (元/MWh)	序号	出力 (MW)	报价 (元/MWh)	序号	出力 (MW)	报价 (元/MWh)	序号	出力 (MW)	报价 (元/MWh)		
1	89	-80	1	89	-80	1	240	-80	1	95	-80	1	90	-80	1	185	-80	1	
2	115	500	2	115	500	2	300	342	2	116	500	2	300	342	2	116	500	2	
3	198	563	3	198	563	3	414	342	3	198	563	3	414	342	3	198	563	3	
4	218	570	4	218	570	4	440	386	4	218	570	4	440	386	4	218	570	4	
5	238	585	5	238	585	5	467	386	5	238	585	5	467	386	5	238	585	5	
6	257	595	6	257	595	6	493	386	6	257	595	6	493	386	6	257	595	6	
7	277	620	7	277	620	7	520	386	7	277	620	7	520	386	7	277	620	7	
8	297	690	8	297	690	8	547	386	8	297	690	8	547	386	8	297	690	8	
9	314	1180	9	314	1180	9	573	1006	9	313	1180	9	573	1006	9	313	1180	9	
10	330	1300	10	330	1300	10	600	1006	10	330	1300	10	600	1006	10	330	1300	10	

对比分析							收起
机组编号	标准十程报价		优化十程报价		对比		
	出清电量 (MWh)	日发电量 (万元)	出清电量 (MWh)	日发电量 (万元)	出清电量 (MWh)	日发电量 (万元)	
#1	7316	25.2 ⁺	31984.61	289.52 ⁺	24668.61	254.52 ⁺	
#2	7316	25.2 ⁺	31984.61	289.52 ⁺	24668.61	254.52 ⁺	
#1	8124	30.2 ⁺	32416.6	293.59 ⁺	25292.6	263.59 ⁺	
#2	8171	34.2 ⁺	33380.51	290.69 ⁺	25209.51	256.69 ⁺	
#1	2625	13.2 ⁺	5235.19	18.48 ⁺	2655.19	5.48 ⁺	

图 33 智慧交易员生成方案分析

二是经济效益与运营成本同步优化。在山东公司多台火电机组试点中，通过多目标动态优化与精准指导，度电边际收益提升 2.1%，年化综合效益预计超 1000 万元；依托“五维雷达”客户评估模型，高价值客户流失率同比下降 10%，销售线索精准度提升 60%，拓客人效提高 50%以上。

三是合规风控与服务质量双提升。发电侧机组物理约束与市场规则嵌入策略生成底层，实现方案 100%自动合规校验；售电侧建立主动预警干预机制，客户满意度有效提升，续约意愿显著增强，推动运营管理由“事后处置”向“事前预防”转型。

本案例通过技术创新与业务实践的深度融合，实现了企业运营效率与经济效益的双重提升，更构建了电力行业智能决策新范式。各项指标均通过实际运行数据验证，展现了人工智能技术在电力行业的巨大应用价值。

(八) 东莞市人工智能大模型中心

1、案例概述

东莞作为全球制造业核心基地，拥有超 22 万家工业企业、1.4 万家规上工业企业，中小企业占比超 98%，覆盖智能终端、模具制造、光伏电子等 20 余个细分领域，形成了 AI 技术落地最肥沃的产业土壤。当前制造业面临人力成本攀升、产能效率瓶颈、质量管控难题等共性痛点，对 AI 赋能的需求已从“可选”变为“必需”。

东莞市人工智能大模型中心（AIPMC）在国家 AI 建设大浪潮下应运而生，由东莞市人工智能与数字经济有限公司携手华为打造面向制造领域的城市级人工智能大模型中心，依托华为的软硬件基础底座，提供公益的普惠 AI 平台，基于制造企业共性需求，为企业提供充足 AI 算力、AI 开发平台和大模型算法服务，让企业用少量数据和低成本便可快速提高模型精度。帮助解决企业想做做不起（新技术、投资大），想做欠能力（缺乏 AI 开发人才，不具备算法规划能力），做了做不好（缺少先进技术赋能，模型开发效果差）的问题。

当前东莞人工智能大模型中心整体已建成 100P 国产化智能算力、上线部署 3 大基础大模型（自然语言大模型、计算机视觉大模型、预测大模型）、2 大制造行业大模型（制造知识大模型和制造视觉大模型）及 DeepSeek、千问开源模型 等，持续打造百模千态的模型生态，更好的满足企业研、产、供、销、服各个环节的业务场景，并且基于华为工厂自身 20+年的生产制造经验，提供给企业丰富的经验支撑，避免 AI 与实际业务结合过程中的问题，“模型能力+丰富实践”帮助企业将大模型更好的与业务结合起来。凭借“低成本、少数据、高精度”的服务优势，中心已在东阳光、优利德、交投集团、新途动力、安美科技等企业形成典型应用案例，有效提升东莞制造业智能化水平。

2、需求分析

1) 原有情况：国家整体 AI 建设浪潮下，东莞作为全球制造业核

心基地，拥有超 22 万家工业企业、1.4 万家规上工业企业，中小企业占比超 98%，覆盖智能终端、模具制造、光伏电子等 20 余个细分领域，形成了 AI 技术落地最肥沃的产业土壤。当前制造业面临人力成本攀升、产能效率瓶颈、质量管控难题等共性痛点，对 AI 赋能的需求已从“可选”变为“必需”。

2) 拟解决痛点及关键问题：

① 中小企业算力成本高：整合全市算力网络，通过算力券等政策普惠全市算力，促进算力回莞；

② 缺乏可落地 L2 场景化模型：凝练千余工业场景，提炼标准 L2 场景模型，提供企业标准服务或微调服务；

③ 企业 AI 人才匮乏：通过大模型中心标准化算力及模型，尽量降低企业自适配难度，同时联合机构等进行培训，培育东莞 AI 相关人才；

东莞市人工智能大模型中心由东莞市人工智能与数字经济有限公司建设运营，作为面向制造领域的城市级大模型中心，项目依托华为坚实的软硬件底座，精准锚定制造企业共性需求，构建起“算力 - 模型 - 服务”三位一体的全链条 AI 支撑体系。目前，中心已按 100PFlops@FP16 标准建成 100P 国产化智能算力集群，已上线自然语言、计算机视觉、预测三大基础大模型，以及制造知识、制造视觉两大行业大模型，计划后续逐步衍生 135 个场景化应用模型，目标服务企业超 330 家，既为东莞企业发展注入强劲 AI 动能，也为城市公共服务管理提质增效，持续支撑多领域用户便捷运用智能算力与大模型服务。

3、项目实施情况



图 34 东莞市人工智能大模型中心架构

东莞市人工智能大模型中心创新采用“1 个中心 + 3 种使能 + 4 类联接”的业务架构，以大模型中心为核心枢纽，通过算力使能、大模型使能、工具使能三大支撑，深度联接产业、科研、生态、人才四大维度，推动产学研与 AI 大模型技术深度融合，加速东莞全域智能化进程，面向制造业领域提供行业云服务及人工智能服务。其核心能力集中体现为三大方向：一是依托昇腾构建自主创新基础设施，以澎湃算力底座支撑城市智能跃迁；二是打造全流程工具链，强化大模型工程化能力，打通 AI 技术进产线的“最后一公里”；三是开放盘古、DeepSeek 等百模千态的大模型资源，助力 AI 在企业场景真正落地，最终构筑共融共生的 AI 技术与产业生态。

建设上遵循“三步走”策略，一期已完成算力模型底座搭建，二期聚焦行业场景适配，终极目标是创建新一代人工智能大模型先锋城市。生态布局方面，中心从研、产、营、服、管全链条梳理 AI 生态方案，通过组织各行业协会有走进华为等圈层活动，挖掘合作线索，深化生态伙伴协同。凭借“低成本、少数据、高精度”的服务优势，中心已在东阳光、优利德、交投集团、珀韵电子、新途动力、安美科技

等企业形成典型应用案例，有效提升东莞制造业智能化水平。



图 35 东莞市人工智能大模型中心建设方向与节奏

一、建成面向人工智能业务的城市级大模型中心

东莞市人工智能大模型中心整体达成 100PFlops@FP16 算力、3 个基础大模型（自然语言处理大模型、视觉大模型和预测大模型）、2 个制造行业大模型（制造视觉类大模型、制造知识类大模型），服务东莞市政府、本地企业，支撑更多领域用户使用公共算力以及大模型应用场景，构建东莞本地人工智能产业生态。

二、聚焦政务服务、智能制造上线人工智能场景应用，推动东莞政府、企业数字化和智能化转型升级

以东莞地区各级政府、关键企业、制造业为核心，打造逾百个人工智能场景应用，加速政府、企业数字化转型，加速数字化和智能化产业集聚，全面服务于政务服务、城市治理、智能制造、科学研究等领域，赋能和驱动东莞数字政府、数字经济、数字社会、智能产业高质量发展。

三、建立健全东莞人工智能产业生态，赋能东莞市千行百业高质量发展，服务大湾区

依托人工智能大模型中心培育、集聚人工智能人才、孵化人工智能产业，建立健全人工智能生态体系，全面拓展“算力+政务”“算力+工业”“算力+医疗”“算力+交通”等行业创新应用，赋能东莞市千行百业高质量发展。以制造业大模型为核心，打造区域人工智能先锋示范区，逐步形成“立足东莞，服务大湾区”的产业格局。



图 36 东莞市人工智能大模型中心赋能产业生态

4、场景效益

东莞市人工智能大模型中心“算力 + 算法 + 场景”模式成为全国制造业 AI 转型样本，获院士肯定为“可复制的东莞经验”，目前已在多个企业形成标杆案例，以下为部分案例成效介绍：

1) 新途动力优化质检质量管理：对锂电池 Pack 组装产线过程中，电芯入支架前的视觉定位检测和电芯入支架后的正负极检测，检测包括电芯型号和支架识别、电芯正负极识别、极耳（极柱）缺陷、电芯定位检测、位置偏移量控制共 5 大类外观质量检查场景。模型上线后为锂电池 Pack 组装产线质检不良事件降低 80%、管理成本减少 40%。

2) 优利德毫米级高精度测量：通过毫米级高精度测量关键尺寸（如外壳轮廓、按键间隙等），结合实时 SPC 分析与自动分拣拦截，完成万用表尺寸测量、实时质量监控共计 2 大功能。助力优利德大模

型训练推理平稳运行，质检效率提升 70%、质检管理效能提升 1.25 倍。

3) 东阳光药物研发模式创新：为东阳光提供靶点发现、苗头化合物发现、先导化合物优化的全流程应用功能，通过科学计算方式，形成药物研发辅助制药（药物分子分析、预测、优化等）的业务闭环。实现药物分子分析效率提升 50%、整体制药效率提升 1.25 倍。

4) 珀韵音圈缺陷自动化质检：因整体产品型号及数量较多，珀韵公司当前产品缺陷检测方式为人工抽检，采样率仅为 10%~15%，存在漏检可能性，影响产品质检可靠性。通过建设 CNN 和图像处理算法实现音圈缺陷自动化检测，实现锦丝线缺陷检测、焊锡偏离缺陷检测、焊锡不良缺陷检测共计 3 大功能，器件质检查全率提升 6.6 倍，器件质检查准率提升 9%。

5) 提供国产化智能算力服务：依托华为昇腾软硬件底座建成 100P 国产化智能算力集群，支持 DeepSeek、盘古等主流大模型高效训推，通过模型蒸馏、微调等技术实现知识迁移与性能优化。创新打造 AI 开发生产线，资源复用率提升 30%，应用开发周期缩短 90%，破解企业“用模难”问题，在智能质检、供应链优化等领域形成技术突破。

随着云计算与 AI 技术的发展和深入,我们将持续跟踪 AI 行业云平台技术及产业的发展情况。欢迎相关领域广大企业、研究机构、专家学者和从业人员与报告编写组联系,反馈您的宝贵意见和建议,帮助我们持续提升报告的针对性和实用价值。

感谢您对云计算开源产业联盟的支持!