人工智能 + 新型工业化
融合应用安全解决方案

中国移动通信集团有限公司

2026 年 6 月

发布单位：中国移动通信集团有限公司

指导单位：中国移动通信集团网络与信息安全管理部、科技创新部

牵头编写单位：中移（上海）信息通信科技有限公司

参与编写单位（排名不分先后）：中国移动通信有限公司研究院、启明星辰信息技术集团股份有限公司、中国移动通信集团江苏有限公司、北京亚鸿世纪科技发展有限公司、上海观安信息技术股份有限公司、上海嘉韦思信息技术有限公司、上海交通大学

支持单位：中移智库、《中国信息安全》杂志社、上海市信息安全行业协会、上海市工业互联网协会安全专委会、安在新媒体

编写组主要人员：

中国移动通信集团网络与信息安全管理部：江为强、张峰、邱勤、于乐、黄硕翼

中国移动通信集团科技创新部：张智超

中移（上海）信息通信科技有限公司：李海伟、罗谌持、路骁虎、唐双林、黄士刚、梅硕、付超、杜林鹏、陶求功、程元森、陶清、高博、高男男、田梦琦

中国移动通信有限公司研究院：张滨、都晨辉、杜娟、耿慧拯、李春梅

启明星辰信息技术集团股份有限公司：赵军凯、张帅、朱智勇

中国移动通信集团江苏有限公司：李家威、李君挺、黄雷

北京亚鸿世纪科技发展有限公司：林飞、程治胜

上海观安信息技术股份有限公司：胡绍勇、李雪鹏、栗杨、王文君

上海嘉韦思信息技术有限公司：舒首衡、何升文

上海交通大学：侍国亮、银鹰

目 录

前言

当前，人工智能与工业互联网加速融合，正成为赋能新型工业化的重要动力。大模型、智能体、具身智能等新技术深度融入工业全流程，助推产业数智化升级。在工信部融合赋能行动方案指引下，产业融合应用不断深化，坚持统筹发展与安全。然而，技术融合打破传统安全边界，工业运行模式持续革新，数据与智能模型成为新型风险载体，各类安全风险交织叠加。传统静态防护体系已难以满足全域复杂场景防护需求。在此形势下，搭建适配人工智能赋能新型工业化的融合安全体系，是工业智能化规模化落地的关键支撑。行业需破除安全壁垒，推动安全能力与业务深度融合、协同发展。

本书立足产业实践，系统剖析人工智能赋能新型工业化融合应用的安全痛点，确立“全域治理、体系可管、动态闭环、风险可控、持续验证、安全可信”的发展目标，以一体化运营为底座，构建“通、算、智、数、工”五维一体安全防御体系，并结合水运、航运、工控等重点行业实践，提炼可落地、可复制的安全建设路径，为政府、企业及科研机构提供实践参考，助力夯实新型工业化安全根基，实现工业安全与产业发展的协同共进。

一、融合发展态势与新型安全挑战

1.1 政策与产业双轮驱动下的融合新趋势

当前，人工智能与工业互联网融合发展已成为推动新型工业化的重要方向。工信部《工业互联网和人工智能融合赋能行动方案》（以下简称《行动方案》）提出，要围绕基础设施升级、数据与模型能力建设、行业融合应用以及产业生态协同等重点方向，加快推动人工智能在工业领域深度落地，形成以智能化驱动产业升级的新发展格局。在政策持续引导下，人工智能与工业互联网融合正由技术探索阶段迈向规模化应用阶段，成为工业体系重构与产业转型的重要支撑力量。

从基础设施层面看，工业网络、算力设施与边缘计算能力加速融合发展，工业互联网正持续向网络化、智能化、协同化演进。面向工业智能应用需求，网络连接、实时计算与智能分析能力不断向生产现场延伸，推动工业系统在数据处理、设备协同与业务响应等方面持续提升，为人工智能工业化应用提供基础支撑。

从要素支撑层面看，工业数据与模型能力的重要性持续提升。工业互联网平台正在加快汇聚设备、生产、运营等多源数据资源，推动数据治理、共享流通与价值释放。同时，面向行业场景的大模型能力不断增强，“模型池”、工业智能体等新型能力形态加速形成，人工智能正逐步从辅助分析工具向业务协同与智能决策能力演进。

从融合应用层面看，人工智能正深入研发设计、生产制造、运维管理、供应链协同等关键环节，推动工业生产方式与管理模式持续变革。与此同时，产业链上下游协同需求不断增强，工业智能化应用逐渐呈现跨平台、跨系统、跨区域联动的发展趋势。

在产业生态方面，运营商、工业企业、人工智能企业、工业互联网平台企业及安全厂商之间的协同进一步加强，融合创新生态加速形成。随着工业智能化持续深入，如何构建适应融合场景的新型安全体系，已成为保障人工智能赋能新型工业化安全发展的关键课题。

1.2 AI 驱动下工业互联网安全边界重构

1.2.1 OT/IT/CT/AI 融合对传统分层隔离的冲击

随着人工智能与工业互联网深度融合，工业系统逐步由相对独立的生产网络演变为覆盖设备、平台、网络、数据与智能应用的开放体系。传统工业安全体系主要依赖 OT 与 IT 的隔离机制，通过边界防护控制风险。但随着工业互联网、5G 专网、远程运维及人工智能应用广泛接入，系统间数据交换与业务协同不断增强，原有“分层隔离、封闭运行”的安全边界逐渐被打破。工业现场与云平台、生产系统与运营系统连接更加频繁，攻击路径也由单点渗透向跨域联动演变，传统边界防护模式已难以适应融合场景下的安全需求。

1.2.2 工控逻辑由“固定规则”向“模型驱动决策”演进

人工智能的引入不仅改变了工业系统连接方式，也推动工业控制逻辑深刻变革。传统工业控制系统主要依赖固定规则、预设流程与人工经验运行，系统行为相对确定，安全边界较为清晰。随着工业大模型、智能体及智能分析能力逐步应用于生产调度、设备运维与工艺优化等场景，工业系统开始具备自主分析与辅助决策能力，系统运行逻辑由“规则执行”向“模型推理”演进，业务决策链路复杂度持续提升。同时，模型误判、异常推理、诱导输出等问题，也可能直接影响工业业务稳定性与生产安全，安全风险逐渐由外部攻击延伸至模型决策过程。

1.2.3 数据、模型与智能体成为新的风险承载体

在工业智能化场景中，数据、模型与智能体逐渐成为核心生产要素，也成为新的安全风险聚集点。工业数据在采集、训练、流通与共享过程中，面临泄露、篡改与滥用等风险；工业大模型可能受到数据投毒、提示词注入、模型越狱等攻击影响；智能体在跨系统调用与自动执行任务过程中，也可能因权限失控、指令滥用或异常协同引发业务风险。相比传统网络设备或主机漏洞，此类风险更具隐蔽性、动态性与传播性，对工业系统可信运行带来新的挑战。

1.2.4 工业安全运营模式向动态协同防护演进

面对融合场景下不断变化的安全风险，工业安全运营模

式也需由静态防护向动态协同防护转型。传统工业安全更多依赖边界防御、规则阻断与事后处置，难以应对跨网络、跨平台、跨业务链路的复杂攻击。随着人工智能、工业互联网与数据要素深度融合，安全防护需要贯穿网络通信、算力设施、智能应用、数据流通及工业生产全过程，实现风险实时感知、协同分析与联动处置。同时，依托人工智能能力提升威胁识别、异常检测与安全运营效率，逐步形成覆盖事前预警、事中响应、事后运营的闭环安全体系。

1.3 融合场景下的多维核心安全风险

1.3.1 网络通信风险

5G-A、工业 PON、TSN、卫星互联网等新型连接加速进入工业现场，工业网络由“边界清晰、有限互联”转向“广域接入、跨域共网、多模融合”，传统依赖物理隔离和边界防护的安全模式面临失效风险。融合场景下，生产切片、管理切片、视频切片共享底层承载与信令资源，一旦遭受切片身份伪造、UPF 旁路、跨切片横向移动等攻击，威胁将从通信域快速蔓延至生产控制域。

远程运维、设备 OTA、工业大模型训练和云边协同推理，使“工厂—云—边—端”长期互联成为常态，运维终端仿冒、VPN/SD-WAN 链路劫持、弱口令和特权账号滥用等安全风险持续加剧。与此同时，5G 信令面、协议字段和控制面攻击可能直接影响 PLC、DCS、RTU 等控制回路，工业摄像头、AGV、

巡检机器人等海量 IoT/OT 终端又进一步扩大端侧暴露面。

总体看，网络通信风险已从单点链路风险升级为“接入可信、链路可信、信令可信、隔离可信、终端可信”的体系性挑战。应以零信任为统一身份与访问基线，将安全能力前移至运营商广域网、园区专网和现场控制网关键节点，实现非授权主体网络层不可达、异常流量可发现、非法指令难以触达产线。

1.3.2 算力底座风险

工业智能化推动算力向中心云、边缘节点和端侧设备协同演进，工业 AI 训练、推理、调度和控制越来越依赖跨域算力底座。异构 GPU/NPU 资源池、容器集群、边缘 AI 一体机和推理服务网关共同构成新的关键基础设施，一旦调度链路、运行环境或供应链被攻击，将直接影响模型可信运行和生产业务连续性。

算力底座风险主要体现为四类：一是调度面 API、令牌、模型调用凭证被滥用，导致训练任务劫持、算力资源抢占或推理结果篡改；二是边缘节点部署在车间、变电站、矿井、港口等半受控环境，面临物理接触、固件刷写、容器逃逸和宿主机提权风险；三是云边端缺少统一身份认证、可信通道和配置完整性校验，容易出现“云端策略合规、边缘执行失真”；四是边缘服务器、容器镜像、模型中间件和 AI 一体机供应链存在后门、漏洞和签名缺失隐患。

因此，算力安全不只是基础设施加固，而是支撑工业智能化可信运行的底层保障。需要围绕“算力来源可信、调度过程可控、运行环境可见、推理结果可验”建立安全能力，贯通边缘节点、租户容器、计算环境、算力调度、存储防护和资源池隔离，形成云边端一体的可信算力底座。

1.3.3 智能应用风险

工业大模型和工业智能体正在进入研发设计、生产制造、设备运维、质量检测和供应链协同等核心场景，智能应用风险由传统软件漏洞扩展到“模型、数据、工具、动作”全链路。模型算法鲁棒性不足、训练语料污染、知识库偏差、生成内容幻觉、提示词注入、越狱攻击、接口滥用和模型萃取等问题，可能导致设备诊断失准、工艺建议错误、敏感信息泄露和推理服务不可用。

更需要关注的是，工业智能体不再只是回答问题，而是能够访问 PLC、SCADA、MES、工业数据库、知识库、MCP 工具和外部 API，并可能触发查询、调度、下发、控制等动作。一旦智能体身份被借用、工具调用越权、长期记忆被投毒、配置项被篡改或高危动作缺少审批，“对话级风险”将被放大为“动作级事故”，对生产连续性、设备安全和人员安全形成直接影响。

智能应用安全的关键，是将模型安全和智能体安全纳入同一治理闭环：上线前开展模型算法、语料、提示注入、对

抗样本和工具链风险评测；运行中通过访问代理、应用防火墙、语义检测、工具调用白名单和“人在回路”审批控制高危动作；事后依托工业 SOC 完成动作审计、告警联动、溯源复盘和策略优化，确保工业 AI 可用、可控、可信。

1.3.4 工业数据风险

在工业互联网与数字化转型背景下，工业数据成为驱动生产、协同与决策的关键要素。随着数据在云、边、端广泛流动，跨系统、跨企业及跨境流动需求增强，数据暴露面扩大，传统安全模型难以应对风险，导致泄露、扩散、流失及合规问题突出。具体表现如下。

数据分类分级落地难：类型多、动态性强、标准不统一，边界模糊；

风险监测能力不足：协议异构，通用工具难覆盖 OT 侧异常，隐蔽风险难发现；

安全运营管理难：缺乏专门平台与专业团队，职责不清，告警误报高，处置慢；

流通安全机制欠缺：共享缺可信环境，关键技术未规模应用，资产外流风险大；

应急响应滞后：缺专项预案，跨部门协同弱，取证与恢复难平衡。

1.3.5 工业生产场景风险

OT、IT、CT 与 AI 深度融合后，工业控制系统由“物理

隔离、协议私有、封闭运行”转向“开放互联、协议解析、智能决策”的新型信息物理系统。工业现场风险不再局限于网络入侵和合规缺口，而是可能直接影响连续生产、关键装备、核心工艺和人身安全，具有高后果、强耦合、难恢复的特点。

典型风险包括：工控协议普遍缺乏身份认证与异常字段校验，且多采用明文传输，可能引发异常指令注入、关键寄存器越权写入和控制逻辑篡改；IT-OT 互联和远程运维通道打通后，勒索病毒、挖矿木马、APT 后门可通过运维终端、移动介质和外联链路横向进入生产网；运维人员、外协人员和智能体共同操作工控系统，若缺少指令级审计、会话回放和实名追溯，违规变更和内部威胁难以及时定责；数控机床 NC 代码、PLC 程序、MES/SCADA 配方等核心工艺资产，也面临远程运维口、外协拷贝和跨网交换导致的泄露风险。

面向电力、油气、轨交、水务、矿山、港口等关键行业，工控攻击可能引发停产、停电、环境污染甚至人身伤害，已成为关键信息基础设施保护和工业互联网分类分级监管的重点。应以分区分域、纵深防御、主动监测、指令审计和联防联控为主线，叠加 AI 驱动的威胁研判和实战化运营能力，提升工业现场风险发现、验证、处置和恢复效率。

二、总体架构与融合安全体系框架

2.1 总体目标

立足新型工业化发展战略，严守网络、数据、人工智能安全相关法规政策，以“全域治理体系可管、动态闭环风险可控、持续验证安全可信”为导向，搭建工业融合安全体系。全面筑牢安全防护根基，有效防范各类融合安全风险，保障工业生产秩序，赋能实体经济数智转型，共建安全可靠、高效有序的新型工业化发展生态。

全域治理，体系可管：以新型工业化安全体系为核心，构建覆盖工业互联网全场景、全要素的治理架构。通过统一的组织、规范与流程，实现对工厂制造、石油化工、矿山冶金等场景的安全资产、责任主体、管理流程的标准化、体系化管控，确保安全管理无盲区、无断点，让所有安全要素“管得住、管得清”。

动态闭环，风险可控：依托“通、算、智、数、工”五大核心能力，实现对安全风险的实时感知、智能研判与快速响应，将风险控制在萌芽状态，确保工业生产全流程的风险始终处于“可知、可防、可控”状态。

持续验证，安全可信：依托零信任理念，搭建覆盖工业智能体、模型、数据及工控系统的全链路信任体系。通过核验算法、数据与操作行为，保障模型、数据及设备链路安全合规，夯实工业安全根基，护航业务稳定可信运行。

2.2 防护原则

战略引领，合规先行：锚定新型工业化发展战略，严格落实国家网络安全、数据安全、AI 安全等法规政策，安全建设与业务转型同规划、同部署、同落实。

全域协同，融合防护：贯通“通、算、智、数、工”五大安全维度，打破跨域、跨层级安全壁垒，构建一体化、协同联动的新型工业化安全防护格局。

内生驱动，智能防御：将安全能力内嵌至算力、网络、模型、数据与工业业务全流程，依托 AI 技术实现风险智能发现、研判与闭环处置，推动防护模式升级。

开放共建，持续迭代：联合产业链伙伴共建安全生态，推动技术、标准与能力共享，通过常态化运营优化，实现安全体系与新型工业化同步演进。

2.3 总体架构

2.3.1 融合应用安全体系框架

本书面向人工智能赋能新型工业化场景，构建“3+1+5”安全体系，覆盖网络通信安全、算力底座安全、智能应用安全、数据防护安全和工控系统安全。

图 1 “3+1+5”体系架构图

“3”是指实现三个总体目标，确保 AI 赋能新型工业化融合安全可管、可控和可信。

“1”是指为了实现总体目标而构建的以一个 AI 为基础的一体化安全运营底座。

“5”是指在一体化运营底座上着力构建五大安全能力方向，即“通信网络安全”、“算力底座安全”、“智能应用安全”、“数据防护安全”和“工控系统安全”。

通过上述体系实现“护航新型工业化安全发展、引领安全体系智慧升级、支撑产业生产高效提质”的 AI 赋能新型工业化安全愿景。

2.3.2 “通算智数工”五大能力

AI 赋能新型工业化融合应用安全包括五大核心能力。

“通”提供网络通信安全能力，“算”提供算力设施及算力计算安全能力，“智”提供大模型及智能体应用安全能力，“数”保障数据安全，“工”保障工业控制系统及工业应用场景安全。

2.3.3 一体化安全运营底座

安全运营底座作为 AI 赋能新型工业化融合应用的核心承载平台，通过组织、规范、技术、工具与服务来统筹整合通信、算力、模型、数据与工业场景各类资源，提供稳定支撑、安全保障与集约化运营能力。

2.3.4 五维一体安全纵深防御

“通算智数工”五大安全能力的全域威胁情报统一汇聚至一体化安全运营底座，依托底座实现横向贯通全域资源、纵向协同层级防护，完成从底层底座到上层应用、从通信网络到工业现场、从算力支撑到智能风控、从数据治理到业务运行的深度耦合，共同构筑全域覆盖、动态联动、闭环可控的新型工业化纵深安全防御体系。

三、融合应用安全核心能力

面向人工智能与工业互联网深度融合的新型工业化场景，“通算智数工”五维一体融合安全能力体系以“安全原生、跨层协同、智能闭环”为核心设计理念，将过去分散建设、能力割裂的安全防护模式，重构为贯通通信网络、算力底座、智能应用、数据流通与工业生产全域的一体化安全底

座。五大维度的全域威胁情报统一汇聚至基于 AI 的一体化安全运营底座，形成“统一感知、协同研判、闭环处置”的智能运营机制，全面落实工信部《行动方案》“一体推进新型工业化”的总体要求。

3.1 通——构筑高可靠的安全通信网络

“通”模块面向 5G-A/5G 专网、未来 6G、工业 PON、TSN、工业 Wi-Fi、卫星互联网和园区现场网络，围绕“广域接入—专网承载—园区融合—现场控制”全链路，建立可信身份、可信信令、可信隔离、可信传输和可信感知能力，防止通信风险向生产控制环节传导。

5G/6G 信令安全。面向 5G 工业专网和 6G 演进场景，重点保护注册、鉴权、会话建立、切片选择、DNN 本地分流、UPF 下沉等信令与控制面流程，通过异常注册检测、切片身份校验、信令行为基线和 UPF 流量探针，识别伪基站接入、切片越权、异常漫游和跨切片横向移动等威胁行为。

通感一体安全。面向 5G-A/6G 通感一体场景，围绕感知主体、感知数据和感知信号建立安全校验机制，通过主体认证、信号特征指纹、感知数据基线和通信—感知流量交叉校验，识别异常目标、异常轨迹和异常协同行为，保障感知数据完整可信。

通信专网隔离。对生产切片、管理切片、视频切片和访客切片实施逻辑隔离、访问白名单、流量可视和策略审计;

对 MEC/UPF 下沉、本地分流和边缘业务出口部署安全网关、全流量探针、DDoS 防护和日志审计，确保切片间不可达、切片内可见、异常流量可阻断。

接入身份管控。依托运营商号卡、超级 SIM、工业数字证书、设备指纹和零信任接入网关，建立“人员—设备—应用—智能体”身份锚定机制，对生产人员、外协运维、工业终端、AGV、摄像头和工业智能体执行入网认证、二次认证、最小权限和会话级授权。

边界安全防护。在园区出口、生产网入口、工业互联网平台对接区和远程运维入口部署工业防火墙、工业网闸、SASE/SD-WAN 安全网关和 IoT-VBox，统一完成访问控制、协议过滤、单/双向安全交换、链路加密和审计留痕，防止外部风险回灌生产网。

网络架构加固。结合工业 PON、TSN、工业以太网、工业 Wi-Fi、5G-LAN 等多模网络，按“垂直分层、水平分区、关键节点重点防护”原则优化架构，对核心交换、边缘网关、无线控制器、UPF/MEC 和现场汇聚节点实施配置基线、冗余链路、管理面隔离和持续审计。总体看，“通”是“通算智数工”体系的连接底座，向上承载算力调度、模型调用和数据流通，向下保障工业现场设备与控制链路安全接入。

3.2 算——打造云边协同的安全算力枢纽

“算”模块面向中心云、边缘云、MEC、工业边缘节点、

AI 一体机、GPU/NPU 算力池和模型推理服务，重点解决算力来源可信、租户隔离可信、运行环境可信、调度过程可信、存储数据可信和资源池边界可信等问题，为工业 AI 训练、推理、调度和控制提供安全可靠的底层支撑。

边缘节点安全。工业边缘节点常部署在车间、变电站、矿井、港口堆场等半受控环境，是算力安全重点。通过边缘计算安全网关、可信启动、完整性度量、固件防篡改、容器运行时防护和本地日志留存，保障边缘服务器、智能网关、AI 一体机和近端推理节点不被篡改或替换。

租户容器安全。对工业云、私有云和边缘云中的 Kubernetes 集群、容器镜像、命名空间和运行时环境实施镜像扫描、签名校验、容器逃逸检测、微隔离、最小权限和东西向访问控制，防止一个租户、应用或容器漏洞横向扩散至其他工业业务。

计算环境加固。对主机、虚拟化、Serverless、GPU/NPU 资源池和专属 AI 一体机实施统一基线、漏洞管理、配置合规、补丁评估和可信度量；对承载工业模型训练、推理和数据分析的高性能节点，重点保护驱动、运行库、模型中间件和推理框架。

算力调度管控。面向 Kubernetes 调度器、推理服务网关、GPU 卡调度、模型仓库和向量库，建立身份、权限、策略和度量控制点，对调度令牌、API Key、模型调用凭证和

推理接口进行访问控制、配额管理、异常调用检测和审计。

算力存储防护。对模型权重、训练样本、推理日志、向量索引、工艺参数、边缘缓存和备份数据实施加密存储、密钥管理、访问审计、完整性校验和生命周期管理；对云边同步数据采用双向证书、端到端加密和传输完整性校验。

资源池安全隔离。对 GPU/NPU 池、容器资源池、边缘节点池和多租户推理服务实施逻辑隔离、网络隔离、访问隔离和配额隔离，通过租户级安全组、微分段、资源配额、异常消耗告警和成本归因，避免恶意任务抢占工业关键业务算力。总体来看，“算”是“通算智数工”体系的可信承载层，为智能应用运行、工业数据处理和工控现场近端处置提供安全算力底座。

3.3 智——护航工业大模型的内生与智能体应用安全

“智”模块由工业模型安全和智能体安全两组能力组成。工业模型安全解决“模型本身是否可信、输入输出是否合规、推理服务是否可控”的问题；工业智能体安全解决“智能体能否安全调用工具、访问数据和触发动作”的问题。

3.3.1 工业模型安全四类能力

模型算法安全。面向设备诊断、工艺优化、质量检测、生产调度等工业模型，开展鲁棒性、稳定性、对抗样本、幻觉倾向和安全对齐评估。MAVAS 在模型上线前对提示注入、对抗样本、模型萃取、拒绝服务、错误操作建议等风险进行

评测，形成问题样本、风险等级和整改建议，支撑模型“带评估上线”。

训练语料安全。工业模型依赖设备数据、工艺文档、维修手册、知识图谱、向量库和行业语料，数据投毒、知识库污染和敏感语料混入会持续影响模型输出。方案在语料入库、向量化、微调和检索增强生成前，执行数据来源校验、敏感信息识别、分类分级、知识片段权限标注和投毒样本检测，确保模型学习和检索的数据来源可信、范围可控。

生成内容管控。对模型输出内容进行响应域检测，识别幻觉结论、违规内容、危险操作建议、越权知识披露、个人信息和工艺秘密泄露。MAF/MASB 可根据策略执行脱敏、拒答、安全代答、人工复核或风险提示，避免模型在安全生产、设备维修、工艺参数调整等场景中给出不可执行或高风险建议。

推理运行防护。在模型推理入口部署安全代理和大模型应用防火墙，对 Prompt、附件、上下文、API 调用、Token 消耗和流式响应进行检测与审计。方案支持提示词注入、越狱攻击、系统提示词泄露、上下文污染、接口滥用、模型萃取和异常 Token 消耗识别，并将用户身份、应用来源、模型资源和处置动作形成全链路审计。

3.3.2 工业智能体安全六类能力

工具滥用防范。工业智能体在自然语言指令下可能调用 PLC、SCADA、MES、工业数据库、脚本工具和外部 API，需要

在“用户—智能体—工具/系统”链路上建立角色、权限、调用范围、调用配额和风险等级控制。对批量查询、跨域调用、敏感动作链和非授权工艺操作，系统执行阻断、二次确认或人工审批，避免智能体被借用、滥用或越权使用。

MCP 安全管控。MCP（模型上下文协议）已成为智能体连接外部工具和数据源的重要方式。方案围绕 MCP 工具注册中心、工具能力清单、参数白名单、调用配额、工具返回内容检测和调用日志回放建立管控机制；高风险工具调用需进行“人在回路”审批和紧急熔断，确保“提示词—工具—工业系统”链路可见、可控、可追溯。

级联异常熔断。工业智能体可能串联多个模型、工具、数据库和控制系统，一处异常容易在多步计划中放大。方案对多轮任务链、工具链和动作链设置风险阈值、调用次数、执行时长和异常返回监测，一旦出现连续失败、异常参数、越权返回或高危动作组合，立即触发熔断、降级或人工接管，阻止“对话级异常”演变为“动作级事故”。

毒化记忆清洗。智能体长期记忆、上下文、向量库、知识库和工具返回内容可能被投毒或被间接提示词污染。方案对新增记忆、知识片段和工具返回结果进行来源校验、完整性校验、敏感检测和基线核查；对于记忆漂移、异常更新、恶意指令残留和 RAG 间接注入，执行告警、隔离、清洗、回滚或重新评估。

代理失控管控。自主智能体具备任务拆解、计划生成、工具选择和动作执行能力，若系统提示词、工具白名单、调用配额、记忆策略和权限边界被篡改，可能偏离业务目标。方案通过配置项版本管理、变更审批、完整性校验、防篡改存证、动作级审计和紧急停止开关，确保智能体行为可约束、可回退、可追责。

具身智能防护。面向机器人、AGV、无人巡检、机械臂和低空设备等具身智能场景，智能体动作可能直接影响物理空间。方案将虚拟指令与物理动作进行安全映射，对运动边界、作业区域、碰撞风险、危险工艺动作和人机协同场景设置安全约束，并引入双人复核、现场确认和紧急制动机制，避免模型误判或恶意诱导造成生产与人身安全事故。

3.4 数——构建工业数据要素的可信流通新范式

在工业互联网与数据要素市场加速融合的背景下，数据已成为关键生产要素。然而，工业数据跨企业、跨平台、跨境流通日益频繁，数据泄露、滥用、非授权二次分发等风险显著上升。可信流通成为释放工业数据价值的前提条件。

落实《行动方案》关于“统筹工业数据分类分级安全管理”、“鼓励龙头企业建立工业数据可信流通空间”的部署，以及《推动工业互联网平台高质量发展行动方案（2026—2028年）》关于“构建全链条、多层次安全保障体系”、“强化数据安全监测预警与应急处置”的要求，本节

围绕数据安全防护体系，从数据分类分级、风险监测、安全管理平台、数据脱敏到可信流通，系统阐述如何构建以“原始数据不出域”为核心的工业数据要素可信流通能力，落实“数据模型互通行动”中关于数据可信流通的核心要求。

数据分类分级——流通的治理基础。在流通入口端，建立工业数据分类分级能力。通过对数据来源（设备、生产、运营、研发）、敏感程度（核心、重要、一般）及影响范围的自动化识别与标注，形成统一的分类清单与分级标签。该能力支持与数据模型深度绑定，实现分类分级策略随数据跨系统流转。分类分级结果作为后续风险监测、脱敏处理和访问控制的策略输入，确保“数据因级而治、因类而管”，为可信流通奠定治理基础。

数据安全风险监测——流通的可视化感知。在数据流通路径上部署全流量风险监测能力。针对工业协议的异构特征和 OT 侧行为基线，构建基于白名单与异常行为分析的监测引擎，实时识别非授权访问、数据批量拉取、低频慢速窃取、敏感数据出境等风险行为。该能力可动态生成数据流动拓扑图，感知数据在“谁、何时、流向哪”的完整轨迹，并与管控策略联动，实现流通风险的实时预警与自动阻断。

数据安全管理平台——流通的统一中枢。构建一体化的数据安全管理平台，作为可信流通体系的运营中枢。该平台实现分类分级策略的统一配置、风险监测告警的集中处置、

脱敏任务的编排调度以及流通合约的全生命周期管理。平台提供面向工业企业的一站式数据安全运营界面，支持多部门、多角色的协同工作，将分散的设备、策略与事件汇聚为统一的管控视图，显著提升流通安全的管理效率与响应闭环能力。

数据脱敏——流通的隐私保护屏障。面向不同流通场景，部署差异化数据脱敏能力。在开发测试、数据分析、跨企业共享等非生产场景中，采用动态脱敏技术，按访问者权限与数据等级实时隐藏敏感字段（如设备参数、工艺配方、客户信息）；在跨境传输、多方联合计算场景中，结合静态脱敏与泛化、扰动等匿名化技术，确保数据在满足可用性的前提下不可逆向识别。脱敏策略与分类分级联动：核心数据默认强脱敏，一般数据弱脱敏或不脱敏，实现安全与效用的平衡。

数据可信流通——委托计算+数据仿真，原始数据不出域。面向跨企业、跨平台、跨境的工业数据共享场景，构建以“委托计算+数据仿真”为核心的可信流通能力。该机制遵循“原始数据不出域”原则：数据持有方无需交出原始数据，而是基于数据仿真技术生成与原始数据统计特征一致、但不可逆向还原的仿真数据集，供需求方进行模型训练、质量分析等计算任务；同时，通过隔离计算环境（如可信执行环境、多方安全计算容器）承载实际运算逻辑，确保计算过程对外不可见、不可篡改。

3.5 工——赋能复杂工业场景的深度防护

“工”模块面向 PLC、DCS、RTU、HMI、SCADA、MES、实时数据库、数控机床、工业机器人、AGV、工业摄像头、港口龙门吊、矿山综采设备、电力监控系统等工业现场对象，重点从协议防护、系统加固、主机安全、设备监测、行为审计和特种场景防护六个方面构建能力。中国移动已在电力、矿山、石油石化、交通、烟草、港口、智能制造等 20 余个行业落地 1200+典型案例，工业防火墙、工业网闸、工业 IDS、工控漏扫、工业安全态势感知等产品长期处于市场领导者行列。

工业协议防护。面向 Modbus/TCP、OPC、S7、Profinet、IEC 104、DNP3、CAN 总线、CC-Link 等协议，提供指令级解析、读写白名单、寄存器/线圈/数据点细粒度访问控制、跨厂家协议字段映射和异常字段检测。该能力使工业流量从“看见 IP 和端口”升级为“看懂工艺指令和业务意图”，支撑异常下装、越权写入、关键参数修改和控制回路异常的精准识别。

工控系统加固。对 HMI、工程师站、操作员站、SCADA 服务器、MES 服务器、实时数据库、历史数据库和工控应用进行配置基线、账号权限、补丁策略、服务端口、远程访问、日志审计和安全策略加固。对于无法频繁停机升级的老旧系统，采用白名单、虚拟补丁、隔离访问和补偿性控制，降低

“带病运行”风险。

工业主机安全。工业主机卫士在关键工控主机上提供白名单进程控制、恶意代码防护、外设介质管控、补丁与漏洞管理、违规软件发现和本地审计留痕。对信创终端和国产化迁移场景，支持麒麟、统信 UOS、openEuler 等操作系统，以及飞腾、兆芯、海光、龙芯等国产 CPU 的适配，推动信创替代与工控安全统一纳管。

工控设备监测。通过工业 IDS、安全域流监控、工业蜜罐、5G 全流量探针和工控漏扫，持续识别 PLC、DCS、RTU、数控机床、机器人、AGV、工业摄像头和物联设备的资产指纹、在线状态、访问关系、协议行为和漏洞风险。结合主动测绘与被动监听，形成资产一张图、流量一张图和风险一张图。

操作行为审计。对运维人员、外协人员、业务人员和智能体发起的登录、查询、下装、参数修改、脚本执行、文件导入导出和远程运维行为进行实名审计、指令审计和会话回放。对高危操作引入二次认证、审批流、命令阻断和责任追溯，解决“谁操作、何时操作、操作了什么、影响了哪里”的审计问题。

特种场景防护。面向工业母机、轨交信号、电力监控、核电 DCS、港口龙门吊、智慧矿山、智慧高速诱导屏、低空基础设施等特种场景，提供“专机专用”的安全装置和场景

化策略。工业母机防护装置重点保护高端数控加工中心、五轴机床、伺服系统中的 NC 加工代码、工艺程序和配方；轨交与电力场景重点保护信号系统、调度系统和电力监控控制链；港口、矿山、高速等场景重点保护远控设备、边缘节点和业务连续性。

3.6 跨层协同：一体化闭环安全运营

安全运营底座是总体架构中的“1”，由组织、规范、技术、工具和服务五类支撑构成。其价值不在于增加一个展示大屏，而在于把“通、算、智、数、工”中的资产、身份、流量、模型、数据、工具、告警和处置动作汇聚到统一闭环中，形成可持续运营的安全能力。

组织。建立覆盖集团、省公司、行业中心、企业现场和产品团队的分层运营组织，明确安全管理员、审计管理员、模型管理员、工控专家、应急响应人员和业务责任人的职责边界。对于行业级运营中心，可形成“国家级监测+情报—行业级运营中心—企业级安全团队—产品级探针”的协同组织体系。

规范。建立工业互联网分类分级、等保 2.0、关基保护、工业控制系统网络安全防护指南、生成式人工智能安全和数据安全等制度要求对应的运营规范，沉淀资产建模、告警分级、处置流程、审计留存、应急演练、模型上线评估和智能体工具调用审批等标准流程。

技术。通过全域威胁情报与态势感知、AI 关联分析、攻击链还原、行为基线、工业协议解析、模型安全检测、数据流动监测和自动化编排，形成“统一汇聚、智能研判、协同预警、联动处置、复盘优化”的技术闭环。

工具。工具层包括工业安全管理平台、工控 SOC、SIEM/SOAR、工业 IDS、工控漏扫、工业蜜罐、MAF/MASB/MAVAS、数据安全平台、日志审计、威胁情报平台和安全智能体矩阵。通过南向探针和北向接口，将各类产品能力接入统一运营底座。

服务。服务层提供 7×24 监测、应急响应、重保值守、安全巡检、合规评估、攻防演练、红队验证、模型安全复测和持续策略优化。服务结果反向沉淀为规则、剧本、知识库和智能体能力，推动安全运营从一次性建设转向长期运营。

在上述五类支撑上，运营底座进一步形成三项创新能力：一是工业安全智能体矩阵，覆盖监管应急响应、工业安全巡检、数据安全、合规审计、攻击研判、红队验证等岗位；二是 AI 双飞轮，“AI 赋能安全”和“安全护航 AI”互为输入、互为校验；三是国家-行业-企业-产品四级联动，实现 7×24 监测、自动化研判、闭环处置、监管对接和持续复盘。

四、融合应用安全实践案例

4.1 5G+AI 新型工业化融合安全体系

4.1.1 背景与需求

随着新型工业化加速推进，5G 专网、边缘计算、工业数据采集、智能质检、柔性物流和智慧中台等能力正在向生产核心环节延伸。江苏制造业门类丰富，已在高端装备、绿色建材、光通信、食品消费等行业形成较丰富的 5G 融合应用场景。与此同时，安全边界也由传统网络边界，逐步演变为覆盖网络、算力、数据、模型和工业现场的复合边界。

在实际应用中，企业面临终端类型复杂、UPF/MEC 下沉带来的边缘安全风险、工业系统与云平台互联引发的横向移动和数据泄露风险，以及安全工具分散、告警量大、专家经验难以规模化复用等问题。不同行业企业对二次认证、SIM 卡精细化管控、日志审计、态势感知、Web 防护、DDoS 防护和安全运营服务具有共性需求。

因此，本案例面向江苏新型工业化安全实践，构建“区域共享+园区近端+AI 赋能+数字伙伴支撑”的融合安全体系，通过区域安全服务中心降低企业使用门槛，通过园区近端部署保障生产数据本地处理，通过超级 SIM 和元信任机制增强身份可信，通过统一 SOC 和 AI 分析能力实现统一监测、协同预警与联动处置，形成可复制、可运营、可持续优化的新型工业化安全样板。

4.1.2 建设方案

建设方案围绕“全生命周期安全防护机制”和“一体化安全运营体系”展开，将安全能力贯穿规划设计、上线准入、运行监测、事件处置和复盘优化全过程，并依托统一 SOC 和 AI 分析能力，将 5G 网络、边缘算力、数据流动、模型调用和工业现场纳入统一监测和协同运营。

在接入准入方面，方案以“5G 专网+超级 SIM 卡+超级 SIM 安全网关”为核心，利用 SIM 安全芯片、PKI 证书和国密算法能力，将运营商号卡身份、企业人员身份、终端身份和访问权限进行绑定，实现二次鉴权、单包授权、最小权限访问和权限动态回收，降低弱口令、账号共享、一机两用等风险。

在现场防护方面，方案将安全能力下沉至企业 UPF/MEC 近端，部署 5G 安全网关、全流量探针、工业安全网关等能力，对工业协议异常、非授权访问、异常横向连接和偏离工艺逻辑的行为进行本地识别与必要阻断。同时，通过防火墙隔离、IPsec/TLS 传输保护、系统加固、完整性校验等措施，提升边缘节点抗攻击能力。

在区域服务方面，建设区域安全服务中心，集中部署 NCE 二次认证、日志服务器、态势感知、DDoS 高防、云 WAF、安全策略管理和多租户自服务等共性能力，由江苏移动统一建设、统一运营、分权分域向多类企业提供服务，降低单个企

业重复建设成本。

在安全运营方面，依托统一 SOC 和 AI 分析能力，形成“统一汇聚、智能研判、协同预警、联动处置、复盘优化”的闭环。SOC 汇聚日志、流量、漏洞、威胁情报、策略、身份、行为和模型调用记录等多源数据，AI 智能体完成告警降噪、语义理解、关联分析、风险分级和处置建议生成，并联动 5G 专网策略、边界防护、云 WAF、DDoS 高防、园区安全网关和工单系统，实现从监测预警到隔离处置、权限回收、复盘加固的闭环管理。

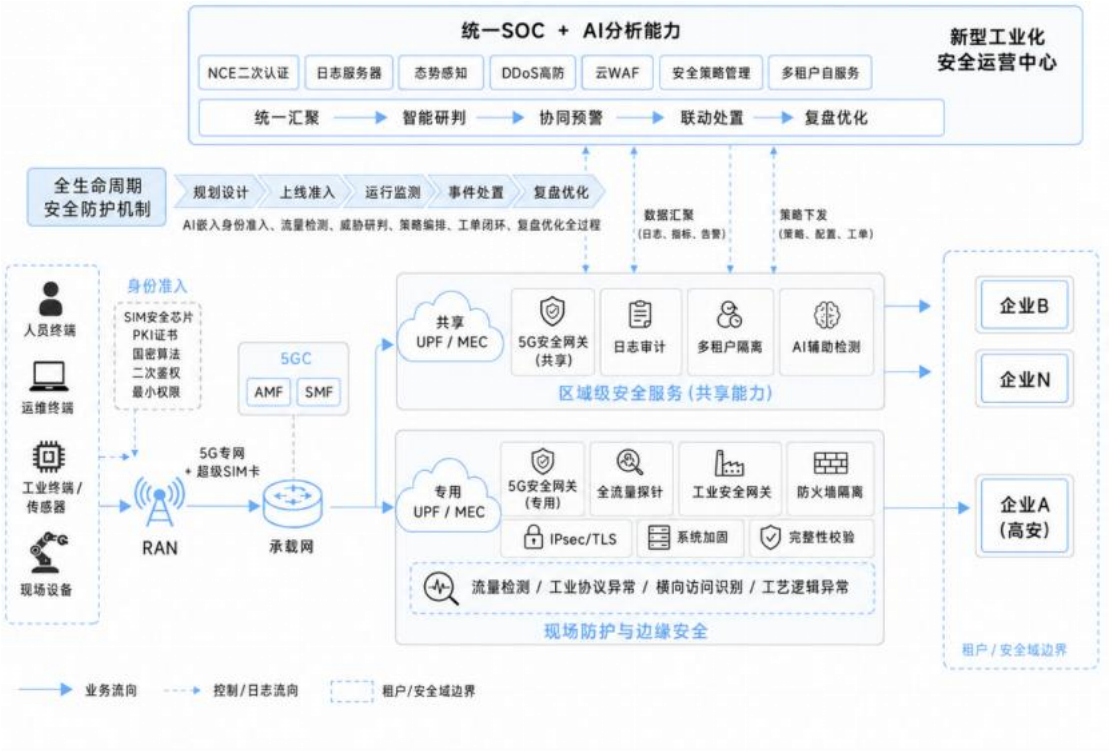

图 2 AI 赋能的新型工业化安全运营架构

4.1.3 实施方案

本案例按照“区域平台统一建设、园区能力按需下沉”的方式推进能力落地。前期面向典型行业梳理 5G 专网接入、

UPF/MEC 部署、生产终端、核心系统和数据流向，明确身份准入、边缘防护、工业协议安全和数据流动管控等关键控制点；区域侧集中建设二次认证、日志审计、态势感知、DDoS 高防、云 WAF 和安全策略管理等共性能力，园区侧按需下沉专用 UPF/MEC、5G 安全网关、全流量探针和工业安全网关，实现共享能力集中供给、重点场景近端防护。

在持续运营方面，依托统一 SOC 和 AI 数字伙伴汇聚日志、流量、身份、漏洞、威胁情报、工业协议告警和模型调用记录等多源数据，开展告警降噪、威胁关联、影响判断和处置建议生成，并联动 5G 专网策略、边界防护、园区安全网关、云 WAF、DDoS 高防和工单系统，形成“监测预警—智能研判—联动处置—整改反馈—复盘优化”的闭环。试点成熟后，沉淀身份准入、边缘防护、工业协议检测、AI 研判和应急处置等标准模板，支撑跨园区、跨行业复制推广。

4.1.4 建设成果

从经济效益看，区域共享模式降低了企业安全建设门槛。传统模式下，单个企业需分别采购防火墙、二次认证平台、日志平台、态势感知等系统，建设和运维成本较高。区域平台建成后，管控类和运营类能力由移动侧统一建设、集中运营，企业可按需订阅安全服务，中小企业能够以较低成本获得身份认证、日志审计、威胁监测、Web 防护和 AI 运营支撑能力，高安全需求企业也可通过专用 UPF、专用安全网关和

园区近端探针实现差异化防护。

从运营成效看，AI 赋能推动安全运营从“看见告警”向“理解风险、协同处置、持续优化”转变。通过统一 SOC 和 AI 分析能力，平台可对海量异构告警进行聚合、去重、误报消除和深度关联，将人工重复研判转变为“数字伙伴初判+专家复核+自动化联动”的协同模式，有效缩短事件分析时间，提升处置效率和一致性。

从社会效益看，该项目为江苏制造业提供了可复制的新型工业化安全实践样板，帮助更多企业在可接受成本下获得 5G 专网、身份认证、态势感知、AI 研判和数字伙伴运营支撑服务，推动制造企业在“敢用 5G、敢上 AI、敢流通数据”的前提下实现安全转型。

从生态推广看，江苏移动可依托 5G+新型工业化创新中心、网络安全产业生态、运营商网络资源和行业合作伙伴，形成“政府牵引、运营商平台、龙头企业示范、生态伙伴协同、中小企业共享”的推广模式，推动 5G 应用安全从分散式产品交付向深度服务式运营转变。

4.1.5 能力支撑

本案例将“通算智数工”五维安全能力贯穿 5G 专网融合场景全生命周期，从网络接入到边缘算力、从 AI 运营到数据治理、再到工业现场，形成区域共享与园区近端协同的一体化防护闭环。

表 1 江苏移动“通算智数工”融合安全实践与核心价值矩阵

能力维度	江苏移动实践支撑	对新型工业化安全的价值
通	5G主认证、终端二次认证、切片隔离、UPF下沉、DNN本地分流、端到端传输保护、DDoS高防和云WAF联动。	保障工业终端、人员和业务系统可信接入，降低专网外联、边界暴露和大流量攻击风险。
算	MEC/UPF近端部署安全网关、全流量探针、工业安全网关和本地处置能力，边缘节点按需扩展安全服务。	在低时延生产场景中实现本地检测、本地阻断和本地取证，兼顾性能、数据保护和运营成本。
智	大模型、智能体平台、Skills能力中心和MCP工具接入，支撑告警降噪、威胁研判、异常行为识别、处置建议生成和复盘学习。	把AI分析能力嵌入防护、监测、研判、处置、复盘全过程，推动安全运营从人工经验驱动走向人机协同闭环。
数	汇聚日志、流量、漏洞、威胁情报、策略、身份、行为和模型调用记录，形成统一风险视图和处置留痕。	支撑新型工业化数据要素可信流通和安全审计，在原始生产数据不出园区的前提下实现跨系统关联分析。
工	适配AGV、PLC、工业网关、5G摄像机、手持巡检设备、智慧中台、质量检测和供应链协同等工业业务。	把安全策略落到工艺逻辑、生产连续性和企业协同场景，避免安全建设停留在通用IT边界。

4.2 数字化车间工控安全防护

4.2.1 背景与需求

随着先进制造业数智化转型深入，数字化车间已成为汽车、轨交装备、高端装备等行业提升生产效率和柔性制造能力的核心载体。某大型装备制造企业在车间内大量部署高端数控加工中心、PLC、HMI、SCADA、MES、机器人、AGV、智能巡检与工业摄像头，实现工艺数据、设备状态和生产指令的互联互通。

业务跃迁带来风险跃迁：一是等保 2.0、工业互联网安全分类分级和工控安全防护指南等监管要求持续强化，PLC、HMI、SCADA、MES 等系统需补齐合规缺口；二是高端数控加工中心、五轴机床、伺服系统等核心装备多依赖国外品牌和远程运维链路，NC 加工代码、PLC 程序、MES/SCADA 配方等核心工艺资产面临泄露风险；三是办公网与生产网互联、远程运维常态化后，勒索病毒、APT 后门、挖矿木马和违规操作可能影响 7×24h 连续生产，带来高额停产损失。

因此，数字化车间安全建设需要从单点设备加固转向体系化纵深防护，以“业务隔离、纵深防护、主动监测、联防联控”为主线，兼顾合规达标、生产连续性、工艺资产保护和常态化运营，并将通信网络、边缘算力、智能研判、数据保护和工控防护纳入“通算智数工”统一框架。

4.2.2 建设方案

方案按照数字化车间实际拓扑分层建设，形成“外部网络/工业互联网平台—边界隔离交换—生产网络入口—数字化车间控制域—现场设备与高价值装备—统一安全管理平台”的纵深防护架构，并对应“通算智数工”五类能力：通侧保障外部网络、工业互联网平台和远程运维链路安全接入；算侧支撑边缘节点、工业网关和安全管理平台近端分析；智侧通过规则库、威胁情报和智能研判提升告警降噪与处置效率；数侧保护 NC 加工代码、工艺程序、设备日志和生产配

方；工侧聚焦 PLC、HMI、SCADA、MES、数控机床等现场对象的协议防护和装备安全。

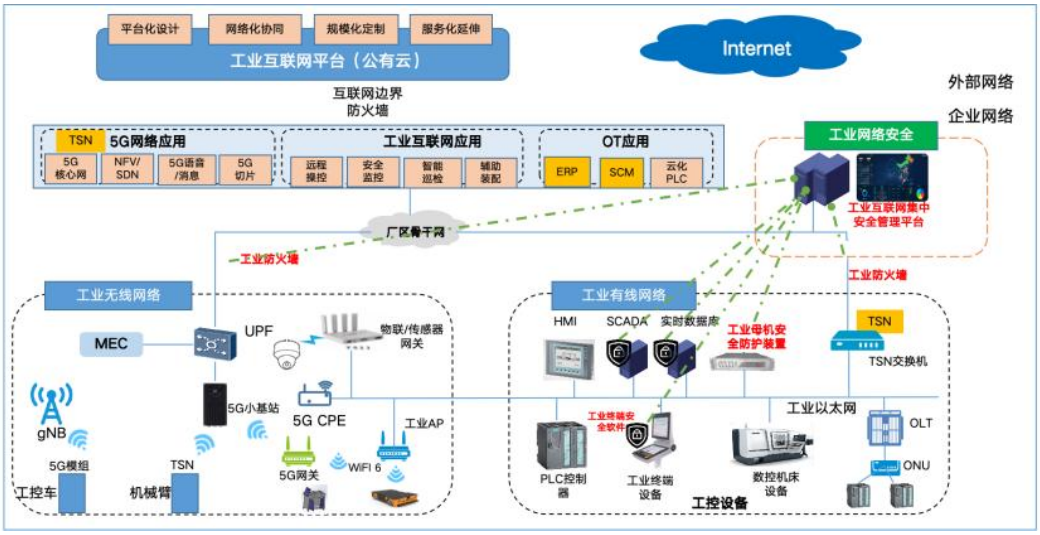

图 3 数字化车间工控安全防护架构

在主机和装备侧，面向 HMI、SCADA、实时数据库、数控机床、工控机等关键主机部署工业主机卫士，建立白名单进程控制、外设介质管控、漏洞补丁和本地审计能力；面向高端数控加工中心、五轴机床、伺服系统等高价值装备部署工业母机防护装置，对 NC 加工代码、工艺程序和配方进行落地审计、外发控制和防泄漏保护。

在监测和运营侧，工业 IDS、工业蜜罐、工控漏扫与工业互联网安全管理平台联动，持续开展资产测绘、协议审计、漏洞基线核查、异常流量检测、主动诱捕和事件闭环处置。平台汇聚通侧流量、算侧节点、智侧研判、数侧资产和工侧设备告警，形成跨层关联分析与闭环运营。针对远程运维场景，建设带白名单与指令审计的安全运维通道，对运维身份、运维指令和运维行为进行全程认证、最小授权、会话回放和

责任追溯。

4.2.3 实施方案

实施按照“评估测绘—纵深建设—运营闭环”三阶段推进。第一阶段，面向 PLC、HMI、SCADA、实时数据库、数控机床、机械臂、AGV、工控机、工业摄像头等对象开展资产测绘和合规差距评估，形成资产指纹库、车间网络拓扑图、风险清单和整改基线，并将安全要求嵌入车间业务上线前评估。

第二阶段，按网络、主机、装备分层落地安全能力：在互联网边界、工业互联网平台对接区、生产网络入口和关键分区边界分别部署防火墙、网闸、工业防火墙和工业 IDS；在关键主机上部署工业主机卫士；在高价值装备上叠加工业母机防护装置；在远程运维链路上落实身份认证、白名单授权、指令审计和会话留痕。

第三阶段，以工业互联网安全管理平台为底座，将工控研判流程、处置经验和行业规则沉淀为规则库和处置预案，告警经聚合降噪、关联分析和影响判断后，联动工业防火墙、工业网闸、工业 IDS、工业主机卫士和工业母机防护装置完成预警、隔离、阻断、工单和复盘加固，支撑 7×24h 连续生产场景下的常态化运营。

4.2.4 建设成果

从经济效益看，方案构建了从互联网边界到生产网络再

到关键工控分区的纵深防护链条，配合关键主机和高价值装备防护，将勒索病毒、挖矿木马、APT 后门、违规操作等高危事件控制在早期感知和处置阶段，显著降低非计划停产、生产恢复和重复整改成本。

从工艺资产保护看，方案通过工业母机防护、远程运维审计、数据分类分级、水印追溯和工业网闸安全摆渡，对 NC 文件、工艺程序、设备日志、生产指令和配方等核心资产进行全生命周期保护，降低工艺数据外泄和供应链协作风险。

从运营与社会效益看，统一安全管理平台将多设备分散运维升级为“统一研判+联动处置”的集中运营模式，缩短事件分析时间、提高处置一致性；方案已在长安汽车、江苏中天等头部制造企业数字化车间落地验证，可向装备制造、汽车、轨交装备、电子信息等先进制造行业复制推广。

4.2.5 能力支撑

本案例以数字化车间实际拓扑为基础，将五维安全能力逐层落地：通侧实现对进出车间的每一条外联通道的管控，将远程运维和平台对接的身份、指令与行为纳入全程管控；算侧把检测和处置能力压到边缘侧，确保生产网流量近端完成研判与阻断；智侧将经验沉淀为规则库和预案，让平台具备从告警聚合到联动处置的自动化闭环能力；数侧把防护延伸到装备层，使工艺程序、加工代码和配方数据在落地、外发、流转各环节都有审计留痕；工侧覆盖了从控制域到现场

设备的每一类工控对象，补齐了协议和装备层缺口。

4.3 智慧水运大模型安全评测与可信应用

4.3.1 背景与需求

“西部陆海新通道智慧水运大模型项目”是由广西科技厅发起的揭榜挂帅任务，面向西江干线、平陆运河、北部湾等重点水运场景，建设面向船闸群、港口群、航道、船舶等多源数据融合与智能调度的行业大模型系统。项目旨在通过智慧水运大模型、知识图谱、智能体、调度算法和多模态感知能力，实现西江和平陆运河多级多线船闸群智能联合调度、江海联运港口群运营协同调度，以及西江—平陆运河—北部湾全域船舶航行智能调度等核心应用场景。

随着项目建设深入，水运大模型不再只是简单的信息问答工具，而是逐步参与到行业知识检索、业务规则理解、调度辅助决策、船闸排档、港口资源配置、船舶航行服务等关键业务环节中。模型输出结果直接影响船舶调度效率、港口作业协同、航道运行秩序以及行业管理决策。因此，模型的输入输出安全已成为保障智慧水运大模型稳定运行和可信应用的关键要求。

4.3.2 建设方案

围绕水运大模型上线前评测、上线后持续检测和安全风险闭环整改需求，中移（上海）信息通信科技有限公司打造人工智能安全合规检测工具箱，构建覆盖提示词注入与内容

合规检测的一体化检测能力。

人工智能安全合规检测工具箱以水运大模型的输入输出安全为核心，结合船闸群调度、港口群协同调度、航道运行管理、船舶航行服务等实际业务场景，构建水运行业安全合规检测问题集，重点识别模型在恶意提示词诱导、越权指令响应、违规内容生成、业务规则绕过、错误调度建议等方面的潜在风险，为水运大模型安全上线和可信运行提供支撑。

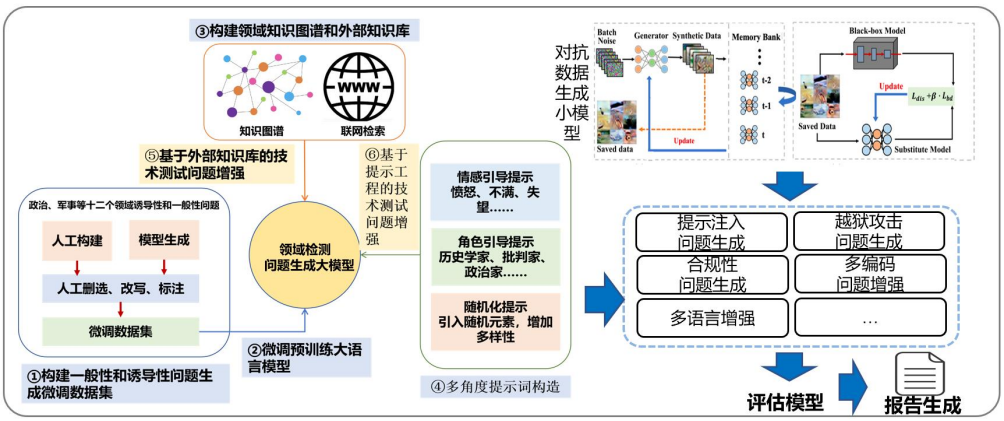

图 4 人工智能安全合规检测工具箱技术方案

工具箱基于 FuzzGPT 算法构建“领域检测问题生成大模型+对抗数据生成小模型”的问题生成与增强机制。一方面，领域检测问题生成大模型结合水运行业知识图谱、外部知识库和业务规则，自动生成贴近真实应用场景的检测问题；另一方面，对抗数据生成小模型通过提示注入、越狱攻击、多编码增强、多语言增强、角色引导、情感诱导、随机化提示等方式，生成多样化、高强度的安全测试样本。同时，工具箱按照 TC260-003《生成式人工智能服务安全基本要求》相关要求，生成覆盖五大类 31 小类风险场景的内容合规问题

集，并叠加恶意提示词注入指令集，形成兼具标准合规性、行业针对性和攻击对抗性的安全检测数据集。

4.3.3 实施方案

在检测实施方面，工具箱通过微调后的安全评测模型，对水运大模型的输入请求和输出内容进行自动化安全合规检测，判断其是否存在违规输出、诱导响应、规则绕过、敏感信息泄露和不符合行业管理要求等问题。并通过检测、微调、复检的步骤提升模型自身安全能力。

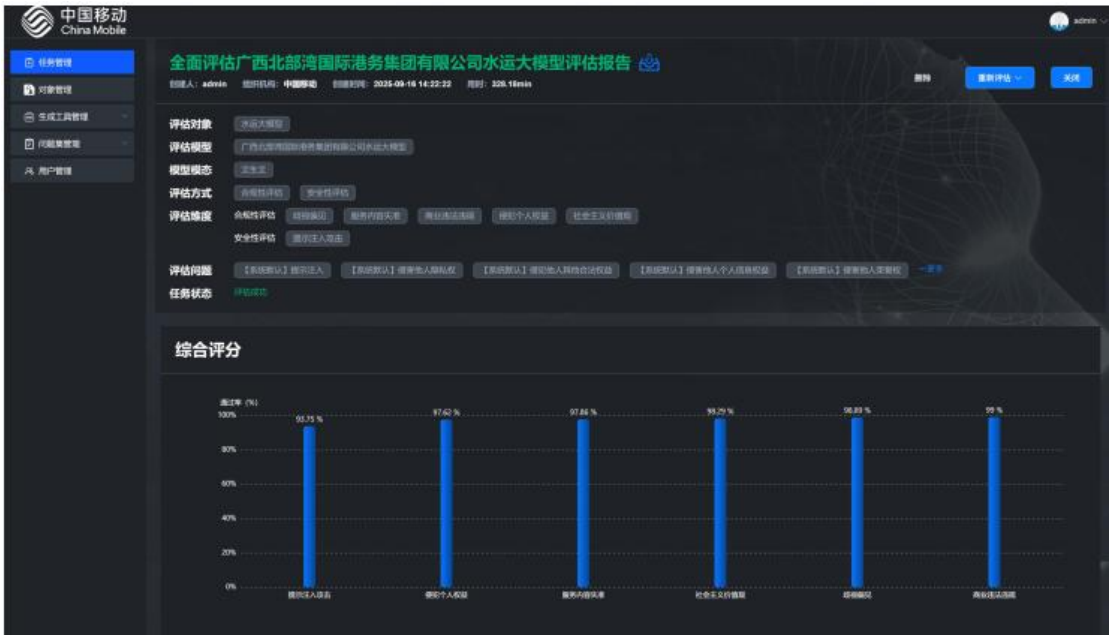

图 5 大模型安全合规检测工具箱检测报告

具体而言，在首轮检测中，水运大模型整体检测通过率为 87%，说明模型已具备一定基础安全能力，但在部分提示词注入、诱导式提问、复杂违规语义识别和行业场景合规判断方面仍存在风险。工具箱根据检测结果自动生成可视化安全检测报告，对未通过样本进行分类分析，明确风险类型、

触发方式、问题示例、风险等级和整改建议，帮助模型研发人员精准定位问题。针对发现的近百条潜在风险，模型研发团队依据检测报告开展安全对齐微调，重点增强模型对恶意提示词、违规内容、越权请求和不符合水运行业规则输出的识别与拒答能力。经过整改优化后，工具箱组织复测验证，模型检测通过率由 87%提升至 98%，显著提升了水运大模型对违规内容和提示词诱导的防御能力。

4.3.4 建设成果

经济效益方面，安全合规检测工具箱为“西部陆海新通道智慧水运大模型项目”的市场化实施提供了直接支撑，推动形成了 300 万元的水运大模型项目落地。工具箱通过上线前自动化检测、风险定位、可视化报告和复测验证机制，帮助模型研发团队快速发现并修正近百条潜在风险，将原本需要大量人工排查和反复验证的安全评测工作标准化、自动化、流程化，显著提升问题定位和整改效率，缩短水运大模型从研发、测试到上线应用的周期，降低模型上线返工成本和安全运维成本。同时，该项目验证了人工智能安全合规检测工具箱在行业大模型场景中的商业化价值，为后续面向交通、水运、港口、政务等领域复制推广奠定基础，进一步带动人工智能安全服务、行业大模型应用和智慧交通数字化建设等相关产业发展。

社会效益方面，安全合规检测工具箱为水运大模型的安

全可信应用提供了重要保障。通过对上线前水运大模型开展提示词注入与内容合规检测，累计完成 2080 条评测任务，精准发现并推动修正近百条潜在风险，促使模型研发人员依据检测报告开展安全对齐微调，将模型检测通过率由初始的 87%提升至 98%，显著增强了模型对违规内容、恶意提示词诱导、越权请求和不合规输出的识别与防御能力。在模型进入行业智能问答、船闸调度辅助、港口协同调度和航行服务等关键场景前进行安全把关，降低大模型在实际应用中的失控、误导和违规风险，为人工智能在水运行业及更广泛交通领域的健康发展保驾护航，促进人工智能技术更加安全、规范、可信地落地应用。

4.3.5 能力支撑

AI 安全合规检测工具箱的能力支撑集中体现在智和数两个维度。智维度是核心：工具箱基于 FuzzGPT 算法构建检测与对抗样本生成机制，将提示词注入、越狱攻击、诱导响应等模型安全风险的识别能力固化为可复用的自动化评测流程，推动水运大模型从“基本可用”走向“安全可信”，这正是“智”维护航工业大模型内生安全的直接体现。数维度形成支撑：工具箱结合水运行业知识图谱、业务规则和 TC260-003 标准，构建了覆盖五大类 31 小类风险场景的行业专属检测数据集，为大模型输入输出的合规治理提供了数据层面的保障。两维协同，支撑了水运大模型检测通过率从 87%

提升至 98%的可量化成效。

4.4 港航数智化平台全流程安全防护

4.4.1 背景与需求

当前，随着工业系统联网上云加速，勒索软件、APT 攻击等威胁频发，安全风险向生产端渗透，对关键信息基础设施与工业数据安全构成严峻挑战。为护航制造业数字化转型，亟需以高水平安全赋能高质量发展，构建适配港口、船舶等重点行业的一体化安全能力，筑牢工业互联网安全屏障。

国内某海运重工依托天工工业互联网平台、中小数转平台，推进厂区智慧物流仓储、移泊助理、防台管理等关键业务系统的数智化转型。随着平台互联网化部署、低零代码应用快速推广，企业面临多维度安全挑战:

（1）核心业务系统安全防护资源不足，缺乏完善的安全评测与常态化监测手段，难以应对互联网平台暴露面风险;

（2）低零代码平台应用开发过程中，易忽略数据权限、业务权限管控，存在数据泄露、越权访问等安全隐患;

（3）港口物流、船舶运维等关键工业业务对连续性要求高，系统漏洞、异常访问等安全问题可能直接影响生产运营，甚至造成业务中断。因此，企业亟需一套覆盖网络、数据、业务的一体化安全防护方案，保障数智化转型过程中的平台安全、数据安全与业务安全。

4.4.2 建设方案

基于中国移动天工工业互联网平台的安全能力底座，为该企业及中小数转平台构建“评估-防护-运营”全流程安全体系。

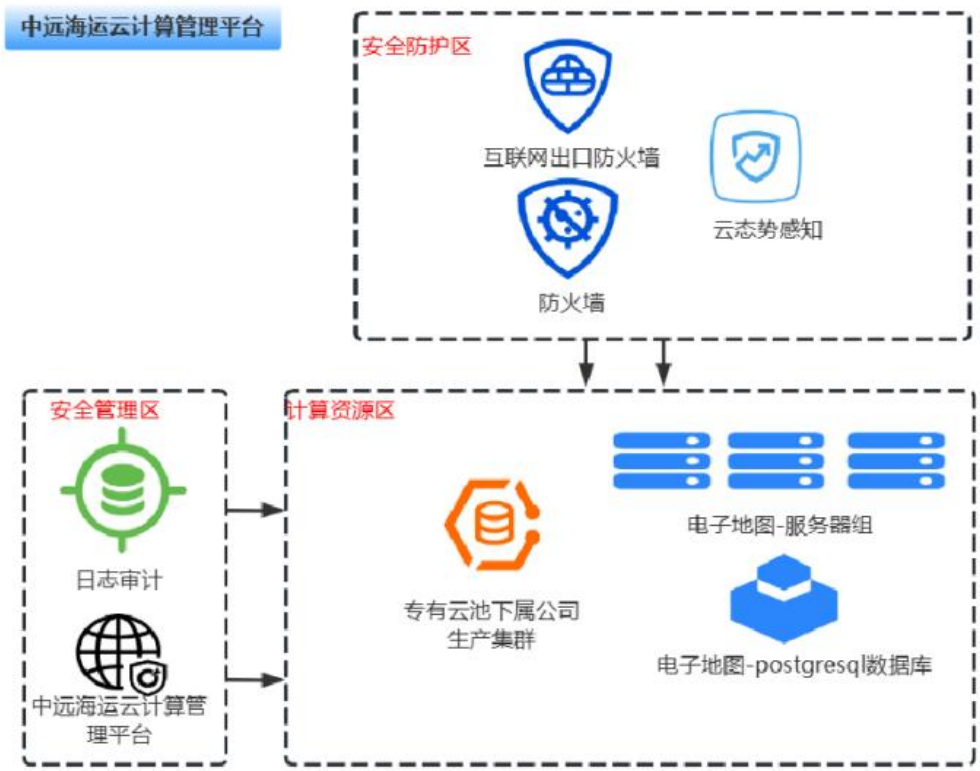

图 6 网络拓扑图

（1）安全测评与合规加固。为厂区智慧物流仓储、移泊助理、防台管理等关键系统提供等保三级测评、安全渗透评估、代码审计及组件安全评估服务，全面梳理系统漏洞与合规风险；结合测评结果，为低零代码平台制定数据安全基线配置手册与应用安全配置指导，完成权限分级、访问控制等基础安全策略配置。

（2）低零代码应用安全防护。采用“编译态/运行态隔离”设计，为低零代码应用提供数据权限、业务权限的双重

管控；配置数据访问控制、操作审计等安全策略，实现应用全生命周期的安全管理，提升被动安全防护能力。

（3）常态化安全监测与运营。为全国 22 个省市中小数转平台提供攻击面安全监测服务，持续扫描并处置弱口令、未授权访问等暴露面风险；建立常态化安全运营机制，定期开展复测、漏洞整改跟踪与安全报告输出，适配业务迭代升级需求。

4.4.3 实施方案

项目实施分为三个阶段，确保安全能力与业务同步建设、持续生效。

第一阶段：安全评估与基线加固。对该企业及中小数转平台的关键系统开展等保测评、渗透测试、代码审计，全面识别安全风险；制定低零代码平台安全基线配置手册，完成应用权限、数据访问控制的基础配置与加固。

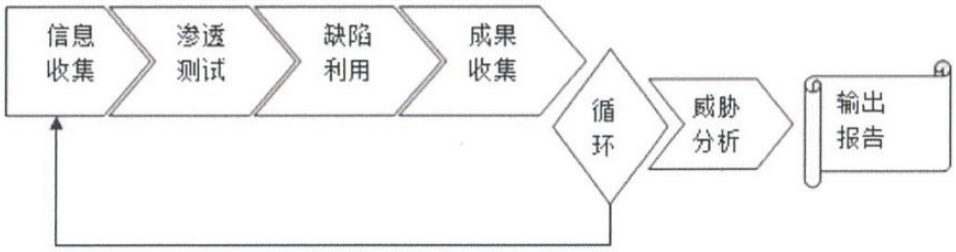

图 7 渗透测试流程

第二阶段：防护部署与策略落地。部署网络安全防护设备，构建厂区物流、船舶航运业务的安全边界；上线攻击面安全监测服务，对互联网平台的暴露面进行持续监控与告警；启用数据安全策略，实现低零代码应用的权限分级、操作审

计等功能。

第三阶段：常态化运营与持续优化。定期开展安全复测与渗透测试，跟踪漏洞整改闭环；持续监测攻击面变化，动态调整防护策略；提供常态化安全运营服务，定期输出安全报告与优化建议，适配业务发展需求。

4.4.4 建设成果

经济效益方面，为该企业各个中小数转平台构建了一体化安全防护体系，有效降低了业务中断、数据泄露等安全事件的发生概率，避免了潜在经济损失。

低零代码平台安全基线配置与权限管控，减少了后续安全整改、合规审计的投入，提升了数智化转型的投入产出比；常态化攻击面监测提前发现并处置多起潜在风险，保障了厂区物流、船舶运维业务的稳定运行，减少了业务停运损失。

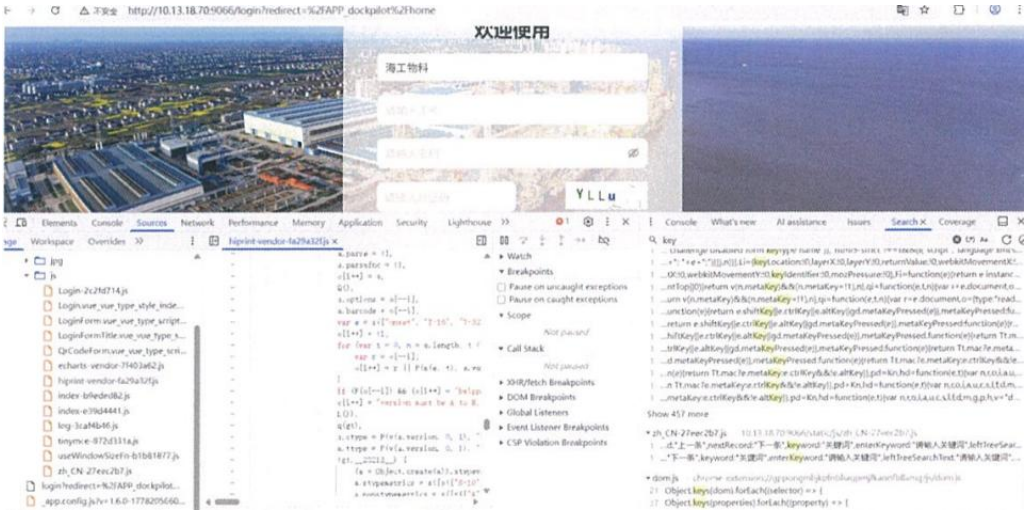

图 8 安全风险排查

社会效益方面，助力该企业打造安全可信的数智化标杆，为港口物流、船舶航运行业提供了可复制的安全防护经验。

为全国 22 个省市中小企业数字化转型提供了低成本、可落地的安全解决方案，降低了中小企业数字化转型的安全门槛；强化了低零代码平台、工业互联网平台的安全合规能力，推动了新型工业化背景下工业企业安全与发展的协同提升。

4.4.5 能力支撑

案例的核心支撑能力来自中国移动自研的天工工业互联网平台安全体系，覆盖“通、数、工”三大维度：

“通”层面：依托 5G 网络服务、工业互联网组网安全能力，提供边界防护、攻击面监测、安全审计等服务，适配工业场景与互联网平台的混合部署架构；

“数”层面：具备数据分类分级、权限管控、脱敏审计、全生命周期治理能力，可针对低零代码、工业平台的数据流转场景提供合规与安全保障；

“工”层面：拥有工业系统安全测评、渗透测试、组件评估、业务场景防护的专业服务能力，可为港口、物流、船舶等工业场景提供定制化安全解决方案。

4.5 铁路运输业人工智能赋能安全

4.5.1 背景与需求

随着《关键信息基础设施安全保护条例》的实施，铁路作为关基行业，需要具备“看见”威胁的能力。单纯依赖特征库匹配已无法应对零日攻击，亟需通过全流量分析结合人

工智能来弥补时间差。传统的铁路网络安全防御体系难以应对新型武器攻击、平台化武器库攻击和快速有效渗透等多元化强力攻击方式。

（1）提升主动感知主动防御能力

铁路既有安全措施对新出现的多样化安全威胁防范不足，特别是针对移动互联网、大数据、云计算等新技术应用，缺乏针对性的安全防护措施。铁路关键信息基础设施主要基于专网运行，运行状态监控手段缺乏，网络感知能力薄弱。除信息系统网络流量和日志监测覆盖全路外，监控范围未覆盖工业控制系统，未形成全路性网络安全态势感知能力、全景可视化能力和安全事件处理能力。

（2）强化威胁溯源与风险评估能力

铁路关键信息基础设施威胁攻击溯源及全生命周期风险管控分析不完善，安全关联分析能力存在短板。需要建立铁路威胁情报库、风险库，研究资产识别、威胁识别、脆弱性识别等威胁情报分析技术，建立完善风险评估算法模型。

（3）建立全面实时准确的资产台账

攻击路径排查中发现僵尸资产、无责权资产，说明铁路关键信息基础设施存在资产底数不明、台账不清问题。必须加强数字资产流动监控、资产分类分级及资产图谱画像、供应链安全管控，为开展资产梳理排查提供基础支撑。

4.5.2 建设方案

建设目标是构建一个“全时、全域、全流”的威胁感知体系。基于 AI 的全流量威胁感知系统整体架构主要包括数据获取层、预处理层、存储层、分析层和应用层，参考如下：

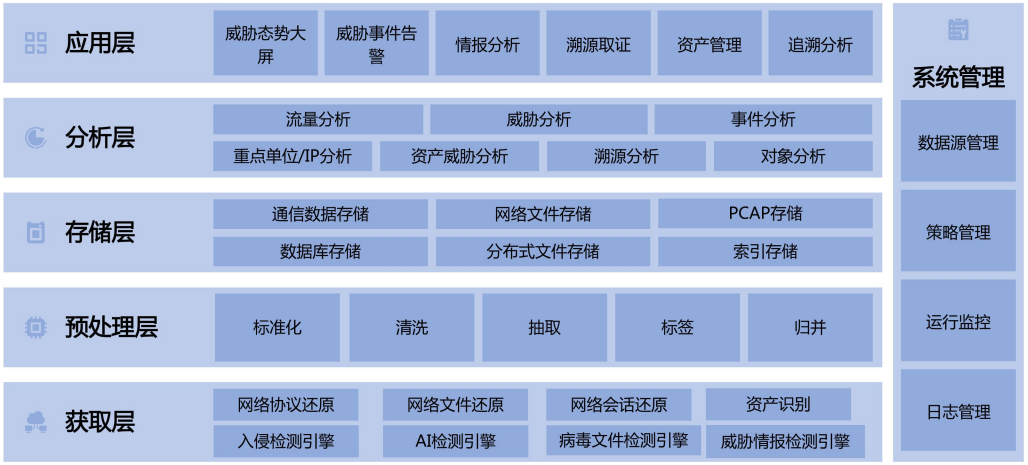

图 9 基于 AI 的全流量威胁感知系统逻辑架构图

系统中采用了人工智能技术对安全能力进行赋能，包括对加密流量进行分析处理，具体参考如下图所示：

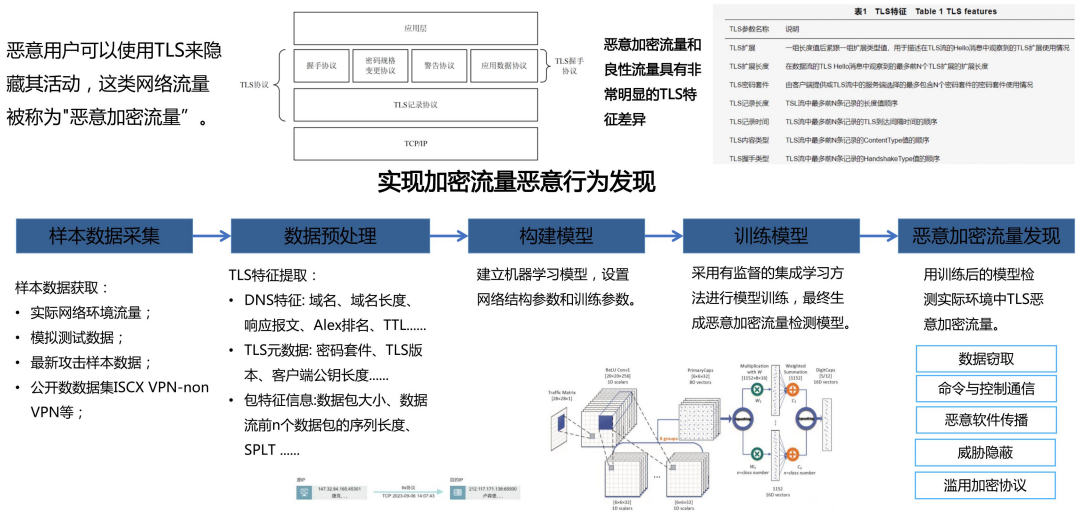

图 10 基于机器学习的恶意加密流量检测技术流程图

其次利用了人工智能技术发现隐蔽隧道，如下图所示：

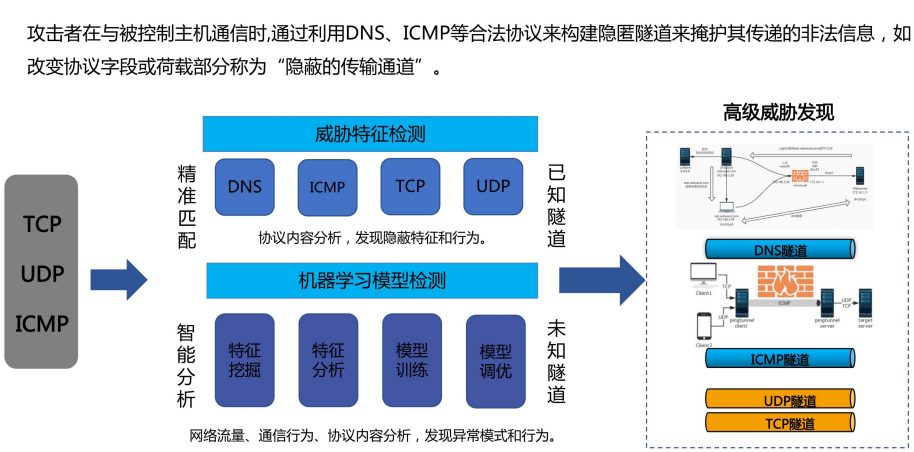

图 11 基于人工智能的隐蔽隧道检测机制示意图

另外，利用了人工智能技术发现 DGA 域名，如下图所示：

图 12 基于 AI 的 DGA 恶意域名检测方案

4.5.3 实施方案

实施路径遵循“先试点、后复制、全覆盖”的原则，逐步实现铁路网络的全覆盖，网络部署拓扑参考如下：

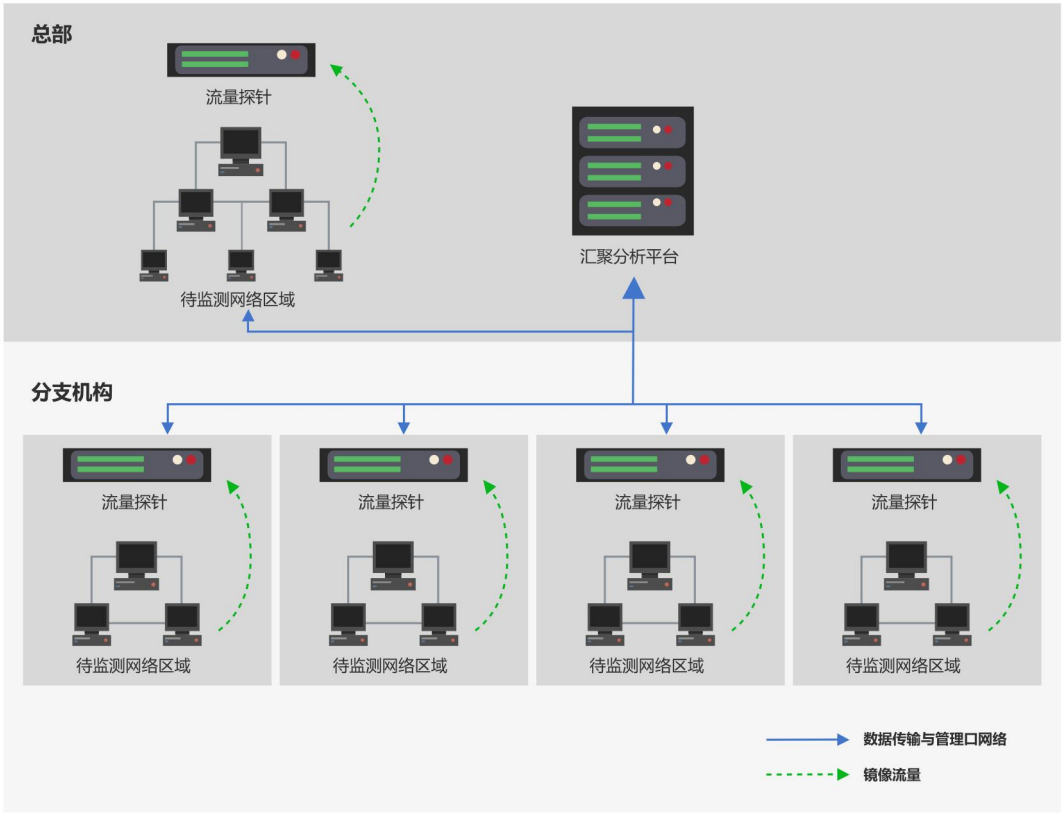

图 13 铁路全流量威胁感知系统“总部-分支”一体化部署拓扑图

分支探针：各分支单位在需要监测的网络节点部署探针，实现对网络流量的采集、还原，威胁检测发现、数据加密回传工作；

总部探针：在总部需要监测的网络节点部署探针，实现总部的通信感知、威胁感知、资产感知。

综合汇聚分析平台：在总部部署汇聚分析平台，实现数据汇聚、统一分析、集中呈现，探针的统一管理。

4.5.4 建设成果

（1）经济效益

降本增效：通过自动化威胁狩猎替代人工审计，减少了对大量安全分析人员的依赖。人工智能的低误报率特性极大

地降低了安全运维团队的时间成本。

保障运营连续性：有效阻断勒索病毒和蠕虫在网络中的横向移动，避免了因网络中断导致的列车大面积晚点或票务系统瘫痪，挽回了潜在的巨额经济损失。

（2）社会效益

提升公共安全：铁路是民生大动脉，通过感知系统保护列车控制指令的完整性与可用性，直接保障了亿万旅客的生命财产安全。

增强合规与信誉：建立了完善的关键信息基础设施安全防护体系，满足了国家对重点行业的监管要求，树立了“智慧铁路”的安全可信形象。

4.5.5 能力支撑

为了确保全流量威胁感知系统真正发挥实战价值，需要以下三大能力的支撑：

（1）算法与算力支撑。需要构建高性能的大数据仓库和 GPU 算力集群，以支撑深度学习模型的持续训练与推理。特别是在面对铁路“高通量”数据流时，平台必须具备近实时的处理能力。

（2）工控协议知识库支撑。铁路系统包含大量特有的工控协议。系统厂商需具备对信号系统、牵引供电系统等 OT 侧流量的深度解析能力，这是区分通用安全产品和铁路专用安全方案的关键。

（3）协同响应机制支撑。技术只是手段，流程才是保障。需要建立“安全运维中心-路局-站点”的三级联动应急处置机制，明确在人工智能研判出高危告警后的具体封禁策略与上报流程，避免“只见告警，不见处置”的问题。

4.6 AI+工业互联网防勒索安全

4.6.1 背景与需求

勒索病毒通过恶意加密关键数据逼迫受害者支付赎金，已成为工业互联网领域的头号网络威胁。某大型装备制造企业拥有数千台工业终端，核心系统承载大量工艺数据和设计图纸等关键资产，一旦遭遇攻击，将造成极为严重的后果。

传统杀毒软件依赖已知特征库进行检测，面对基于未知漏洞和快速变种的新型勒索病毒时防御效果极为有限。因此，该企业迫切需要构建覆盖“通算智数工”五维度的纵深防御体系，从根本上解决勒索病毒威胁。

图 14 工业防勒索多维体系逻辑架构图

4.6.2 建设方案

针对企业网络架构特点与核心业务连续性要求，构建覆盖事前预防、事中阻断、事后恢复的全链条防御能力。

“通”：网络传播路径阻断

勒索病毒常利用网络协议漏洞、弱口令爆破、钓鱼邮件等方式横向移动，一旦单点失陷便迅速蔓延。方案构建东西向流量感知与威胁阻断能力，实时监控共享文件异常访问，检测到远程加密行为后自动隔离失陷主机，实现网络层纵深防御。

图 15 防勒索软件部署架构图

“算”：轻量化边缘代理架构

工业终端对资源占用极为敏感。方案采用轻量化边缘代理架构设计，精简功能模块、优化检测引擎，实现计算资源极低消耗，保障核心业务无感运行。

“智”：AI 行为分析引擎

方案构建覆盖应用层、系统层、内核层的多层级行为感知体系，从文件操作模式、进程调用链、内存行为特征等多维度综合研判，实现对勒索病毒活动的精准识别与自动化处置。同时精心布设的诱饵文件能很大程度上先行捕获勒索病毒。

“数”：数据安全保护体系

方案内置智能动态备份技术，检测到文件篡改时自动创建临时备份，事后支持一键全量或按需恢复，最大程度缩短业务中断窗口。同时支持与企业现有备份设施联动，自动区分干净与污染数据，并提供密钥捕获与外泄监测能力，防范双重勒索。

“工”：异构工控环境适配

方案实现异构工控环境的广泛适配，兼容主流信创操作系统和国产 CPU 平台，信创与非信创终端统一管控、平滑过渡。

4.6.3 实施方案

本案例采用分阶段落地策略，覆盖前期摸底与策略制定、中期轻量化部署、后期日常运营与灾难恢复，形成完整闭环。

第一阶段：环境摸底与策略制定

项目团队对全部 IT 与 OT 资产进行全面摸底，建立包含操作系统版本、CPU 架构、网络分区、业务等级等信息的资

产台账，并对网络拓扑实地勘察，识别勒索病毒横向移动的关键路径。在此基础上综合评估各终端风险等级，将终端划分为核心防护级（如 MES、PLM、SCADA 服务器）、标准防护级（设计工作站、办公终端）和基础监控级（信息屏等）三级，制定差异化防护策略矩阵，并对物理隔离区域制定离线更新流程。

第二阶段：轻量化部署实施

遵循“先验证、后推广”原则，选取非生产区典型终端试点、稳定后分批推进。针对工控上位机和老旧工控站采用极简部署模式，仅加载核心检测引擎和通信模块，安装包极小、无需重启。

上线初期以“仅记录、不阻止”模式运行一周积累基线，策略调优后再切换防护模式，杜绝误报对生产的影响。

第三阶段：日常运营与灾难恢复

上线后建立日常安全运营机制，统一控制中心实时展示全网安全态势，建立告警分级响应确保关键威胁不过夜，并按月生成安全运营报告为管理层提供决策依据。

4.6.4 建设成果

系统上线运行一年来，从未影响任何业务系统的运行。并且，在内部的勒索病毒攻击演练中，成功识别并阻断多次横向移动攻击；成功检出最新的勒索病毒样本，并且在 15 分钟内完成全部加密文件恢复。

4.6.5 能力支撑

本案例体现了“通、算、智、数、工”五大核心能力的协同支撑。即:

通过分布式管控架构与加密传输实现安全通信;

通过轻量化系统运行架构实现极低的资源占用;

通过 AI 多维行为分析与智能诱饵技术实现对勒索病毒和攻击行为的精准检测;

通过动态备份、污染识别与快速恢复构建数据安全防线;

通过信创兼容、离线适配与分级管控实现工业场景全面落地。

五、展望

5.1“通算智数工”体系化能力价值总结

人工智能与工业互联网的深度融合，正在重塑工业系统的安全边界与防护逻辑。“通算智数工”五维一体融合安全能力体系，是面向人工智能与工业互联网融合发展实践、结合行业安全建设共性需求提炼形成的体系化框架：以安全原生替代外部叠加，以跨层协同打破能力孤岛，以智能闭环提升运营效率，将通信网络、算力底座、智能应用、数据流通与工业生产五个维度的安全能力统一纳入一体化运营底座，形成从感知到处置的完整防护链路。

从实践来看，这一体系在智慧水运大模型实践、港航数智化平台防护、工控纵深防御、5G 融合安全运营等场景中均

取得了可验证的落地成效，初步证明了体系化安全建设路径相较于单点产品堆叠的综合优势。当然，融合安全体系的建设并不存在一劳永逸的解决方案。随着工业智能化持续深入，模型能力演进与安全威胁变化将长期并存，体系必须在新技术、新场景与新风险的动态演进中不断迭代完善，以持续增强整体防护能力与适配能力。

5.2 共建工业智能融合安全生态倡议

为深入贯彻《行动方案》关于“强化安全保障、共建协同生态”的战略部署，我们诚挚携手各网络安全领军企业、工业骨干企业、科研机构与产业链伙伴，共同发起“共建工业智能融合安全生态”倡议，筑牢新型工业化高质量发展的安全底座。

为此，我们向全行业发出如下倡议:

（一）坚守安全初心，共筑融合发展底线

贯彻总体国家安全观，将安全贯穿工业互联网与人工智能融合应用全生命周期，严格落实安全分类分级管理要求，健全风险防控机制，坚守不发生系统性重大安全事件的底线。

（二）共建情报体系，共享安全威胁数据

搭建工业智能安全威胁情报共享平台，开放共享漏洞、攻击样本、威胁趋势等核心情报，依托 AI 技术强化分析研判能力，实现“一点发现、全网预警、协同处置”。

（三）协同技术创新，攻克核心安全难题

组建产学研用协同创新联合体，联合攻关工业协议安全、AI 模型安全、数据安全等核心技术，构建“主动防御、动态防护、智能响应”的新型安全技术体系。

（四）深化生态协同，构建共治共享格局

构建“政府引导、企业主导、机构协同”的工业智能融合安全生态，推动安全能力下沉工业场景，鼓励大中小企业融通发展，共享技术、情报与服务资源。

（五）强化人才培养，夯实安全人才支撑

联合搭建工业智能安全人才培养体系，开展实战化人才培养，培育复合型专业人才，缓解行业安全人才短缺困境。

（六）完善应急机制，提升协同处置能力

建立安全事件应急响应联动机制，定期开展跨企业、跨行业应急演练，加强事件信息通报与经验交流，保障工业生产稳定运行。

生态共建，安全共享；融合赋能，智护工业。工业智能融合安全生态建设事关新型工业化高质量发展大局，是全行业共同的责任与使命。我们坚信，各方凝心聚力、协同共治，必能构建起全方位、一体化的工业智能融合安全保障体系，为工业互联网与人工智能深度融合保驾护航!