中国移动
China Mobile

AI+工业互联网
安全解决方案白皮书

中国移动通信集团有限公司
2024年11月

发布单位：中国移动通信集团有限公司

牵头编写单位：中国移动信息安全管理与运行中心、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）：启明星辰信息技术集团股份有限公司、工业互联网创新中心（上海）有限公司、北汽福田汽车股份有限公司、上海赛博网络安全产业创新研究院、上海观安信息技术有限公司、上海嘉韦思信息技术有限公司、北京亚鸿世纪科技发展有限公司、上海交通大学、《中国信息安全》杂志社

协助单位：上海市工业互联网协会、中移智库、安在新媒体

编写组主要人员：

中国移动信息安全管理与运行中心：袁捷、张峰、江为强、邱勤、董航、于乐、郭中元、王光涛、王国宇

中移（上海）信息通信科技有限公司：陈豫蓉、乔建设、路骁虎、周威、唐双林、赵威、李峻天、王轩轩、周子涔、郭清华、张玉欣、付超、任咏、程元森、仲其伟、代硕勋、潘永高、刘梅华、程赟、马兵、陈玉娟、岳峰

启明星辰信息技术集团股份有限公司：吴道虎

工业互联网创新中心（上海）有限公司：郑忠斌、秦峰、杨俊、张洋

北汽福田汽车股份有限公司：王鹤、张丽

上海赛博网络安全产业创新研究院：张喆、王佳雯

上海观安信息技术有限公司：王文君， 阮子禅

上海嘉韦思信息技术有限公司：舒首衡、郗上才、何升文

北京亚鸿世纪科技发展有限公司：林飞、程红

上海交通大学：侍国亮、周志洪、银鹰

《中国信息安全》杂志社：位华、李刚

目录

“AI+工业互联网”发展概述及主要应用场景

1.1“AI+工业互联网”发展现状

人工智能（AI）与工业互联网的结合正引领着第四次工业革命，通过机器学习算法优化的自动化生产线，工业互联网作为新一代信息技术与制造业深度融合的产物，正推动着工业制造、石油化工、矿山冶金、电力能源等多个领域向智能化、数字化转型。在工业制造领域，AI 技术通过智能监控、精细化管理、质量控制等手段，提升生产效率和产品质量，降低成本和风险，同时促进创新和发展，增强企业竞争力。在石油化工领域，利用 AI 进行研发创新、生产效能提升和安全环保治理，实现生产过程的优化和环境的可持续发展。在矿山冶金领域中，AI 技术的应用覆盖了资源勘探、生产过程优化、安全管理等全流程，提高矿产资源开发的效率和质量，推动精准采矿和工业安全管理的进步。在电力能源领域，通过 AI 实现电力系统的优化调度、新能源发电预测、智能运维和虚拟电厂管理，确保电力供应的稳定性和可靠性。总体来看，工业互联网的发展正通过 AI 技术的应用，为传统行业带来革命性的变化，不仅提高了生产效率和安全性，还促进了资源的优化配置和环境的可持续发展，展现出巨大的潜力和广阔的应用前景。随着技术的不断进步和应用的深入，预计 AI 将在工业互联网中发挥更大的作用，推动工业产业的高质量发

展。

1.2“AI+工业互联网”主要应用场景

AI 技术正日益深入应用于工业制造、石油化工、矿山冶金、电力能源等多个工业领域，成为工业互联网场景智能化的关键驱动力。这些技术的应用不仅显著提升了工业企业的生产效率，而且加速了企业的数字化转型进程。同时，它们还促进了整个产业的升级，提高了整体的运营效率和竞争力。

1.2.1“AI+工业制造”场景

在工业制造领域中，AI 技术的应用场景涵盖了制造业从生产优化到安全管理的多个方面:

智能监控与预测性维护: AI 技术通过大数据分析和机器学习算法，实时监控工业设备的运行状态，预测设备可能出现的故障，并提前进行维护，减少设备停机损失并提高使用寿命和效率。

精细化生产流程管理: AI 技术对生产流程进行智能优化，分析生产过程中的瓶颈和问题，提出改进方案，实现生产资源的合理配置，提高生产效率，降低成本。智能质量控制与检测: AI 技术收集生产过程中的数据，分析并预测产品质量趋势，及时发现潜在问题，提升产品质量水平，增强企业竞争力。

产品设计与生产制造：AI 技术在产品设计环节提升设计仿真度，提高设计效率和准确性，助力产品快速迭代。并加强信息实时收集、处理、执行能力，通过赋能智能排产、设备管理、质量管控、仓储配送等环节，提高生产质量并节约成本。

智能化运营管理：AI 技术在供应链管理、销售预测、市场营销等细分场景提升管理工作效率，帮助制造企业构建以用户为中心的经营模式。

1.2.2“AI+石油化工”场景

AI 技术在石油化工领域的应用场景涵盖研发、生产、管理等各个环节，不仅可以提高生产效率和产品质量，降低成本和风险，还可以促进企业的创新和发展，提升石化企业的竞争力。

助力化工研发技术创新：在分子设计与合成方面，AI 技术可以通过对大量化学数据的学习和分析，预测和设计新的分子结构，加速新材料和新化学品的研发进程。例如，利用深度学习算法，可以模拟化学反应过程，预测反应产物和最优反应条件，从而减少实验次数和成本。在催化剂开发与优化方面，AI 技术帮助筛选和优化催化剂，提高化学反应的效率和选择性。通过分析催化剂的结构、性能和反应条件等因素，AI 技术可以预测催化剂的活性和稳定性，为催化剂的设计和改进提供指导。

助力化工生产效能提升：利用机器学习算法预测化工反应的产物，优化生产工艺参数，减少废品率。实时监测生产过程中的各项指标，如温度、压力、流量等，及时发现异常情况并进行调整，确保生产过程的稳定性。同时，通过 AI 技术对设备运行数据的分析，可以预测设备的故障和维护需求，提前安排维护计划，降低设备故障率和非计划停机时间。例如，利用传感器采集设备的振动、温度、电流等数据，通过机器学习算法进行分析，预测设备的故障类型和发生时间。

加强生产安全管理与环保治理：利用 AI 图像识别技术，可以自动检测生产现场的安全设施是否完好，员工的操作是否符合规范，对生产过程中的安全风险进行评估和预警，及时发现潜在的安全隐患并采取措施进行处理。此外，监管部门可以通过 AI 自动化监测企业的污染排放情况，预测污染物排放的趋势和影响，为企业制定环保措施提供依据。例如，利用传感器采集废气、废水等污染物的数据，通过机器学习算法进行分析，可以预测污染物排放的浓度和变化趋势。

1.2.3“AI+矿山冶金”场景

AI 技术在矿山冶金领域的应用正变得越来越广泛，它通过提高效率、降低成本、增强安全性和优化决策过程，为这一传统行业带来了革命性的变化。

智能安全分析识别管控：基于AI安监产品，结合大模型、定位、物联网等技术，建立对矿区、冶金园区与作业现场的人员、设备、环境进行违规行为、危险源等危险要素的视频识别与融合管控，确保人机环管安全合规，具备危险告警后的系统联动处理与应急处置能力，基于多模态大模型能力，实现通过交互回答方式回溯告警事件、生成监测报告，提升矿冶安全监查和应急处置效率。

智能化设备预测性维护：设备故障智能分析诊断、故障预测等依托设备机理模型、故障模型与实时数据通过机器学习方式进行训练计算，依托人工智能大模型升级现有系统，结合设备参数、设备异常数据等训练未来趋势的判断，结合设备知识库、专家经验等新增设备根因分析、获得知识性问答、健康状态评估、辅助决策等功能，提升系统算法场景覆盖度和整体检测精度，增加智能交互体验。

智能井下作业巡检：矿山巡检员借助本安型手机APP进行井下安全巡检，如机电硐室配电装置指示灯是否正常，人工智能大模型可基于作业现场视频拍照自动识别违规行为，生成违规巡检卡，智能推送干系人或下发工单，简化操作，缩短整改时间。

皮带机预测性维护与管控：带式输送机是矿冶生产中十分常用的物料转运设备，造价高、运输数量大、速度快，作业过程中容易造成皮带跑偏、打滑、划破、撕裂、

磨损等问题，而常规人工运维管理又存在监测不及时、效率低、招工难的问题，针对场景痛点，基于安监大模型，结合感知数采终端，基于数字化与图像 A1 技术对皮带机组进行监测分析与管理控制，提供皮带机组的智能监测与预测性维护方案覆盖皮带物料识别、皮带打滑跑偏、撕裂监测等安全监管功能，有效防护皮带机的作业安全、设备资产安全、提升故障处理效率、降低故障损失。

钢制品智能质量检测：传统冶金产线普遍采用人工目视或抽样的检测方式实现产线产品质量的检验，这种检测方式过度依赖人工，检测率低，漏检错检可能性高，对产线的把控力不强，工人在产线旁易产生安全事故，针对此情况，利用 5G 网络低时延、大带宽、高可靠的特性，结合人工智能机器视觉技术，实时采集传输多个检测点的表面高清图像至 MEC 边缘云平台进行算法比对，并下发指令对产线上的钢成品进行质量检测，控制重复缺陷的持续产生，提高产品质检效率和精度。

1.2.4“AI+电力能源”场景

在电力能源领域，AI 技术主要覆盖以下几个关键应用场景：

电力系统优化调度：AI 技术通过分析电网的实时数据和预测数据，能够提高电力调度的效率和准确性。例

如，南方电网推出的调度云超算平台，利用 AI 技术进行功率预测和实时控制调度，显著提升了电力决策效率。

新能源发电功率预测：AI 技术在新能源发电领域，如风电和光伏发电中，通过分析气象数据和历史发电数据，优化发电功率预测模型，提高预测速度和准确度，从而保障电网的安全运行和电力的可靠供应。

智能运维与巡检：AI 技术的应用使得电力设备的巡检工作更加智能化和自动化。例如，使用无人机搭载高清摄像仪和红外传感器，完成对输电线路和设备的运行状态监测和安全评估，提高了巡检的效率和准确性。

虚拟电厂和微电网：AI 技术在虚拟电厂（VPP）中发挥重要作用，通过聚合分布式清洁能源、可控负荷和储能系统等资源，作为一个“虚拟”的电厂参与电力市场和电网运行，实现电力资源的优化配置和智能调度。

电力系统稳定评估与决策：AI 技术在电力系统稳定评估中应用，通过深度学习和强化学习等高级机器学习技术，对电力系统的稳定性进行实时评估和决策，提升电网的安全和调控能力。

这些应用场景展示了 AI 技术如何帮助电力行业提升效率、降低成本，并实现更安全、更智能、更环保的电力系统运营。随着技术的不断进步，预计 AI 在电力能源领域的应用将更加广泛和深入。

1.3“AI+工业互联网”发展中存在的问题

在“AI+工业互联网”中，人工智能技术的应用正带来革命性的变化，同时也伴随着一系列安全风险。数据泄露和隐私侵犯成为主要问题，因为 AI 技术依赖大量数据，一旦保护不当，可能导致敏感信息外泄，给企业带来损失并对运营造成影响。AI 算法本身的安全性也是关注的焦点，算法中的漏洞可能被攻击者利用，影响生产流程，甚至引发更严重的安全事件。随着工业互联网的发展，大量智能设备的接入增加了网络安全风险，这些设备可能成为攻击的切入点，威胁整个工业网络的安全。工业控制系统也面临着新型攻击方式的挑战，需要更加严格的安全措施来保护系统不受虚拟机逃逸和跨虚拟机侧信道攻击等威胁。此外，平台数据安全风险涉及数据在各个阶段的安全，包括数据的侦听、拦截和篡改等问题。技术成熟度和数据可用性风险、对抗性攻击、系统漏洞风险以及供应链攻击风险都是工业大模型应用中需要重视的安全问题。这些风险要求我们在享受 AI 技术带来的便利的同时，也要加强对安全防护措施的投入，确保工业制造行业的数据安全和系统稳定运行，以应对日益增长的网络安全威胁。

“AI+工业互联网”安全风险

“AI+工业互联网”安全风险主要体现在人工智能自身

的安全，特别是以工业大模型为代表的工业领域机器学习模型安全风险，其次是 AI 技术赋能工业互联网各个场景的安全风险。

2.1 工业互联网大模型安全风险

在当今这个信息化飞速发展的时代，工业大模型应用已经成为工业领域提效新质生产力的重要方向，并在自然语言处理、图像识别、预测分析等多个领域展现出了惊人的能力，极大地提升了生产效率。然而，随着应用范围不断扩大，其安全风险也日益凸显。工业大模型自身的安全风险主要集中在以下几个方面:

工业业务服务安全风险:

工业大模型在服务中可能会在没有适当监督的情况下生成不准确或有害的内容，比如涉及恐怖主义、种族歧视、黄色暴力等不当信息，这不仅违反法律法规，还可能对社会稳定和公共安全构成威胁。此外，工业大模型在应用过程中可能与工业安全标准和最佳实践不一致，导致模型输出与实际工业操作要求不匹配，增加操作风险和事故概率。

工业AI技术滥用风险:

工业大模型在训练过程中会接触大量的工业真实数据，容易被利用生成较为真实的工业事故场景画面，进而造成社会恐慌，危害国家社会稳定。

工业模型算法安全风险:

工业大模型依赖的AI算法可能在设计时未能充分考虑其鲁棒性、公平性和可解释性，存在潜在的安全漏洞。这些漏洞可能被攻击者利用，通过逆向工程提取模型信息，从而威胁到大模型的安全性和可靠性。因此，确保算法在这些关键属性上的健壮性对于维护工业大模型的整体安全至关重要。

数据要素安全风险:

工业大模型在训练过程中需要大量数据，例如：工业设计图纸、工业设备控制数据等。这些数据可能包含生产核心机密。如果发生数据泄露事件可能导致企业合法权益被侵犯，甚至遭受工业生产系统的破坏。

平台安全风险:

工业大模型的开发和使用可能依赖于多个供应商提供的组件和服务，这些组件和服务的开发者在开发过程中欠缺安全风险意识，从而使这些组件和服务本身就具有一定的安全风险，进而通过供应链安全风险威胁工业大模型安全。

2.2“AI+工业互联网”主要场景安全风险

在工业互联网中，AI 技术应用带来了显著的效率提升和流程优化，但同时也伴随着安全风险，需要监管方、

协会、企业及个人共同努力，完善法规政策、加强技术研发、提高安全意识，以确保工业互联网健康安全的发展。

2.2.1“AI+工业制造”场景安全风险

在“AI+工业制造”场景中，AI 技术应用后带来的安全风险主要体现在以下几个方面。

数据泄露与隐私侵犯风险：AI 技术的核心是数据，数据安全保护措施不到位可能导致数据泄露，给企业带来巨大损失。同时，攻击者可能利用泄露的数据进行恶意攻击，严重影响企业的正常运营。

AI 算法的安全隐患：AI 算法可能存在安全漏洞，攻击者可能利用算法的缺陷进行攻击，干扰设备的正常运行，甚至篡改控制逻辑，造成严重后果。算法的黑箱特性也增加了安全风险，例如数据泄露、算法歧视或算法滥用。

智能设备的安全风险：随着工业互联网的普及，越来越多的智能设备接入网络，这些设备的安全防护能力参差不齐，容易成为攻击者的突破口，攻击者可以通过控制这些设备，进而对整个工业网络进行攻击，造成严重的损失。

工业控制系统安全风险：工业控制系统在 AI 技术的赋能下，面临着虚拟机逃逸、跨虚拟机侧信道攻击、镜像篡改等新型攻击方式的威胁。

平台数据安全风险：平台数据安全涉及接入平台、平台运行、平台退出三个阶段中的数据安全，包括数据侦听、

拦截、篡改、丢失、窃取等安全风险。

2.2.2“AI+石油化工”场景安全风险

在“AI+石油化工”场景中，AI 技术应用后带来的安全风险主要体现在以下几个方面。

数据泄露风险：石油化工行业涉及大量敏感数据，包括生产工艺参数、控制指令信息、员工健康信息等。AI 技术的应用需要处理这些数据，存在数据泄露和滥用的风险。

网络攻击风险：随着 AI 技术在石油化工行业的应用，网络攻击的风险增加，包括 APT 攻击、软件供应链攻击等，可能导致关键基础设施的破坏和生产事故。

数据质量与可靠性风险：人工智能算法在石油炼化过程中的原料供应链风险评估中应用，但存在数据质量和可靠性、模型的解释性和可操作性等挑战。

2.2.3“AI+矿山冶金”场景安全风险

在“AI+矿山冶金”场景中，AI 技术应用后带来的安全风险主要体现在以下几个方面。

数据采集与处理风险：通过物联网设备实时获取矿山各类环境和生产数据，包括设备状态、人员位置、气体浓度等，这些数据的采集和处理需要保证及时性和准确性，不一致或过时的数据可能导致决策混乱和滞后。数据采集不仅限于单一来源，而是通过多维度、多渠道的数据获

取，可能会面临数据泄露和篡改的风险。

分析与预测风险：传统矿山在应对突发事件和管理风险方面面临诸多困难，比如数据共享不足、决策效率低、风险预判能力欠缺等，难以识别出潜在的风险隐患和未来可能发生的灾害事件。

风险评估与决策支持风险：矿山企业在实施风险评估系统时，可能会遇到数据整合不充分的问题，这会影响到风险预测模型的精准性和风险管理体系的标准化。面对矿山环境的复杂变化，现有的预测模型可能无法准确预测并有效管理风险，这对于矿山安全管理构成挑战。

技术依赖风险：随着 AI 技术的应用，矿山安全管理越来越依赖于技术，这可能导致对技术的过度依赖，从而在技术出现故障或误判时，增加安全风险。

网络安全风险：AI 系统的引入可能会使矿山网络面临更多的网络安全威胁，如黑客攻击、数据泄露等，这些安全问题可能会影响矿山的正常运营和生产安全。

设备和控制安全风险：AI 技术的应用可能需要与现有的设备和控制系统进行集成，这可能引入新的安全漏洞，如未经授权的访问和控制。

人员安全风险：AI 技术的应用可能会改变工作流程和操作方式，需要对工作人员进行新的培训和教育，以确保他们能够安全地使用新技术。

2.2.4“AI+电力能源”场景安全风险

在“AI+电力能源”场景中，AI 技术应用后带来的安全风险主要体现在以下几个方面。

数据安全与隐私保护：电力系统的安全不仅关系到社会稳定，还涉及军事国防安全，需要加强数据管理及隐私安全保护，防止数据泄露。在 AI 模型的训练和测试过程中可能会造成模型与数据隐私泄漏，需要采用数据隐私保护措施，如模型结构防御和信息混淆防御。

单一任务决策限制：现有的 AI 模型往往只能针对单一任务进行决策，缺乏“多任务”模型，这限制了 AI 技术在电力系统中的广泛应用。

鲁棒性缺乏：攻击者可能通过对输入样本添加微小的异常扰动，使模型输出错误的预测结果，影响电力系统的运行和安全。

“AI+工业互联网”安全风险治理方案

为落实工业互联网中的人工智能安全治理相关要求，保障工业互联网人工智能全流程安全可控，在国家工业互联网安全标准体系和《中国移动人工智能安全白皮书》的指导下，结合工业互联网领域的 AI 安全实践经验，构建“1266”中国移动“AI+工业互联网”人工智能安全体系架构。如图 1 所示。

图 1：“AI+工业互联网”人工智能安全体系框架

“1”是指规划一个“AI+工业互联网”工作体系架构。

“2”是指着力两个工作发力方向，即“工业互联网 AI 安全风险防范”和“AI 赋能工业互联网安全”。

第一个“6”是覆盖安全管理及安全技术的工业互联网 AI 安全防范举措，包括“AI+工业互联网”安全运营管理、工业 AI 业务服务安全、工业 AI 技术合规、“AI+工业互联网”算法安全、“AI+工业互联网”数据要素安全、“AI+工业互联网”平台安全。

第二个“6”是指 AI 赋能工业互联网数据安全、应用安全、网络安全、控制安全、设备安全和平台安全六大安全领域。

通过上述措施实现“让工业互联网更安全，让工业互联网安全更智慧，让工业生产更高效”的工业互联网安全愿景。

3.1 总体目标

“AI+工业互联网”安全首先要建设完善的**安全体系**，通过安全运营管理进一步强化**安全责任**；其次要防范工业AI业务服务安全、工业AI技术合规、“AI+工业互联网”算法安全、“AI+工业互联网”数据要素安全和“AI+工业互联网”平台**安全风险**；再次要通过AI赋能工业互联网数据安全、应用安全、网络安全、控制安全、设备安全和平台安全来提升工业互联网**安全能力**，提高AI在工业制造、石油化工、矿山冶金、电力能源等工业互联网场景中安全性和可靠性；最后坚持安全防护基本原则，确保工业企业高效且井然有序地**安全生产**。

3.2 安全防护基本原则

统一领导，分级管理：在《加强工业互联网安全工作的指导意见》（工信部联网安〔2019〕168号）中提出了筑牢安全，保障发展的原则，强调安全与发展并重，确保工业互联网安全和发展同步规划、同步建设、同步运行。这意味着在“AI+工业互联网”的安全防护中，需要有一个统一的领导机构来统筹规划和管理安全事宜，同时根据不同

的业务需求和风险等级，实施分级管理，以确保重点领域和关键环节得到有效的管理和防护。

AI 安全三同步：根据《加强工业互联网安全工作的指导意见》，工业互联网安全和发展应同步规划、同步建设、同步运行，这与 AI 安全的“同步规划、同步建设、同步运行”原则相呼应。在 AI 技术的应用过程中，安全措施需要从一开始就被纳入规划，确保在技术发展的同时，安全防护措施也能得到相应的发展和完善。

生态合作，协同发展：在全国网络安全标准化技术委员会发布的文件《人工智能安全治理框架》中提到了开放合作、共治共享的原则，强调了多方参与和共同治理的重要性。在 AI+工业互联网的安全防护中，需要不同利益相关方，包括政府、企业、科研机构和公众等，共同参与到安全治理中来，通过合作形成共识，共同提升安全防护能力。

中国移动“1264”人工智能安全原则：一是人工智能安全风险防控技术，通过整合 IPDRR 各阶段的安全技术，覆盖“1264”中的 6 个 AI 安全风险防控领域，即基础平台、数据要素、模型算法、业务服务、防范滥用；二是人工智能赋能网信安全技术，通过大小模型协同，赋能“1264”中的 4 个安全领域，即基础网络安全、数据安全、内容安全、业务应用安全。这一原则强调了在 AI 技术的应用中，

需要从多个维度出发，构建一个全面覆盖的安全防护体系，确保 AI 技术的可信、可控与可靠。

3.3 工业互联网 AI 安全风险防范

针对工业互联网人工智能技术应用与平台系统自身存在的安全风险，“AI+工业互联网”安全运营管理、工业 AI 业务服务安全、工业 AI 技术合规、“AI+工业互联网”算法安全、“AI+工业互联网”数据要素安全和平台安全六个方面，按照下文所定义管理与技术防护措施，实现工业互联网人工智能应用全流程安全可管可控可信。

3.3.1“AI+工业互联网”安全运营管理

安全政策规范：根据工业和信息化部发布的《“工业互联网+安全生产”行动计划（2021-2023 年）》要求，企业应将工业互联网与安全生产同规划、同部署、同发展，并构建基于工业互联网的安全感知、监测、预警、处置及评估体系，提升工业企业安全生产的数字化、网络化、智能化水平。这要求企业在制定网络与信息安全及生产安全管理制度中，必须将人工智能安全纳入考量，确保企业安全管理与 AI 技术同步发展，以促进安全生产水平的持续提升。

安全管理组织：工业企业应当建立专门的安全管理组织，负责监督和执行安全政策，确保工业互联网环境下的

安全风险得到有效管理。包括建立安全生产监管平台，实现安全生产全过程、全要素、全产业链的监管。

人员安全管理：人员安全管理涉及提升从业人员的安全意识和技能。工业企业应实施全员 AI 知识普及与技能培训，提高员工对 AI 的理解与接纳程度，消除对新技术的陌生与抵触。同时，倡导开放、包容、创新的企业文化，鼓励员工主动学习，营造积极变革、创新的良好氛围。

AI 风险管理：AI 风险管理是在现有的工业企业风险管理体系中嵌入相关风险管控。企业需要对人工智能模型开展算法备案、安全评估、大模型上线备案等工作。安全评估包括通用安全、设计开发安全、测试安全、部署与运行安全、退役安全等，企业必须确保 AI 模型的公平、透明、负责任。

AI 事件管理：AI 应急处置能力聚焦事前演练排查和事中快速响应能力，通过制定多层平台联动框架和标准，指导解决方案团队建设工业安全生产事件案例库、应急演练情景库、应急处置预案库等，并基于行业级、企业级监管平台建设系统风险仿真、应急演练和隐患排查能力。

AI 技术管理：人工智能技术在工业互联网安全的应用体现在主动防御、威胁分析、策略生成、态势感知、攻防对抗等多个方面。企业需要利用 AI 技术，通过智能算

法对原始数据进行预处理，降低安全分析人员数据处理压力，辅助安全分析人员判断。

3.3.2 工业 AI 业务服务安全

工业安全策略：在工业大模型业务服务中，确保其安全性和价值观对齐至关重要。这主要通过两个核心策略实现：一是提高训练数据的安全性，二是改进训练算法。数据安全性是确保模型输出符合社会主流价值观的基础，因此，需要采用数据脱敏、去标识化和数据掩码等技术来保护个人隐私和防止敏感信息泄露。此外，优化训练算法也是关键，通过基于反馈的方法和对抗训练来增强模型的鲁棒性，并利用知识融入训练来减少模型错误，确保其输出符合人类期望。这些措施共同作用，旨在使人工智能技术对社会产生积极影响。

输入输出安全：大模型的输入输出安全主要包括涉及防御性提示设计和对抗性提示检测的输入模块安全，消除毒性与偏见、幻觉的缓解以及防御模型攻击的模型模块安全，应用检测，干预和水印技术的输出模块安全。

为了防范这些安全风险，可以采取以下措施：

提升问题安全检测过滤能力：采用启发式检测方法和黑名单与白名单机制，快速过滤掉潜在的恶意输入。

增强安全语义分析引擎：利用自然语言处理技术对输入问题的语义进行深度理解，识别其背后的意图和潜在风险。

构建多维度安全检测模型：结合问题的多个特征，如模糊度、长度、关键词、语义等，构建综合安全检测模型。

加强时间敏感性检测：通过时间戳分析和核心意图提取，判断问题是否属于潜在的恶意攻击。

内容输出安全合规性再检测：在模型生成输出后，通过后处理机制对输出内容进行再次检测，确保其符合安全合规性要求。

优化分词方式与困惑度分析：针对特定语言特点，优化分词算法和分词粒度，提高模型对输入问题的理解能力。

构建关键词特征库：建立包含敏感词汇和关键词的特征库，对输入问题进行快速关键词检测。

模型再训练与微调：通过对抗性训练和微调优化，提高模型对恶意输入的抵抗能力。

隐私保护与数据加密：对输入数据和输出内容进行加密处理，保护用户隐私和数据安全。

工业威胁情报：工业威胁情报风险主要包括勒索软件攻击、高级持续性威胁（APT）、网络钓鱼、DDoS 攻击等。这些攻击可能导致企业数据泄露、设备运转异常，甚至引发安全事故。为了防范这些风险，首先要加强数据安全保

护，其次要注重算法的安全性和鲁棒性，通过严格的测试和验证，确保算法能够抵御各种攻击手段，从而提升 AI 算法的安全性，再次可以利用人工智能技术处理不确定信息，检测未知威胁，提升安全检测中预测、防范、检测等各个风险环节的自动化和智能化程度，最后可以构建智能化安全防护体系，通过人工智能机器学习和知识图谱等技术，对工业数据和安全数据进行汇聚、清洗、分类、抽象，借助工业互联网安全知识库和知识图谱所形成的安全知识，检测、判别安全风险与威胁，并作出响应处置决策和行动。

3.3.3 工业 AI 技术合规

AIGC 检测：工业 AIGC（人工智能生成内容）检测技术是针对人工智能生成的图像、视频和文本等内容进行真伪鉴别的技术。随着 AIGC 技术的发展，合成内容越来越难以被肉眼识别，因此需要专门的检测技术来识别。目前，检测技术主要依赖于深度学习和图像分析技术。例如，利用深度神经网络（如 CNN）来识别合成图像中的细微特征和异常模式；采用频域分析技术，如傅里叶变换，来检测图像中的周期性伪影；以及利用自然语言处理技术来分析文本内容的一致性和逻辑性。此外，还可以通过检测图像的统计特性，如局部梯度分布和纹理特征，来识别合成图像。这些技术的综合应用，提高了检测的准确性和可靠性，

使得即使在 AIGC 内容越来越逼真的情况下，也能有效地识别和区分。

深伪检测：工业领域深度伪造技术的应用带来了严重的安全风险，因此迫切需要有效的反制技术来检测和防范。这些技术包括但不限于:

深度学习检测算法：利用卷积神经网络（CNN）和其他深度学习模型来识别深度伪造内容的特征，如不自然的纹理和光照异常。

异常检测技术：通过分析图像或视频的统计特性，如像素分布和频率特征，来识别与真实内容不一致的异常。

行为分析技术：监测和分析用户行为模式，以识别可能的深度伪造攻击行为。

数字水印技术：在内容创建时嵌入不可见的水印，以便在内容分发后进行真伪验证。

区块链技术：利用区块链的不可篡改性来记录和验证内容的来源和完整性。

多模态分析：结合图像、音频和文本等多种数据模态，通过交叉验证来提高检测的准确性。

这些技术的综合应用，可以提高工业领域对深度伪造内容的检测能力，有效保护工业数据和系统的安全。

虚假数字人检测：工业数字人主要用于模拟真实员工的工作流程和行为，以提高生产效率、降低成本和风险。

它们可以执行重复性高、危险或需要精确操作的任务。虚假数字人可能用于欺诈或误导，因此需要专门的检测技术来识别。这些技术包括:

行为分析: 通过分析数字人的行为模式，识别与真实人类行为不一致的地方。

语音和面部识别技术: 利用深度学习模型分析语音和面部表情的自然度，检测合成特征。

物理模拟检测: 检查数字人在物理交互中的表现，如光线反射、阴影和物理碰撞的合理性。

深度学习图像分析: 使用卷积神经网络（CNN）检测图像中的异常纹理和像素级不一致性。

多模态一致性检查: 结合视觉、音频和文本数据，检测不同模态间的不一致性。

这些技术的综合应用有助于提高对虚假数字人的检测能力，确保工业环境的安全和可靠性。

3.3.4 “AI+工业互联网”算法安全

鲁棒性: 鲁棒性是指算法在面对错误输入或故意攻击时仍能保持性能的能力。在工业互联网中，AI 算法需要能够抵御各种攻击手段，包括对抗性攻击和数据污染。为了提升算法的鲁棒性，可以通过严格的测试和验证来确保算法的稳定性和安全性。此外，鲁棒性的提升也涉及算法的容错能力，即在部分组件失效时仍能保持系统运行的能力，

这对于工业互联网中的连续生产过程尤为重要在机器学习中，一种常用的方法是对抗性训练。这种方法通过在训练数据中引入对抗性样本来增强模型的鲁棒性。

公平性：公平性是指算法在决策过程中不因个体的某些属性（如性别、种族等）而产生歧视。在工业互联网中，AI 算法的公平性尤为重要，因为它们可能会影响生产资源的分配、员工的绩效评估等。为了实现算法的公平性，需要在数据收集、模型训练和算法部署的各个阶段预防和减少偏见。例如，确保训练数据的代表性和质量，以及在算法设计中考虑公平性指标，如机会均等和资源平等。

可解释性：可解释性是指算法的决策过程和结果能够被人类理解和解释。在工业互联网中，算法的可解释性对于建立用户信任、进行故障诊断和合规性检查至关重要。提高算法的可解释性可以通过采用透明度更高的算法模型，或者开发算法解释工具来实现。这些工具可以帮助用户理解算法的工作原理和决策依据，从而增加算法的透明度和信任度。

逆向萃取：对于投入使用的工业大模型，不法分子可以采取逆向攻击等手段，违规获取已部署的人工智能模型算法的详细信息，包括参数、结构、功能等，导致知识产权被侵犯或商业机密泄露等风险。如果被恶意篡改模型的参数、结构，或者嵌入后门，就会导致模型推理过程不可

信、决策错误、生成错误结果，甚至导致系统崩溃或无法正常运行。工业企业应建立完善的数据安全保护机制，包括数据加密、访问控制、安全审计等措施，确保数据的机密性、完整性和可用性来防止逆向萃取攻击。

3.3.5 “AI+工业互联网”数据要素安全

数据要素安全风险存在于数据收集、存储、使用、加工、传输、提供、公开、删除等数据全生命周期活动中。“AI+工业互联网”主要存在以下数据要素安全风险。

违规采集：违规采集涉及未经同意收集、不当使用数据和个人信息的安全风险。例如，未向用户充分披露收集和使用个人信息的目的，或者基于用户同意的业务目的收集的个人信息被用于模型训练，且模型训练和使用目的与原目的无关，或者超出用户的隐私期待。为了应对这一风险，工业企业应遵循数据收集使用、个人信息处理的安全规则，严格落实关于用户控制权、知情权、选择权等法律法规明确的合法权益。

数据异常：数据异常检测对于工业系统的安全和稳定生产至关重要。传统的异常检测方法需要大量标记样本，且不适应高维时间序列数据。为了解决这些问题，可以采用基于 LSTM 自动编码机的无监督异常检测模型，该模型通过学习正常样本的特征和模式来进行异常检测。这种方

法能够处理高维度时序数据，较好地适应实际工业互联网环境。

投毒污染：投毒污染是指在训练数据中植入恶意样本或修改数据以欺骗机器学习模型的方法。这种攻击可以使工业大模型 AI 算法产生错误的判断，并且由于算法黑箱和算法漏洞的存在，这些攻击往往难以检测和防范。为了应对投毒污染，需要在数据预处理阶段进行数据清理，使用自然语言处理技术来过滤掉包含不当语言或有害内容的评论，并采用人工审查高风险的数据源。

数据泄露：数据泄露风险涉及因数据处理不当、非授权访问、恶意攻击等问题，可能导致关键生产数据和用户个人信息泄露。例如，个人信息在传输过程中被黑客拦截并泄露，或者云端存储未加密导致数据被外部黑客窃取。为了防范数据泄露，应对个人信息进行加密，尤其是在传输和存储过程中，并采用强密码和多因素身份验证以确保只有授权人员可以访问数据。此外，还可以使用差分隐私技术，在模型训练时加入噪声，防止从模型结果中反推生产数据或个人数据。

3.3.6 “AI+工业互联网”平台安全

智算设施安全：智算设施作为 AI+工业互联网的基础设施，其安全性至关重要。智算设施安全涉及工业大模型算力网络的一体化协同调度，以及算力的互联效力。为了

提升智算设施的整体能效，需要推动“AI+”设施升级，筑牢数智服务基础底座，包括算网大脑强化资源的一体化协同调度，提升通算、智算、边缘算力的互联效力，加速智算成网，构建泛在融合的智能化综合性信息基础设施。此外，智算设施的安全还包括对海量生产数据的训练和推理，以及 5G+光网的运力支持，这些都是智算设施安全的重要组成部分。

AI 框架安全：AI 框架安全关注的是使用 AI 模型时平台架构、算法、系统的安全性，解决 AI 安全架构风险、算法后门嵌入、代码安全漏洞等问题。例如，用于某工业生产的机器学习开源框架平台和预训练模型库可能因开发者蓄意破坏或代码实现不完善而面临安全风险。为了提升 AI 框架的安全性，需要对预训练模型和机器学习开源框架平台进行安全检测，并及时修复发现的安全问题，以提前感知风险，降低安全事件发生的概率。

供应链安全：供应链安全是 AI+工业互联网中的另一个重要方面。供应链涉及的实体和环节多样，直接套用传统网络安全技术会导致防护效果不佳，需针对工业互联网供应链安全防护对象开展核心技术攻关，抵御日益复杂的网络攻击。提升工业互联网供应链技术安全保障能力，需要分析工业互联网供应链安全防护对象的新特征和新需求，发展可统筹兼顾工业互联网供应链全环节安全的技术

体系。此外，还需要重视工业互联网供应链的渠道安全，应对工业产品供应渠道和软件升级劫持攻击。

3.4 AI 赋能工业互联网安全

在当今数字化、智能化的时代背景下，工业互联网作为连接工业全要素、全产业链、全价值链的新型基础设施，其重要性日益凸显，为制造业、能源、煤矿、电力、医疗等支柱产业的数字化转型升级提供了有力支持。当前，我国在工业领域正在进行智改数转，工业互联网、新兴信息技术等深度融入工业数字化转型过程之中，网络与信息化环境更为复杂，工业互联网的安全显得更加复杂且重要。一方面工业控制系统的安全直接关联到生产安全和设备完整性，一旦受到攻击可能导致重大的安全事故和经济损失，其次工业互联网的数据安全同样不容忽视，数据泄露或被恶意篡改会对企业造成巨大的信誉和经济损失。如此背景下，人工智能（AI）技术的应用成为加强工业互联网安全的新趋势。AI 技术，尤其是机器学习和深度学习，通过对大量数据的分析和学习，能够有效识别和预防潜在的安全威胁，同时提高安全事件的响应速度和处理效率。AI 的这些能力使其成为提升工业互联网安全的有力工具。当然，AI 在工业互联网安全应用中也面临诸多挑战。技术的复杂性、数据隐私问题以及对抗性攻击等，都是当前需要解决的关键问题。如何在保障安全的同时，发挥 AI

技术的最大优势，是工业互联网发展过程中必须面对的重大课题。AI 技术在工业互联网安全方面的应用既是一场决胜，也充满挑战。这不仅需要技术的不断创新和优化，还需要行业、企业、政府等多方面的共同努力和协作，以确保工业互联网的健康、安全、可持续的发展。

3.4.1“AI+工业互联网”数据安全

工业元素智能识别：通过 AI 技术，可以实现对复杂制造图纸与文本信息的全面解读，精准捕获图纸中的文字描述和关键参数，提取关键视觉特征，如线条、符号、尺寸标注等，构建图纸的结构化描述。从海量复杂文件中智能识别敏感工业元素，为数据安全防护提供依据。

数据智能分类分级：数据智能分类分级技术可以运用 AI 先进的自然语言处理能力、上下文理解能力、跨领域知识学习能力，对生产文件和数据库中的敏感生产数据进行精准定位，提高数据安全分类分级的实施效率、降低实施成本，为数据安全精准防护提供支撑，为数据共享、数据流通消除潜在的安全隐患，促进数据安全有序流动，助力工业企业数字化转型。

AI 数据脱敏：AI 数据脱敏技术是指在保护用户隐私和敏感技术信息的前提下，对数据进行处理，使其在分析和共享时不暴露原始数据的具体内容。通过这种方式，可

以在不泄露敏感信息的情况下，对数据进行分析和挖掘，从而保护工业企业和个人的数据安全。

AI 数字水印：AI 数字水印技术可以在 AI 生成的内容中嵌入数字标记以识别其来源。这种技术对于版权保护、预防信息泄露具有重要意义。例如，DeepMind 推出的 SynthID 工具，能够在 AI 生成的内容中嵌入数字水印，帮助识别内容。数字水印可以是可见的或不可见的，用于确认各种数字对象的真实性，包括制造图纸、音频文件和演示短片。这种技术的应用，增强了信息的可信度，对抗错误信息和不当内容归属。

3.4.2“AI+工业互联网”应用安全

AI 质检：AI 质检通过深度学习技术，能够自动从样本图片中抽取和对比复杂特征，实现从人工设计特征规则到 AI 自动学习的突破。这使得 AI 质检在识别随机缺陷、复杂场景的缺陷检测方面具有明显优势，提升了工业质检的自动化和智能化水平。工业 AI 质检已经在汽车制造领域大规模应用，通过自动扫描和数据分析，将检测时间缩短，精度提高，显著提升了效率与质量。

AI 安监：AI 技术在安全监测方面可以处理不确定信息，对生产制造中的未知威胁具有较强检测能力。AI 具备自学习能力，能够不断提升知识水平，提高安全检测中预测、防范、检测等各个风险环节的自动化和智能化程度。

此外，AI 技术具备快速反应及精准识别能力，可以在第一时间发现和识别预防威胁，并立即启动应急响应。

智能问答：智能问答系统能够提供即时的生产制造信息查询和问题解答服务，提高工作效率和准确性。这些系统通常基于自然语言处理技术，能够理解和回应各种查询，减少人工干预，特别是在需要快速响应的工业环境中。

AI 设备巡检：AI 智能监控巡检系统通过视频智能分析进行自动巡检，实现对海量巡检点位的全天时、无人化、智能化巡检。这种系统有效弥补了人工巡检在视觉感知和实时响应上的局限性，为企业降本增效。例如，在电力和水利领域，AI 巡检系统能够实时监测和识别重要设备状态，对检测到的异常情况实时告警。

智能网关：智能网关在工业互联网中扮演着数据收集和初步处理的角色。它们能够支持视频监控、7*24 小时录像，并提供录像、检索、回放、云存储、云报警关联等功能。这些网关通过结构化摘录视频内容信息，实现数据的分布式存储和备份，保障了生产设备资源的可靠性、安全性及事故可追溯性。

产业分析：AI 技术在产业分析方面的应用，通过大数据分析和机器学习算法，能够对生产流程进行智能优化，分析生产过程中的瓶颈和问题，提出针对性的改进方

案。这有助于实现生产资源的合理配置，提高生产效率，降低生产成本。

3.4.3“AI+工业互联网”网络安全

基于 AI 的工业网络攻防：AI 技术通过自动化渗透测试，提升了工业网络攻防系统的推理能力和任务调度效率，构建了一个自适应、智能化的多层次安全攻防体系。例如，AI 可以用于自动化漏洞挖掘、恶意代码检测、威胁流量分析等，提高安全技术的及时性与准确性。

基于 AI 的网络安全管理：AI 技术能够自动分析威胁，接受多来源警报，实现合规性自动化迅速检测、响应，帮助内部网络安全团队管理和排除潜在风险。此外，AI 支持的安全事件管理自动化改进了网络事件响应流程，使用人工智能算法来分析和关联实时数据，从而能够及早发现威胁并更快、更有效地响应安全事件。

边界隔离：通过在工业制造企业部署安全设备，建立基础边界防护，精细划分安全域，并灵活配置安全策略，阻断机台之间的非法通信，大幅缩小威胁扩散范围，有效守护工业互联网重要设备或资产安全。

3.4.4“AI+工业互联网”控制安全

工控协议安全机制：AI 技术可以帮助增强工控协议的安全机制。例如，通过基于某种算法与数字证书的技术对工控协议进行安全改造，使其具备双向身份认证和报文

加密的能力，从而弥补了工控协议的安全缺陷，并满足实际工程应用需求。这种安全机制的增强有助于防止协议报文被窃取或篡改，确保工控系统的安全。

控制软件安全加固：AI 技术可以用于检测和防御对抗性攻击，提升控制软件的安全性和鲁棒性。通过精准检测和拦截对抗攻击、科学评估工业大模型鲁棒性、实时监控新型对抗攻击等措施，可以提升系统抵御对抗攻击的能力，帮助开发人员构建更安全的 AI 系统。这包括对控制软件进行安全加固，以防止恶意软件和攻击者对工业控制系统的破坏。

指令安全审计：通过自学习业务行为基线、异常行为检测、深度流量检测和自学习白名单策略来提高审计的准确性和效率。AI 能够解析工业协议，生成行为基线，识别潜在风险，提供业务安全告警，并透明化通讯数据。它利用工业漏洞库进行精细分类和审计，以资产为核心展示安全状态，评估资产安全风险。此外，AI 还能通过分析网络数据包，生成网络交互信息列表，形成行为基线，帮助识别潜在安全风险。通过深度解析工业协议，AI 能够针对特定行业提供业务安全告警，发现异常行为。AI 的深度检测技术基于应用层流量检测，透明化通讯数据，并在人机界面展示通讯信息，实现全方位展示和事后审计。通过这些

技术，AI 提高了工业控制协议中指令安全审计的准确性和效率，增强了工业控制系统的安全性。

3.4.5“AI+工业互联网”设备安全

工业一体化全程可信：根据中国移动发布的《5G+工业互联网一体化全程可信“元信任”安全解决方案白皮书》，中国移动提出了从身份可信、网络可信、终端可信、数据可信、应用可信、AI 可信、软件供应链可信、运营可信、“元信任”网络安全保险 9 个方面构建安全防护机制，推动 5G+工业互联网安全由“单点可控”迈向“全程可信”。这种全程可信的安全解决方案，利用 AI 技术在预测维护、过程优化、质量控制等领域提高生产效率和安全性，同时通过增强模型算法稳定性、防范模型算法窃取攻击和防范模型算法篡改攻击三种技术手段保障工业互联网中人工智能技术的安全性。

固件安全增强：AI 技术的应用，如大语言模型（LLMs），可以自动修复制造软件中的安全漏洞，包括固件中的漏洞。通过建立一个自动化流程，从发现漏洞到生成修复代码，再到测试和人工审查，这一流程能够有效加速并提高固件修复的质量和速度。这种自动化的漏洞修复流程不仅提高了修复效率，还增强了固件的安全性。

设备漏洞修复：AI 技术在自动化发现和修复软件漏洞方面展现出巨大的潜力。例如，某著名互联网公司的安

全工程团队利用大语言模型建立了一个自动化的漏洞修复流程，这一流程不仅能自动发现和隔离漏洞，还能生成修复代码供人工审查，极大提高了修复效率和速度。这种自动化的漏洞修复能力对于工业互联网中的设备安全至关重要，因为它可以快速响应和修复新出现的安全威胁。

3.4.6“AI+工业互联网”平台安全

流量监测：AI 技术可以通过大数据分析和机器学习算法，对工业互联网中的网络流量进行实时监控和分析。这种智能监控能够快速识别威胁并找到潜在安全风险之间的联系，消除人为错误。通过智能检测，AI 可以帮助快速识别威胁并进行模式识别。

风险识别：AI 技术的应用使得安全风险辨识评估更加全面和精准。AI 可以处理不确定信息，对未知威胁具有较强的检测能力，并且具备自学习能力，能够不断提升知识水平，提高安全检测中预测、防范、检测等各个风险环节的自动化和智能化程度。此外，AI 技术可以在第一时间发现和识别预防威胁，并立即启动应急响应，提升风险防范的预见性和准确性。

态势分析：利用数据融合、数据挖掘、智能分析和可视化等方式，AI 技术可以对工业互联网安全数据进行归并、关联分析、融合处理。通过大量安全风险数据进行关联性安全态势分析，综合分析网络安全要素，评估网络安全状

况，借助可视化呈现、预测网络安全态势，构建智能化工业互联网安全威胁态势感知体系。

安全预警与处置：AI 技术的应用提升了安全防护的主动性和智能化。工业大模型技术提供商正在利用机器学习、人工智能等技术提升威胁检测效率和安全处置自动化水平，尤其在审计和管理平台类产品中的体现尤为显著。AI 可以帮助工业企业从海量日志中迅速、精准地识别安全事件，并进行预警和处置。

故障恢复：AI 技术可以通过智能算法对原始生产数据进行预处理，降低安全分析人员数据处理压力，辅助安全分析人员做出决策判断，包括故障恢复决策。此外，AI 技术还可以预测设备可能出现的故障，并提前进行维护，减少设备意外停机带来的损失，提高设备的使用寿命和整体效率。

四、“AI+工业互联网”应用安全案例

4.1 工业大模型安全风险治理实践

中国移动为落实人工智能安全治理相关要求，保障人工智能全流程安全可控，中国移动参考国内外的安全治理经验，结合“AI+”战略及业务实际情况，构建了“1264”人工智能安全体系架构。

图 2：中国移动“1264”人工智能安全体系架构

“1”是指规划一个工作体系架构。

“2”是指着力两个工作发力方向，即“AI 安全风险防控”和“AI 赋能网信安全”。

“6”是指落实基础平台、数据要素、模型算法、业务服务、防范滥用、人员组织六大安全防护措施。

“4”是指赋能基础网络安全、数据安全治理、内容安全治理、业务应用安全四大安全领域。

这套方案不仅适用于传统的大模型也适用于面向工业互联网领域的行业大模型。通过上述措施，实现工业互联网“业务合法合规、算法公平公正、数据安全可信、系统可管可控、赋能高质高效”的总体安全目标。

4.1.1 工业互联网大模型安全防护实践

（1）背景与需求

某石化企业已建成DCS控制系统（distributed controlsystem）和SCADA系统（Supervisory Control And Data Acquisition），辅助整体石化产品制造全流程。通过这种集成的自动化控制解决方案，企业能够实现对生产过程的高效监控和管理，从而确保产品质量、提高生产效率并保障生产安全。

目前该企业已基本具备L1多模态大模型能力，并建设了统一的算力资源池，形成了一定的垂直大模型基础；但欠缺工艺技术问答、装置操作问答等大模型场景化应用能力以及对大模型全生命周期安全防护能力。随着业务不断地发展，面临的大模型能力转化为生产价值的挑战以及大模型自身所面临的挑战日益严峻，其安全风险也日益凸显。例如：数据异常、数据泄露、供应链安全风险等。

（2）建设方案

基于以上客户需求，中国移动基于工业大模型和L1多模态大模型，通过知识蒸馏、常减压装置有监督数据微调等技术，结合中国移动（上海）产业研究院沉淀的工业大模型安全治理经验，构建“算数模台用”一体化操作智能辅助大模型，有效提升客户工业生产效率；并在整体L3-操作智能辅助大模型中嵌入监测预警、安全检测、安全防护能力，确保大模型整体的安全运行，推动垂直行业大模型

安全发展。

从模型安全、模态安全、研运安全三个方面，对该企业的 L3-操作智能辅助大模型的全生命周期进行管理，能够覆盖基础设施、模型算法、数据要素、业务内容四个层次，文本、图片、音频、视频四大领域，需求、研发、应用、运营四个阶段，形成“4+4+4”大模型安全防护体系。

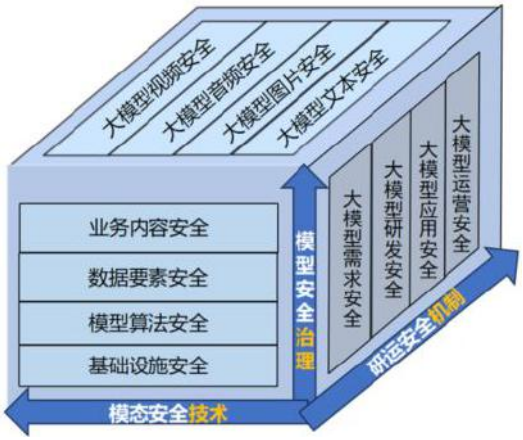

图 3：“4+4+4”大模型安全防护体系

4 层模型安全治理

业务内容安全：在石化制造、压力蒸馏等多模态场景，实时监测过滤智能问答安全，确保合规；

数据要素安全：工业元素智能识别以及数据智能分类分级，确保 AI 工业制造数据在训练阶段、开发测试阶段和部署运营阶段的全流程数据安全；

模型算法安全：避免大模型的算法出现鲁棒性和脆弱性等安全风险；

基础设施安全：保障大模型硬件、算力、网络等基础设施安全。

4 种模态安全技术

大模型文本安全：提供工业生产技艺多模态场景安全防护与合规能力;

大模型图片安全：提供工业事故预案等多模态场景安全防护与合规能力;

大模型音频安全：提供工业生产场景安全防护与合规能力;

大模型视频安全：提供工业质检多模态场景安全防护与合规能力。

4 步研运安全机制

大模型需求安全：根据监管要求及常见安全风险，规划大模型研发的安全需求;

大模型研发安全：通过评测数据集自动化生成、自动化评估等检测手段，保障大模型研发安全;

大模型应用安全：通过安全增强、模型安全对齐等安全保护技术，保障大模型应用安全;

大模型运营安全：通过实时监测预警过滤、恶意行为监测预警等监测技术，保障大模型运营安全。

（3）建设成果

构建 10 项业务应用，业务风险识别准备率不低于 60%，业务回答准确率不低于 80%，并结合“1264”的生成式人工智能服务安全建设思路，形成一套操作智能辅助大模型，

用于辅助企业的业务操作、应急处置、运行和培训，有效提升企业的运行与执行效率的同时，确保了企业大模型的全生命周期安全。数据安全方面，打造工业元素智能识别、数据智能分类分级能力，有力防范了数据异常风险；应用安全方面，打造 AI 质检、智能问答、智能网关等能力，赋能整体业务发展。网络安全方面，通过边界隔离等技术，实现工业网络攻击防护，确保整体网络安全。该项目的大模型安全应用，形成了垂直行业大模型应用标杆案例，有力推进了该垂直行业大模型安全应用的高质量发展。

4.1.2 工业互联网大模型安全风险评估

（1）背景与需求

某工业互联网企业是基于工业互联网公共服务平台，以“互联网+智能制造”为发展方向，提供覆盖产业链全过程和全要素的生产性服务。提供的服务包括企业智能化诊断服务、三维设计 CAD 软件云化、仿真分析 CAE 软件云化、PCP 协同平台云化、SAP BI 业务管理解决方案、远程运维系统搭建、氢能源监控、智能视频分析应用、燃气钢瓶信息记录平台等。

该企业工业互联网在大模型方面有多年的积累，并已经建立自己的大模型系统，通过集成大数据、云计算、人工智能等技术，为工业企业提供智能化的解决方案，且应用到工作中，效果显著。然而随着应用的深入和规模的扩

大，以及其他大模型暴露的一些安全问题，用户担心自己的大模型存在一系列风险，迫切地希望能够基于某种标准对大模型系统进行安全评估，以达到安全、合规的目的，以加速大模型的推广。

（2）建设方案

本次评估主要依据由公安部网络安全等级保护评估中心牵头的《大模型系统安全保护要求》（T/ISEAA 005-2024）和《大模型系统安全测评要求》（T/ISEAA 006-2024）这两项团体标准。该标准于2024年4月30日正式发布，旨在强化大模型自身及其整体系统安全防护体系，应对技术合规、网络安全、隐私保护及社会伦理法律等方面的新风险和挑战。

大模型安全评估从通用安全要求和全生命周期安全要求实施现场测评工作，并采取例如样本集攻击测试等技术手段进行验证，以发现大模型系统存在的例如注入攻击、投毒攻击和模型窃取攻击等相关安全风险，测评结果分别对通用安全、设计开发安全、测试安全、部署运行安全和退役安全五个方面进行安全评价。

测评框架如下图:

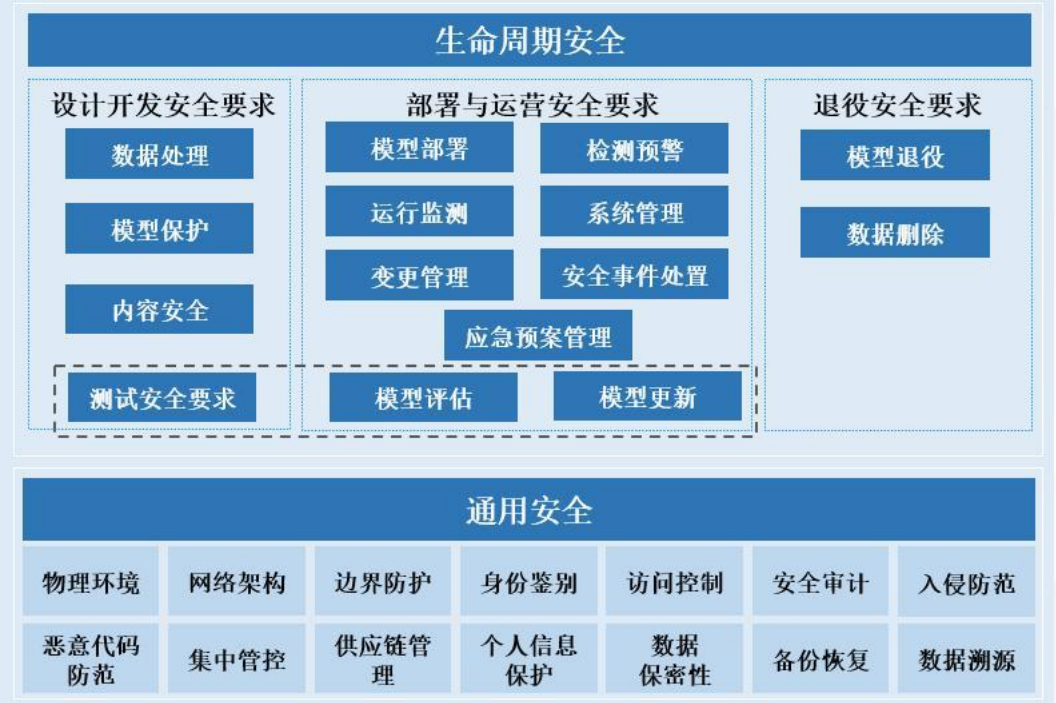

图 4：大模型系统安全测评框架

测评指标包括：

场景	安全类	指标总数	适用性指标数量
通用场景（含ToC和ToB场景）	通用安全	73 项	73 项
	设计开发安全	40 项	40 项
	测试安全	14 项	14 项
	部署与运行安全	21 项	21 项
	退役安全	7 项	7 项
模型本身安全	通用安全	73 项	73 项
	设计开发安全	40 项	40 项
	测试安全	14 项	14 项
	部署与运行安全	21 项	9 项
	退役安全	7 项	0 项

通用安全是大模型运行的软硬件基础设施平台，从物理环境、网络架构、边界防护等 14 个方面进行测评，测评按照等保三级的要求方法进行测评即可。

生命周期安全是大模型自身的安全要求，覆盖设计开发安全要求、测试安全要求、部署与运行安全要求、退役安全要求。所采用的测评方法有:

1、访谈查验：访谈、文档审查、模型流程管理、实地查看、配置检查;

2、漏洞扫描：全方位检测大模型系统存在的脆弱性，检查系统存在的弱口令，收集系统不必要开放的账号、服务、端口;

3、样本集攻击测试：利用大模型测试工具，特意设计一系列模拟恶意攻击行为，验证大模型系统的安全性;

4、渗透测试：尝试发现目标系统中存在的安全漏洞，并验证和展示相关漏洞被利用后可能对目标系统造成的危害。

5、生成报告：汇总各测评单元结果，经过算术加权计算，可以给出各个层面的分类测评结果。

对于访谈查验、漏洞扫描、渗透测试，采用传统的方法评估即可。

样本攻击测试是大模型安全评估的重要部分。目前测试单位的大模型样本攻击测试，已经积累了超过 35w+的各类样本，我们采用这些样本对大模型进行样本攻击测试，测评的内容如下表所示:

序号	安全类	测试类型

1	数据清洗	不良信息清洗过滤
2		保护敏感个人信息 或敏感业务数据
3	模型保护	越狱攻击
4		目标劫持攻击
5		用户提示词泄露攻击
6		系统提示词泄露攻击
7		对抗攻击
8		后门攻击
9		鲁棒性 （随机添加字符、随机删除字符、随机替换字符） 高强度、中强度、低强度
10	内容安全	推理攻击
11		输入/输出内容不良信息检测过滤 （多维度）

（3）建设成果

依据测评结果，对大模型安全做差距分析，并形成《大模型安全差距分析报告》，对评估发现的问题汇总结果如下:

评估类别	评估指标	问题数量
模型本身安全通用安全	73 项	5 项
设计开发安全	40 项	12 项
测试安全	14 项	4 项
部署与运行安全	21 项	7 项
退役安全	7 项	3 项

大模型系统安全测评报告可以给出大模型系统在物理环境、网络架构等各控制点的单元测评结果，并基于单元测评结果给出在通用安全和各生命周期环节的各个分类测评结果。结合测评单元中各测评指标的重要程度和测评实施符合情况，经过算术加权计算，可以给出各个测评单元的单元测评结果。汇总各测评单元结果，经过算术加权计算，可以给出各个层面的分类测评结果。

因为该用户在等保方面拥有丰富的经验，因此在通用安全方面满足要求，但是在大模型系统本身的安全存在很多问题，依据测评结果，用户对系统进行整改后进行复测，最终各层面的测评结果为良好。

从该用户的大模型系统安全评估案例可以看出，当前的各个大模型厂商主要关注大模型本身的训练数据、生成内容质量、简单过滤、性能等方面，普遍缺乏对大模型相关安全标准的了解，在数据清洗、模型保护、内容安全等工作做得很少，导致后期整改的内容比较多。在此我们呼吁各个大模型厂商应熟悉相关标准与要求，并尽快展开大模型系统的安全评估工作，确保在合规的前提下，加速推动大模型技术健康、有序地发展与应用。

4.2 AI 赋能工业互联网案例

4.2.1 AI+工业制造网络安全实践

背景与需求

广东省纺织服装规上工业企业大多处于“手工+设备半自动化”或“设备自动化+人管设备”制造初级阶段。其数字化、网络化、智能化基础薄弱且处于产业集群底层，企业上下游、产业链间协同不足，无法形成产业集群效应。

为此，市政府牵头规划、设计“纺织服装产业数字赋能公共服务平台”。其中，网络安全、纺织服装产业企业信息化安全与智能化管理备受关注，旨在打造“产业主平台+企业次平台+生产/办公安全应用”产业级一体化全程可信安全解决方案。此平台主要需求如下:

随着新一代信息技术的迅猛发展和工业互联网普及程度的不断提升，工业控制系统已经从相对封闭的状态向跨网络互联互通转变。来自公共网络的安全威胁将会蔓延至产业公共服务平台。本项目从高角度对平台+企业的工业数字化、智能化进行探索，并开展了覆盖平台+企业的网络安全一体化全程可信建设工作。

随着工业企业上云、工业 APP 培育等工作持续推进，部分工况状态、产能信息等海量工业数据向产业公共服务平台汇聚，存储状态由离散变为集中，逐渐形成高价值的数据资源池，这些工业数据将日益成为不法分子牟取利益的攻击目标。本项目通过从平台级自上而下的数据智能分析、网络安全智能治理和网络安全组织管理体系建立，明

确网络安全的决策、管理和运行职责，以满足产业公共服务平台长期高质量服务需求。

企业侧主要需求如下:

通过在纺织企业部署安全设备，建立基础边界防护;

通过试点企业收集网络安全事件及信息，建立网络安全的风险监控机制以及监督检查工作机制，整体把握网络安全风险态势，指导网络安全运营工作;

建立和健全网络安全制度与体系，打造网络安全管理长效机制;

建立依托于工业互联网安全运营平台的网络安全持续监控和响应机制，实时采集和智能分析各类安全日志和网络通信数据信息，展现网络安全整体态势。

建设方案

中国移动（上海）产业研究院依据中国移动信安中心一体化全程可信体系架构，提出“产业主平台+企业次平台+生产/办公安全应用”产业级一体化全程可信安全解决方案，在端、边、云三侧进行安全建设。综合考虑等保合规需求，安全建设涉及边界安全、数据安全、终端安全、安全审计、智能安全分析等多个方面。同时，需要同步建立整体网络安全管理制度，明确人员、资产、事件等安全要素的管理流程，提升安全管理水平。

（3）实施方案

建立以治理、风控、合规为驱动的企业级工业互联网安全服务平台分析系统。在其中，确定组织职责，明确决策、管理、运营、监控之间的职责和场景。结合企业级工业互联网安全服务平台分析系统，展示各单位的生产调度、移动业务等业务场景的网络安全态势和风险现状，智能分析、处置和落实各项风险整改措施，以满足法律法规监管要求以及企业内部管控要求，做好安全合规工作。

企业级工业互联网安全服务平台分析系统充分整合来自各个层面、维度的数据、日志等信息。其中包括安全设备、计算终端、网络设备、通信设施、应用系统、数据库、网络流量、资产信息、应用操作行为等日志信息。平台采集 IT 和 OT 的各个层面的设备设施数据，结合 AI 安全资产管理、AI 安全风险分析、AI 全面监控、AI 态势感知展示等各类业务操作数据，同时结合业务场景，通过 AI 能力，建立数据分析模型，最终实时展现安全态势，发现威胁并及时响应。

“轻盾-边界盾安全网关”通过串联/旁挂的部署方式，可实现基于已知与未知病毒、木马与恶意程序及各种加壳的病毒文件等对工业网络边界流量实时监测、检测、阻断面向工业系统的恶意病毒入侵，以避免网络、主机、服务器等感染病毒或遭受攻击。同时，能够识别针对工业网络与工业系统的 SQL 注入、WEB 攻击、恶意扫描、拒绝服

务、跨站脚本、木马后门等攻击行为。这对于网络流量的异常情况具有非常准确、有效地发现预警能力，并且能够对于高危行为进行阻断。

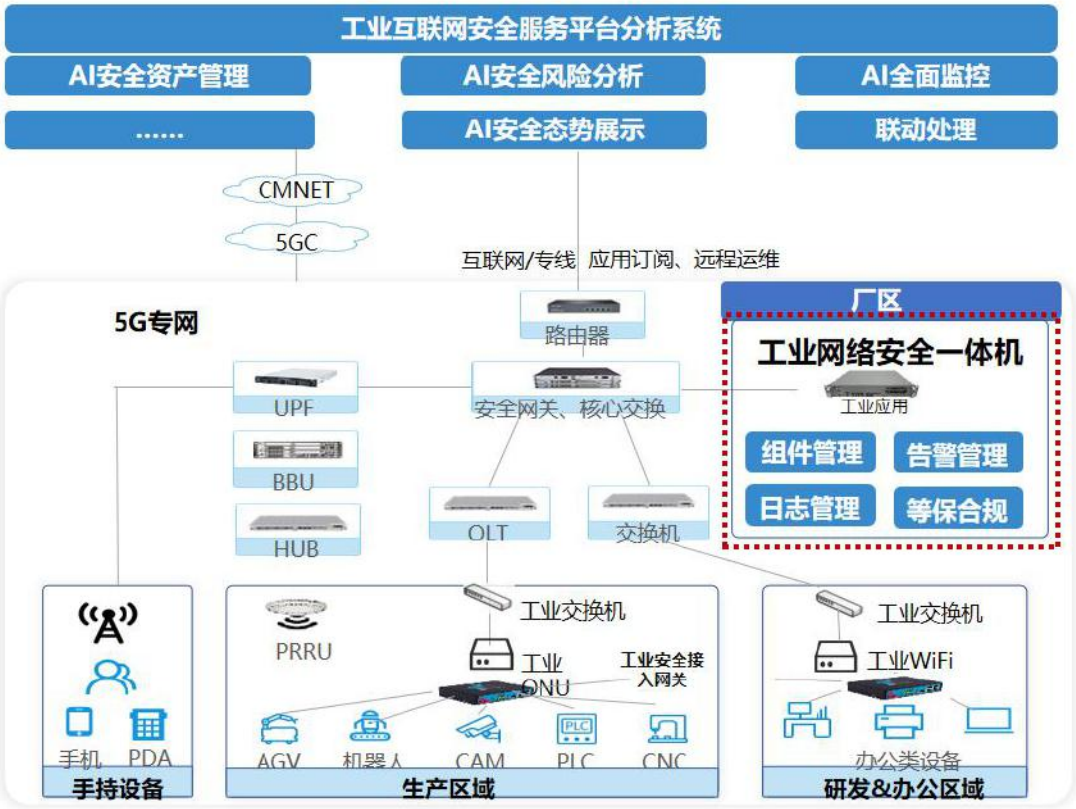

图 5：网络安全设计拓扑图

建设成果

本工业互联网安全解决方案可适用于多种推广场景，撬动安全领域新蓝海业务市场，仅纺织服装产业，其市场规模将达千万级别。同时，该产业集群的安全应用，可复制推广至其他产业领域，牵引撬动更巨大的市场规模。

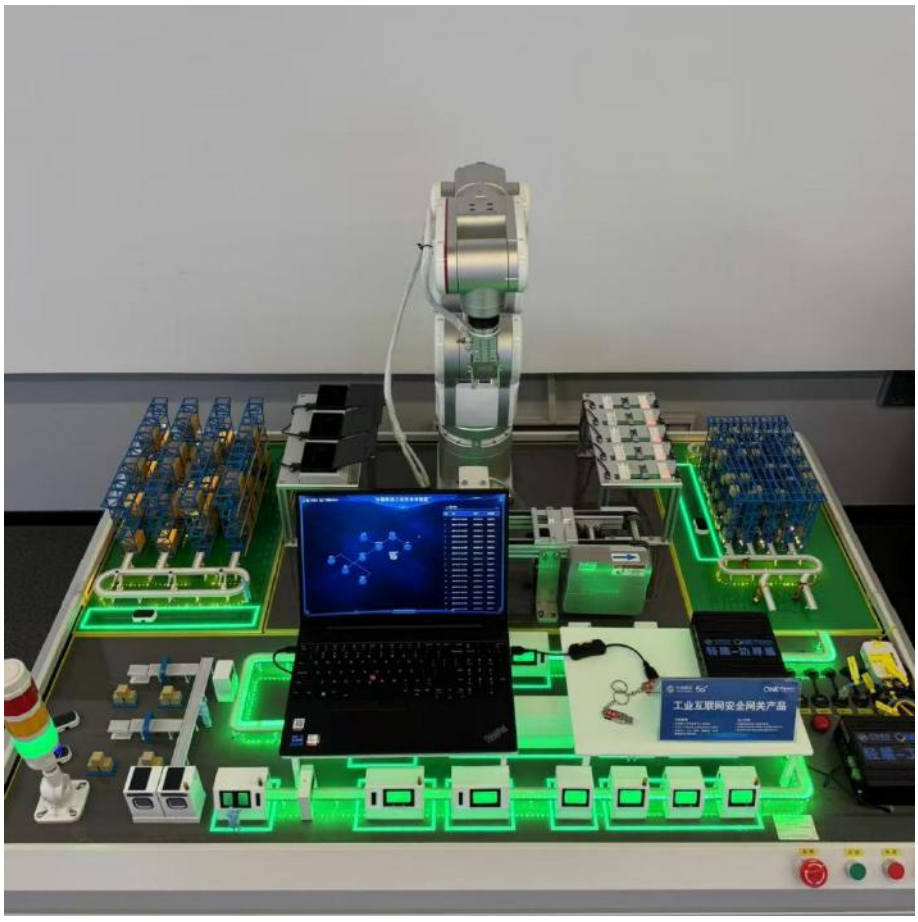

图 6: “轻盾-边界盾安全网关”演示沙盘

经济效益

推广期，面向 300 家潜在的纺织服装工业规上企业；规模化应用期，市场规模将达千万级别。并通过后续政府补贴政策将带来较高的客户使用积极性，可进一步推动产品规模上量、打造平台生态以及实现价值增长。

社会效益

本项目以安全能力赋能公共服务平台，补上公共服务平台的安全短板。通过为产业公共服务平台安全建设的规划设计，打造了“产业主平台+企业次平台+生产/办公安全应用”产业级一体化全程可信安全解决方案，保障平台及

企业侧服务安全平稳运行。本方案还将赋能政府侧“中小企业数字化转型”政策，作为有效抓手推动政策顺利实施落地。进而有效带动纺织服装产业等垂直行业企业的数字化转型，降低生产成本、提高生产连续性、提升企业效率。

4.2.2 AI+石化安全风险治理实践

背景与需求

工业互联网作为新一代信息技术与制造业深度融合的产物，其核心在于通过人、机、物的全面互联，构建起全要素、全产业链、全价值链全面连接的新型生产制造和服务体系。这一体系不仅涵盖了传统的工业控制系统（ICS），还包括了物联网（IoT）、云计算、大数据、人工智能等新兴技术，旨在实现工业生产的数字化、网络化、智能化。

工业互联网运用到油气储运行业时，在架构上，油气行业工业互联网通常分为三个层次：边缘层、平台层和应用层。边缘层负责数据的收集和初步处理，平台层提供数据存储、分析和建模服务，而应用层则基于平台层的数据和服务开发具体的工业应用。安全架构则贯穿这三个层次，确保数据在收集、传输、存储和处理过程中的安全性和可靠性。如中石油的梦想云等已经完成了三层架构的实施与运行。

但油气行业在数字化转型过程中面临的挑战是多方面的，且安全威胁和挑战也在不断增加，在油气行业智改数转过程中，安全威胁具体主要表现为:

1）网络攻击：油气工业互联网的互联互通特性使得网络攻击的风险增加，尤其是针对工业控制系统的攻击，如Stuxnet 和 Havex 等。

2）数据泄露：油气工业互联网涉及大量敏感数据，包括企业运营数据、用户个人信息等，数据泄露可能导致重大经济损失和法律风险。

3）设备脆弱性：油气工业互联网中的设备，尤其是老旧设备，可能缺乏足够的安全防护措施，容易成为攻击的突破口。

4）内部威胁：内部人员可能由于疏忽或恶意行为导致安全事件，如未授权访问、数据篡改等。

另外，油气工业互联网安全是一个复杂的系统工程，安全威胁防御与治理需要从技术、管理、法律等多个角度进行综合考虑和应对。

建设方案

面对油气储运工业互联网安全的挑战，数网安全建设需要采取多层次、全方位的防护措施，参考《工业控制系统网络安全防护指南》《工业控制系统信息安全防护能力

成熟度模型》4 级、《工业互联网企业网络安全分类分级防护系列规范》等，主动防御技术设计包括但不限于:

1）技术防护：采用防火墙、入侵检测系统（IDS）、诱捕技术、基于 AI 的网络异常行为分析、安全信息和事件管理（SIEM）等技术手段，提高系统的防御能力。

2）安全管理：制定严格的安全政策和操作规程，加强人员的安全意识培训，减少人为错误和内部威胁。

3）数网一体化防护：采用 AI 技术实施网络安全主动防御，数据加密、访问控制等措施，保护数据的机密性、完整性和可用性，实现数网一体化防御。

4）AI 安全治理：AI 技术在工业互联网安全中展现出巨大潜力，但也面临着数据泄露与隐私侵犯风险、AI 算法的安全隐患、智能设备的安全风险等挑战。为了应对这些挑战，需要加强数据保护技术，完善隐私保护政策，进行算法审计与测试，提高算法透明度，并对智能设备进行安全认证和加强供应链安全管理。

5）应急响应：建立应急响应机制，一旦发生安全事件，能够迅速响应和恢复。

在油气行业的数字化转型过程中，AI 技术在提升网络安全方面发挥着至关重要的作用。AI 技术的应用不仅能够提高威胁检测的准确性，还能够预测潜在的安全风险，从而实现主动防御。AI 技术在油气行业的数网安全一体化建

设中发挥着关键作用，不仅提升了网络安全防护能力，还优化了数据管理，为油气行业的数字化转型提供了强有力的技术支持，为“人工智能+能源”网络安全建设，推动大模型技术自主可控具有极其价值的应用。

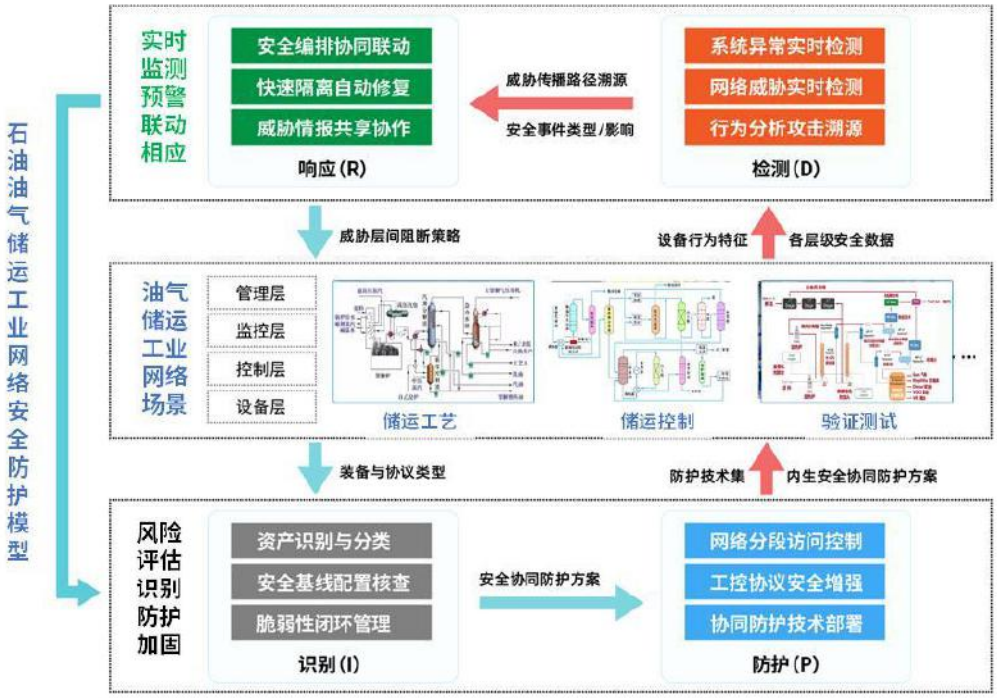

图 7：基于 IPDDR-V 油气储运工业互联网安全防护架构

（1）油气储运业务安全提升

通过三维数字孪生、物联网、大数据、人工智能、地理信息、北斗定位等信息技术，将 AI 与油气工业互联网平台融合，实现油库全面感知、实时监测、智能分析、联动预警的能力，满足油库日常管理的需要。

1）系统构建与实施：根据油气储运 AI 大模型，结合油气工业互联网平台、油田数字孪生安全生产系统能够实时反映油田生产的真实状态，并通过数据分析和模拟预测，为油气储运安全管理提供科学依据。系统通过三维建

模软件结合油田的实际 CAD 图纸、卫星图像等资料，构建出油气储运的三维几何模型，精确反映油气储运的管道与地理地貌、设备布局等物理特征。同时，物联网技术部署大量的传感器和监测设备，实现对油田生产过程的全面感知和数据采集。

2）安全储运：AI 技术通过计算机识别、数据分析和预测提升油气安全储运水平，做好用户季节性需求预测，支撑调度决策。

3）智能监控：在油气站场、管道线路的监控中，AI 技术可以提高监控效率和响应速度，提升安全保障。

4）预测分析：利用 AI 技术综合处理运输储运过程中的温度、压力、流量、设备状态等参数，发现负荷变化异常点，给出优化决策方案。

（2）油气储运工业互联网平台效能能力提升

在油气储运数字化转型的过程中，技术挑战是多方面的，涉及数据采集、处理、分析以及 AI 技术的集成应用。

1）数据采集与整合：油气储运产生的数据量巨大且复杂，包括地质数据、生产数据、市场数据等。为了有效应对这一挑战，借助先进的基于 AI 的油气储运工业互联网平台的数据采集技术和数据整合平台，以确保数据的准确性和可用性。通过部署智能传感器和物联网设备，实现对油

气输送管道的实时监控，提高了数据采集的效率和准确性。

2）数据分析与 AI 应用：数据分析是数字化转型的核心。油气储运利用机器学习和深度学习技术，从海量数据中提取有价值的信息，以支持决策制定。通过 AI 技术，自动分析油气品类、地震及采油数据，提高数据处理效率 30%以上，且与自动化设备运维系统对接，满足业务安全。

3）技术集成与平台建设：油气储运构建统一的数据平台和 AI 应用平台，以实现数据和技术的集成，可以涵盖了智能油气外源、智能工程等多个业务场景，通过构建业务场景模型，优化资源配置、提升工作效率。

（3）油气储运数网安全一体化防御能力提升

随着数字化转型的深入，油气储运面临的网络安全挑战日益严峻，包括数据泄露、服务中断和外部攻击等。

1）数据泄露防护：为了防范数据泄露，建立严格的数据访问控制和数据加密措施。通过 AI 技术，可以实时监控数据访问以及敏感数据流转行为，并进行数字资产智能测绘，及时发现并阻止未经授权的数据访问和泄露。

2）网络攻击防御：建立基于“三化六防”的主动的网络安全防御体系，满足《工业控制系统网络安全防护指南》《工业控制系统网络安全泛化能力成熟度模型》的四级要

求，以抵御外部攻击。通过部署基于 AI 的入侵检测系统和防火墙，可以有效地识别和阻止恶意攻击。

3）安全运营中心（SOC）建设：部署启明星辰与中国移动联合研制的基于 AI 的泰合的安全运营中心，实现对网络安全威胁的实时监控和响应。SOC 可以集中管理安全事件，提供安全分析和响应服务，帮助企业快速应对安全威胁。

4）员工安全意识培训与 AI 辅助教育：定期开展安全演练，提高企业对网络安全事件的应急响应能力。通过基于 AI 技术的启明星辰泰合网络安全运营平台帮助企业理解和遵守复杂的法规要求，确保数据合规。

通过上述措施，油气行业可以在数字化转型的过程中，通过整合先进的信息技术，如大数据、云计算、物联网（IoT）和人工智能（AI），油气行业能够实现全流程优化，同时确保数据的安全性和业务的连续性，有效应对技术挑战和安全挑战，确保数字化转型的顺利进行。数网安全一体化不仅提升了油气储运的生产效率和安全性，还为企业带来了显著的经济效益。

（3）应用效果分析

随着 AI 技术与油气行业工业互联网平台的不断发展和应用，油气行业的数网安全一体化建设将继续深化，工业大模型将成为油气工业智能化的重要范式，推动产业升级，

并实现可持续发展。

1）系统特点与优势：通过AI+油气工业互联网安全技术，实现全数据支撑、全流程可控、全区域可视、全领域透明、全要素预警等，解决了现场安全威胁的自动化监测、安全分析、安全报警等主动安全防御水平不足的问题。系统的应用提升了库区在施工管控、报警处置、业务协同、运行管理等方面的管理水平。

2）应用场景与成效：采用AI+工业互联网平台的智能油库管控系统适用于原油站库、成品油库、液体化工品库、码头库区等多种场景。系统通过自动化作业、智能化管理、智能化数据分析及应用，统一油库信息系统功能结构、软硬件数据接口、核心设备技术要求，做到安全适用、技术先进、经济合理。

3）经济效益分析：提高作业效率、降低安全风险等措施，为企业带来了显著的经济效益。系统的应用不仅提升了油库的自动化、智能化水平，还为企业带来了全新的发展机遇。

4.2.3 AI+矿山冶金数据安全评测案例

背景与需求

在数字化转型加速的背景下，矿山冶金行业的业务应用系统已成为关键信息基础设施。这些系统涉及的数据类型广泛，包括矿产勘探数据、生产过程数据、环境监测数

据以及企业经营数据等。这些数据不仅是矿山冶金企业日常运营的基础，更是提高生产效率、优化资源配置和保障作业安全的重要依据。而作为国家经济的重要组成部分，矿山冶金行业的数据安全防护更是承担着保障社会稳定和工业安全的重要责任。对于矿山冶金行业，数据的安全性和可靠性至关重要，因为任何数据泄露或安全事件都可能对企业运营和公共安全造成重大影响。

在此背景下，建立健全的数据安全治理体系，有效应对不断变化的安全威胁和挑战，提升数据使用效率，推动业务创新与发展，建立健全数据安全协调治理体系，落实数据保护责任，完善数据分类分级，保障数据全生命周期安全是当前数据安全治理工作的核心。

在保障业务系统的数据安全前提下，开展数据安全治理工作主要存在以下几点需求:

管理的数据资产与实际资产的匹配情况不明，敏感数据的种类和访问者的身份不清。矿山冶金企业虽会定期开展网络安全检查工作，但未对数据资产使用权限进行梳理，未落实数据分类分级管理，以及制定相应的权限管控。

数据安全管理落实相关制度流程缺失，缺乏相应的稽核机制。矿山冶金企业涉及大量敏感信息，数据场景日益

复杂化，若无切实有效的制度流程，缺少数据安全管理的抓手，难以掌握数据的流向。

对数据安全风险的具体场景、防护措施及整改方法不明。矿山冶金企业虽符合网络安全防护要求，但数据安全防护较弱，知道存在数据安全风险，但是具体风险有哪些在哪里、如何建设防护、建设路径是什么、建设周期是多久均不清楚，导致数据安全防护建设无从下手。即使发生数据安全事件，也较难及时发现。

建设方案： 矿山冶金领域的数据安全治理主要围绕韧性数据安全体系开展，突破传统数据安全理念，将各类数据安全防护、数据安全管理、数据安全运营等安全能力通过一套技术架构收敛至一个数据安全管理平台、提供数据统一管理门户。在此基础上，引入人工智能（AI）技术，针对数据资产进行快速摸排及发现，通过多种 AI 模型技术赋能数据安全治理，提升数据安全治理的智能化和自动化水平。AI 技术可以在数据安全风险监测方面发挥关键作用，通过机器学习算法实时分析数据流动，识别异常模式和潜在威胁。同时，利用自然语言处理（NLP）技术，可以自动化处理安全事件报告和合规性审计，减少人工干预，提高效率。此外，AI 技术还可以通过自动化响应机制，实现快速响应和处理安全事件，降低数据泄露和安全风险的可能性。借助 AI 技术，动态调整安全策略，以适应不断变

化的环境和需求，从而确保数据在复杂场景下的安全性和合规性。如下为项目建设架构图。

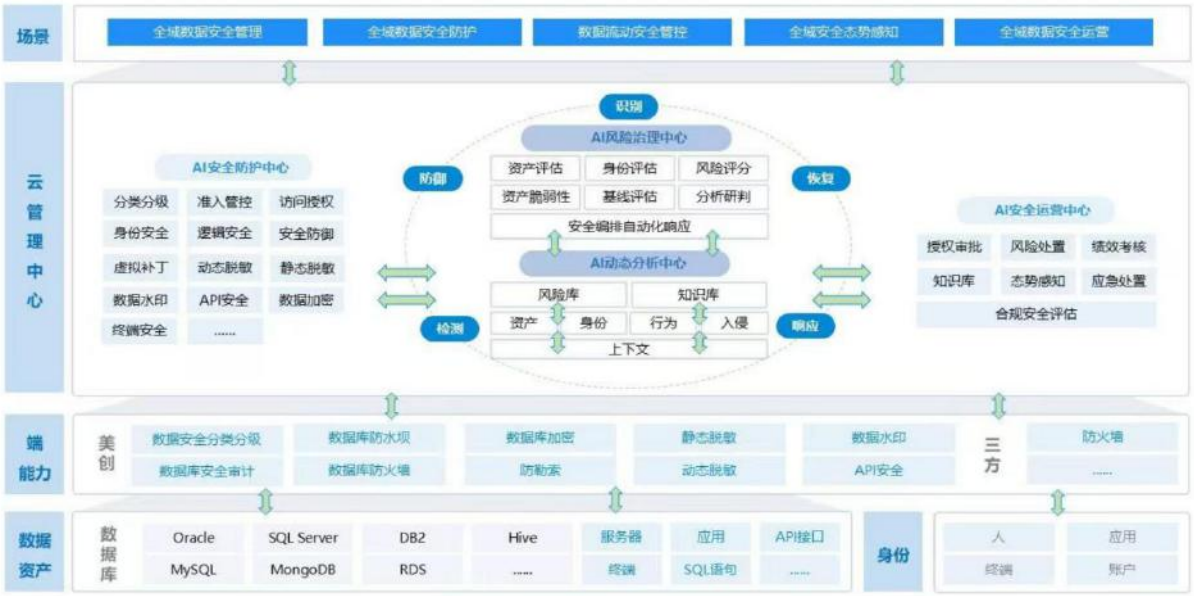

图 8：数据安全治理项目建设架构图

全域数据资产治理

以“资产”为核心，自动发现海量数据资产，结合业务模型进行数据的分类分级，是后续进行风险智控的基础。通过对海量、复杂数据资产统一管理，使用内置丰富的分类分级模型，实现资产梳理、有序化管理与防护，提供 AI 智能大模型技术如 NLP 翻译、特征工程、LLM 语义大模型、语义向量模型技术，赋能数据资产识别及管理。

风险实时智能监测

基于实时采集各网关的日志流量，结合 SIEM（安全信息和事件管理）、UEBA（用户和实体行为分析）技术，对海量日志进行智能建模分析，快速发现各类潜在威胁，并

对各类资产和身份进行安全评分，提前预知各类潜在风险。

风险告警与处置

基于风险来源、风险类型、风险内容等多个维度，通过智能分析来确定后续的应对策略。结合人工智能（AI）技术，能够自动和手动方式针对不同的安全问题进行策略下发，形成数据安全管理平台与安全产品之间的策略协同联动。当系统发现安全风险时，平台能够对风险事件生成相关风险策略，并将策略下发到其他安全设备。

实施方案： 按照政策法规要求，以 DSMM 为抓手，以数据流向和应用场景为切入点开展数据安全治理。数据安全治理内容分为以下三个阶段:

第一阶段: 数据资产梳理。通过问卷调研、工具探查等方式多维度盘点数据资产，厘清各类数据资产现状，并在此基础上进行数据分类分级，为后续数据安全精细化管控提供基础。数据资产梳理完成了如下内容: 数据资产盘点: 通过工具探查数据资产情况，为分类分级实施做好准备工作; 数据权限现状: 盘点清楚用户具备哪些权限，数据可以被哪些用户增删改查，权限过大用户有哪些等; 数据流向梳理: 盘点清楚数据从采集、传输、共享交换到销毁的流向; 数据分类分级: 完成数据分类和分级，共分四

级，其中敏感表格占比约 65%，敏感字段占比约 53%，同时明确各级数据的安全要求。

第二阶段：数据安全风险评估。对数据安全现状进行分析，识别和分析数据资产在全生命周期各阶段风险，并给予中立的加固建议。数据安全风险评估内容包括如下：基础风险评估：通过安全基线检查、漏洞扫描、渗透测试等方式，发现在数据处理环境中存在的安全漏洞，并提供加固建议，通过漏洞整改，可以降低、抵消风险等级，防范安全事件发生，避免信息系统脆弱性被非法利用，保障业务连续性；数据安全能力差距评估：深度分析和评估当前组织安全能力现状，帮助其清楚自身在生命周期各阶段的能力现状与目标的差距；数据安全合规评估：全面解读和分析《数据安全法》《个人信息保护法》以及地方办法条例内容，通过联合对标分析，全面评估组织数据安全合规情况和合规风险； 数据全生命周期风险评估：参考信息安全风险分析方法，从资产和风险两大视角出发，基于数据分级结果，建立组织风险评估模型，识别组织面临的数据安全风险，并通过定性分析和定量分析方法，分析并计算数据资产所面临风险值。

第三阶段：数据安全建设规划。基于数据安全风险评估的结果，结合矿山冶金行业的实际情况，提供针对性的数据安全建设规划方案，明确建设依据、建设规划、建设

路径、建设周期、建设优先级等内容，指引数据安全建设道路。

管理制度建设：结合实际场景，基于合规要求完成《数据安全管理办法》、《数据全生命周期管理制度》、《数据安全应急响应管理制度》等制度建立。

数据安全建设规划：从管理、技术和运营三个维度规划，坚持“长短结合、充分利用、平稳过渡”的方针，开展数据安全建设的短、中、远期规划和建设工作。管理规划包括制度建设、人员管理、用户隐私协议等内容；技术规划覆盖数据全生命周期，分步建设；运营规划基于管理和技术，同步建设，同步运营，从应急响应事件处置、场景化运营、安全意识培训、绩效考核等多个方面逐步建设。

建设成果:

经济效益

通过数据资产梳理与分类分级，矿山冶金企业能够更准确地识别和管理数据资产，减少资源浪费，提高数据管理效率，进而降低运营成本。通过基础风险评估、漏洞整改等措施，降低了潜在安全事件的发生频率，从而减少了因数据泄露或安全事件导致的经济损失和法律赔偿。通过对《数据安全法》和《个人信息保护法》的合规评估，矿山冶金企业能够及时调整管理策略，避免因不合规行为而

产生的罚款和处罚，避免产生损失。数据安全管理的提升不仅能增强用户信任，还能成为相关单位在市场竞争中的一大优势，带来可观的经济收益。

社会效益

通过透明的数据管理与保护措施，提升社会群体对矿山冶金行业的数据安全信任度，促进用户关系的稳定与发展。在数据安全方面的投入体现了对用户隐私和数据安全的重视，展现了社会责任感，增强了矿山冶金企业的社会形象。项目的实施促进了行业内对数据安全治理的重视，有助于推动整个行业的数据安全标准化与合规化发展。通过安全意识培训和绩效考核，提升员工及公众的安全意识，减少因人为因素导致的数据安全事件，促进社会整体安全水平的提升。

4.2.4 AI+电力能源数据安全防护案例

背景与需求

电力数据作为关系到国计民生的重要数据，面临着APT、软件供应链攻击、数据勒索、数据泄露、内部人员作案等诸多安全风险点。同时随着分布式电源、储能设备和电动汽车等新型业务的快速涌现，电力系统的交互主体日益多样化。这种多元化使得整体安全防御关口增多，外部主体的弱安全防护能力可能将网络安全风险传导至电力系统的核心区域，导致边界安全风险陡增。新型电力系统引

入了多元主体，数据交互方式从传统的单向采集转变为双向互动。这种变化使得数据流通共享、交叉访问的需求剧增，引出数据量大、交互频繁、类型多样的安全挑战。数据共享与隐私保护之间的矛盾日益突出，内部和外部攻击呈递增趋势。敏感数据在跨部门、跨系统间留存，任何一个环节或安全防护措施不到位，都可能导致数据泄露。

在这样的背景下，某电网开展数据安全防护体系建设就显得尤为必要。通过构建完善的数据安全防护体系，可以有效应对各种安全风险，保障电力关键基础设施的安全稳定运行，确保电力数据的安全可靠，为国家经济发展和人民生活提供坚实的电力保障。

建设方案

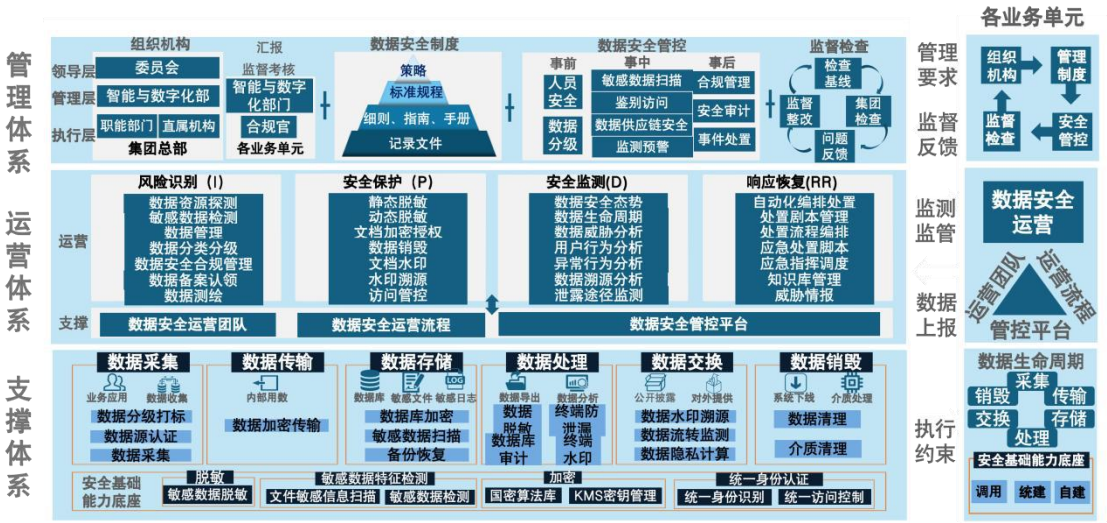

图 9：某电网数据安全防护体系

本方案通过数据安全管理支撑、数据安全运营、数据安全支撑三个要素进行数据安全防护体系建设。

1）数据安全管理

数据安全防护体系建立在对政策法律及标准规范的合规遵从的基础上，做到有法可依、有规可循，并以公司现有安全管理制度与规范为依据，进一步建立、健全和完善公司数据安全制度，为公司数字化转型各业务系统的数据安全保驾护航。

2）数据安全运营

数据安全运营以持续保障公司数据安全为目标，遵循IPDRR模型，从数据资产发现出发，绘制数据资源清单，执行数据分类分级管理，识别重要数据及敏感数据，开展数据安全风险评估，识别风险隐患，并依据评估结果采取相应的安全防护手段，对出现的风险处置。

3）数据安全支撑

数据安全支撑以公司安全服务平台提供的数据安全防护能力为基础，为各类业务应用场景提供统一、合规、一致的数据加密、数据脱敏、数据水印、泄露监测、隔离交换、数据共享、数据发布、备份容灾及数据销毁等能力。

本方案的核心功能如下:

1）数据资源智能识别

对数据库、文件源等静态资源自动扫描发现识别梳理，对流动中的重要业务系统、业务接口、业务数据库等动态资源进行自动化识别和梳理，帮助完成全量数据资产的底账盘查。

2）数据智能分类分级

要通过对敏感数据的识别和分类分级，对数据安全的合规监测，对分级管控的系统和数据进行全方位全流程全链路的安全管控防护。敏感数据识别引擎主要有基于规则的分类引擎和基于 AI 大模型的分类引擎融合完成。

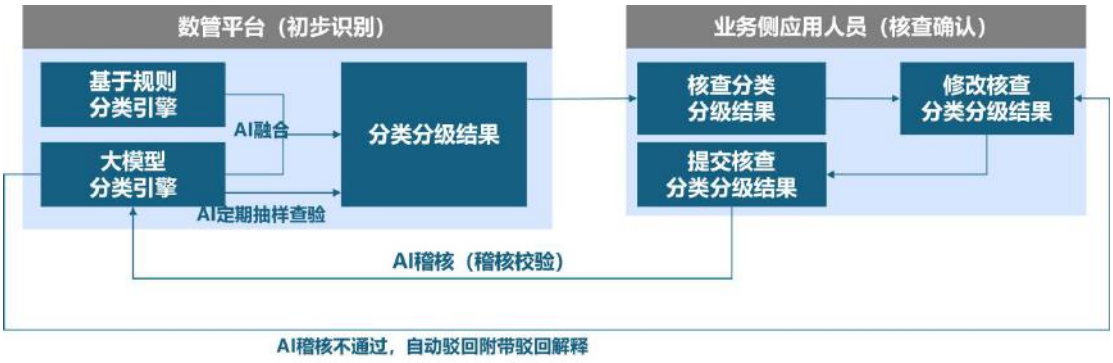

图 10：大模型 AI 分类引擎和规则引擎融合

利用 RAG 等技术为大模型引入外部专业知识代替重新训练和微调，开发设计智能体，采用长短期记忆、反思、工具等技术手段，完成业务功能的开发。数据分类判断是大模型一步一步推导过程，逐步收敛在具体的分类区间里，直到最终判定为哪个分类。同时，为了保证数据的准确性，通过对核查结果的 AI 稽核，自动校验人工核查后的异常数据，实现对提升数据分类分级准确率和加强恶意降级的风险防范。AI 定期的抽样查验的方式自动化验证和检查，生成查验报告，指出需要整改修正的数据类型和数据位置。

通过引入 AI 大模型识别引擎，实现规则引擎和大模型分类引擎的融合，增强数据分类分级的广度和深度，提升数据初始识别准确率和结果核查效率，从而提升整体分类

分级交付效率。

3）数据流转风险监测

数据流转监测采集并解析了从客户端-应用-数据库全链路协议流量，以业务资产为出发点，围绕敏感流量来进行数据的流向、流转路径和流转事件的全面监测。从风险主体（应用、接口、IP、账号）的访问行为，识别出 API 的脆弱性风险、数据流转风险、异常访问行为风险。

4）数据流转测绘

通过流转探针的数据采集，经过数据流转监测引擎分析能自动进行流转测绘。从而掌握数据流动的情况和用数人员访问情况。清晰地看到风险分布情况、风险级别并进行快速的联动处置，当前风险的收敛情况。

5）数据审计溯源

全链路反演溯源，从用户使用场景的闭环逻辑出发，构建了发起溯源任务、溯源结果输出、溯源结果分析、溯源结果报告输出的完整应用场景。

6）数据安全合规检查

支持法律法规录入集中管理后上传、下载形成知识资料库，并且根据对应摘要解读出具体的管理和技术维度的合规项。基于数据全生命周期视角，通过安全合规策略制定不同的模板场景配置任务，最终产生安全合规结果及安全合规报告。

7）数据安全联防联控

支持开放对接各种安全防护能力进行联防联控。开放对接各原子级能力的集合，其中包括敏感识别、分类分级、脱敏、水印、数审、溯源、合规检查等。通过策略中心按需组合配置和编排多种管控原则，实现可编排的安全管控策略，通过下发的管控策略进行多种防护能力的联动，可视化监测多种防护功能的实施完成情况。

（3）建设成果

在数据安全运营方面，该电网形成了以感知、管理、防护为主线的网级数据安全运营中心，构建起迭代渐进的运营闭环能力。同时，提供数据安全运营综合看板，从数据风险总体状况、资产分布情况、系统风险处置情况、全生命周期风险状况等多个维度，对各单位各部门的数据安全状况进行全方位展现。

对于企业而言，这些数据安全防护体系建设举措带来了显著的经济和社会收益。经济方面，通过有效的数据安全防护，降低了因数据泄露、被篡改等安全事件带来的经济损失风险。保障了企业核心数据资产的安全，为企业的稳定运营和持续发展提供了坚实基础，有助于提升企业的市场竞争力和盈利能力。社会方面，作为电网企业，确保数据安全有助于保障国家能源安全和社会稳定运行。提升了企业在公众心目中的形象和信誉度，为企业赢得良好的

社会声誉。同时，也为整个行业的数据安全建设提供了有益的参考和示范，推动行业的健康发展。

4.2.5 AI+工业平台威胁态势监测实践

背景与需求

工业互联网安全作为发展工业互联网的前提和保障，关系着工业安全、经济安全乃至国家总体安全。为解决当前工业互联网安全监管能力薄弱、安全技术手段缺失等问题，通过产学研一体化模式，构建工业互联网安全保障体系，建设工业互联网安全态势感知平台，实时感知市内工业互联网安全态势，实现监测预警、通报，并与国家平台实现对接，实现上下贯通、政企协同、多方联动，满足省级主管部门对工业互联网态势感知、风险预警及信息共享的需要，提升省级主管部门、基础电信企业、工业互联网平台企业、标识解析企业以及工业企业协同应对网络安全事件的能力，提高工业互联网安全保障能力。

建设方案

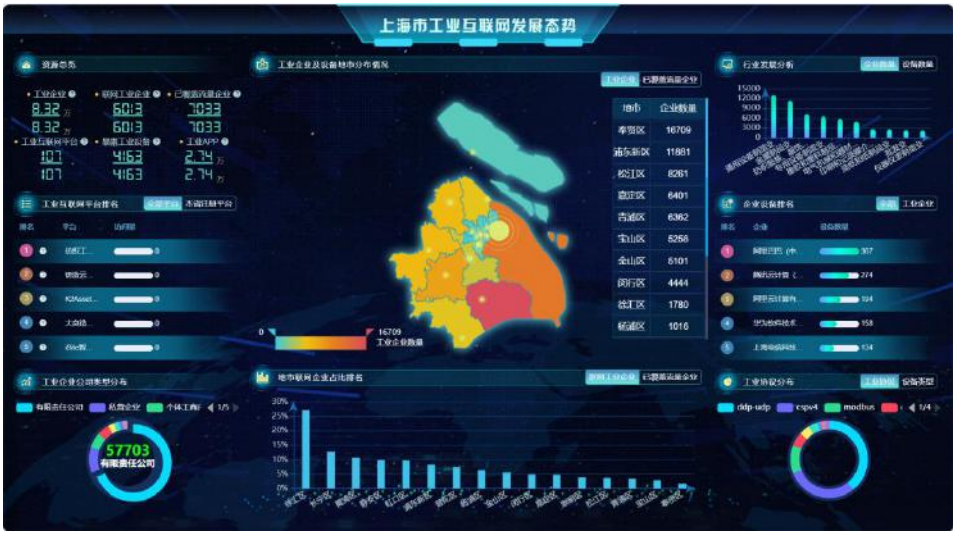

图 11：上海市安全态势感知平台

以上海市态感平台为基础，汇聚区域内关键网络节点、工业企业和工业互联网平台企业的数据，聚合专业化的监测预警、应急处置等手段，开拓安全监管新模式，打造集工业互联网资产识别、威胁监测、态势感知、预警通报和应急处置等于一体的工业互联网安全监管能力，及时洞察网络攻击和意图，准确定位安全风险及隐患，全面掌握区域内工业互联网安全态势。

通过机器学习、深度学习等先进算法，对工业网络中的海量数据进行实时分析，以识别异常行为、预测潜在威胁并采取相应的防护措施。这种技术能够自动学习并适应不断变化的威胁环境，从而提高工业系统的安全防护能力。AI+工业威胁态势监测技术可以实时监测网络流量、设备状态、安全事件等多种数据源，通过智能分析算法快速识别出异常模式和潜在威胁。同时，利用 AI 的预测分析能力，可以对历史数据进行深度挖掘，以预测未来可能出现

的威胁趋势，从而提前制定防御策略。

建设成果

平台面向上海市内重点工业互联网企业开展安全监测，资产隐患排查和风险评估，提升企业自身防护能力。目前平台建立全市的工业互联网资产库，覆盖工业互联网通信协议 120 种，流量解析支持工业协议数量 53 种，支持物联网协议数量 24 种，互联网软硬件产品类型 305 种。平台监管覆盖 8.32 万家工业企业，其中覆盖重点企业 7034，识别工业设备 3105 台，工业互联网平台 107 个，工业 APP 2.74 万个，2024 年发现高危事件 15630 次，向国家平台上报安全事件超过 760 万条，有效避免了企业因网络安全问题可能带来的不良影响和经济损失。

平台向上对接国家平台，向下连接区域内企业平台，实现了“国家—省级—企业”三级联动，为构建上下贯通、政企协同、多方联动的国家工业互联网安全态势感知体系提供重要支撑。基于平台协助主管部门实现每月按照企业名称、域名、IP、端口及安全问题形成安全威胁报告，通过书面或可验证来源的电子方式等形式通告相关单位。并监督被通报企业按期完成网络安全整改工作，形成全市监测月报机制，提升上海市工业互联网安全监管能力。

五、“AI+”在工业互联网的安全展望

5.1 AI 让工业互联网更安全

5.1.1 完善法律法规和安全标准体系

目前我国现有的法律法规正在逐步完善，《互联网信息服务算法推荐管理规定》和《生成式人工智能管理办法》的颁布和实施，确保 AI 技术的安全、透明和道德，同时也在促进技术创新和保护个人隐私。

现有的法律法规对工业互联网从业者在人工智能安全发展大方向上提供了重要指导，让工业互联网人工智能做到了有法可依，更好地组织开展相关领域的工作。工业互联网从业者应严格遵守国家及地方法律法规要求，落实工业互联网人工智能管理相关工作。“AI+工业互联网人工智能安全”的国家标准体系仍待完善，需加快推进相关标准规范的制定。国家层面应以构建人工智能全产业链安全为目标，做好个人数据保护、网络安全、AI 伦理等基础标准的制定，并持续推动工业互联网人工智能安全国家标准化工作，提升国家标准化水平。

5.1.2 推进技术发展，加强自主可控

AI+工业互联网发展的背后，有着众多人工智能技术的支撑，从数据全生命周期管理到信息化系统建设，每一个环节都引入大量新兴技术。从工业发展的角度，AI 赋能网络安全、数据安全和应用安全，保障各类工业业务系统安

全，让不同的服务可以平稳运行在工业制造中，让生产更加高效。同时随着 AI 技术的飞速发展，除传统网络安全技术之外，针对 AI 能力平台的安全技术也亟待发展，对于 AI 大模型技术，例如对抗性训练、模型审计和验证、模型加密、数据掩码等安全技术的综合运用，可以有效地对 AI 技术的安全性进行提升。对待日新月异的人工智能技术，我国应加大核心技术投入，鼓励科研机构和企业在自主可控的关键技术、产品上进行深度研究，推进国产化替代进度。

5.2 AI 让工业互联网安全更智慧

5.2.1 强化运营管理水平，培养队伍

工业互联网人工智能建设过程中，会涉及大量如算力、数据、服务器等软硬件资产，相比于单纯的数据管理内容，人工智能的管理更为复杂，要最大化发挥这些资产的价值，则重在运营。

建立全面的工业互联网管理系统，整合各个行业的数据并实施数据安全保护措施，以支持智能决策和优化生产运营。

培训工业互联网行业管理者和从业人员，提升其对人工智能技术应用的理解和运用能力，加强工业互联网治理水平。要强调的是，工业互联网人工智能运营管理需要与

社会治理结合，搭建工业互联网治理体系，促进信息共享和跨行业协同，实现 AI+工业互联网的可持续发展。

5.2.2 完善 AI 安全体系与治理

我们会继续努力完善由 AI 驱动的工业互联网安全体系，以便为 AI 基础设施和业务应用提供稳固的安全支撑。通过明确各相关方的责任和角色，我们将全面打造 AI 安全治理，致力于满足合法、公正公平、可信赖的数据安全和可控可管的系统等安全目标。