# 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