云计算标准和开源推进委员会

TRUSTED SECURITY 可信安全

安全行业大模型技术应用态势发展报告

中国通信标准化协会

TC608 云计算标准和开源推进委员会

版权声明

报告愿景及目标读者

从国家宏观层面来看，数字经济已经成为我国经济发展的重要战略支柱。随着人工智能技术的迅速发展，大模型在各行各业的应用日益广泛，安全行业也积极使用大模型技术来推动行业的数字化转型和智能化升级，提升整体安全防护能力。在此背景下，深入剖析和梳理大模型在安全领域的应用态势，对于推动安全行业创新、提升安全行业整体防护水平以及促进数字经济健康发展具有重要意义。

本报告从大模型技术与应用概述、安全行业大模型技术应用现状、安全行业大模型技术落地关键点、安全行业大模型技术应用发展趋势与展望几个方面进行洞察和分析，为安全行业从业者、技术开发者以及企业决策者提供参照，促进大模型技术在安全行业的深度应用，推动行业创新，提升整体安全防护能力，为数字经济的健康发展保驾护航。

主要撰稿人

马铭洋、卫斌、李忠权、郭雪、左鹏、高志民、黄超、朱季峰、王肖斌、刘斌、刘刚、王永霞、冯大刚、杨剑、王龑、卞超轶、李栋、马琳、刘春鸣、姚鸿富、李雨含、梁伟、唐佳伟、应缜哲、张运鹏、包沉浮、高磊、李智杰、章玉龙、陈正刚。

（排名不分先后）

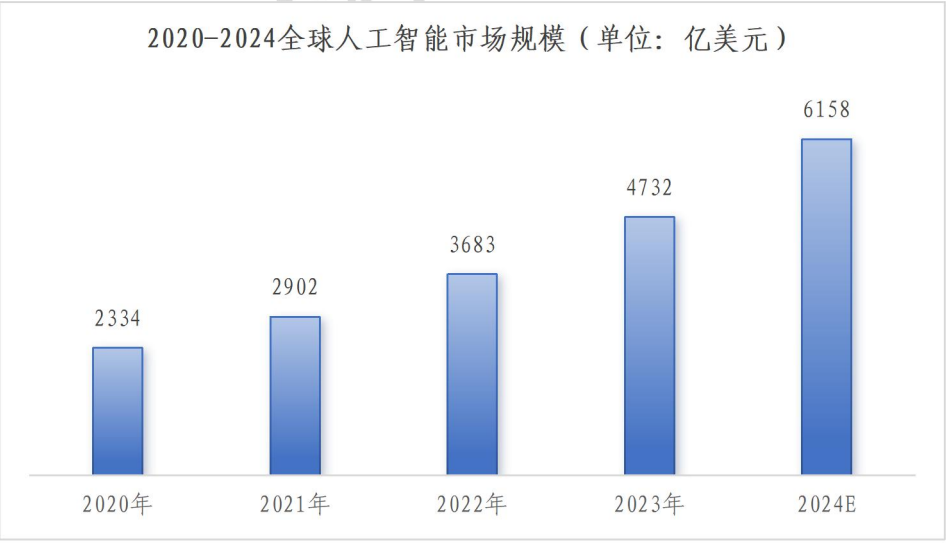
目 录

图 目 录

一、大模型技术与应用概述

（一）大模型技术发展不断演进，引领人工智能迅速发展

数字化时代快速发展，人工智能成为影响经济发展的关键力量。人工智能技术作为科技创新的核心驱动力，成为加快培育发展新质生产力的重要引擎，引领新一轮的科技革命和产业变革。根据公开数据，截至 2023 年 7 月份，我国人工智能核心产业规模已达 5000 亿元，企业数量超过 4300 家。根据《2024 年我国人工智能产业发展形势展望》报告显示，预计 2024 年全球人工智能市场规模将达 6158 亿美元，我国将突破 7993 亿元。



数据来源：《2024 年我国人工智能产业发展形势展望》

图 1 2020-2024 全球人工智能市场规模

大模型作为当今人工智能领域的核心技术之一，技术架构不断

演进。大模型的核心特征在于其庞大的参数规模和高度复杂的网络结构，通过深度学习原理，对大量的数据和计算资源进行训练，学习数据中的深层次特征与规律，实现较高的性能和泛化能力。**一是在 2017 年之前，**深度学习已经在图像识别、语音识别等领域取得了成果。**二是在 2017 年，**Google 研究人员提出了 Transformer 架构，奠定了当前主流大模型预训练算法架构的基础。**三是在 2018 年，**OpenAI 发布了大模型产品，展示了自回归语言模型潜力，能够生成连贯的文本，极大地提高了性能，预训练大模型时代逐渐来临。**四是在 2022 年，**OpenAI 发布的 ChatGPT 具备极强的对话交互能力，展现了大模型在自然语言理解和生成方面的巨大潜力，促进了 AI 应用的普及和公众认知。同时国内大模型迎来爆发期，多家企业和研究机构推出大规模预训练模型。**五是在近年来，**国内外更多大模型产品纷纷发布，更新迭代，大模型开始迈向多模态领域，不仅处理文本，还能理解图像、视频等，进一步拓宽了 AI 的应用场景。

图 2 大模型发展历程

（二）政策促进大模型应用落地，助力行业提升服务能力

我国对人工智能领域的重视程度不断提升，促进人工智能大模型技术快速发展。从顶层设计到具体实施全面布局，将人工智能转化为实际生产力，助力国家数字化战略的推进。**一是国务院于2017年发布我国在人工智能领域进行的第一个系统部署的文件《新一代人工智能发展规划》。**重点对我国新人工智能发展的总体思路、战略目标和主要任务、保障措施进行系统的规划和部署。**二是科技部等六部门于2022年印发《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》。**旨在贯彻落实党中央、国务院关于推动人工智能发展的决策部署，统筹推进人工智能场景创新，着力解决人工智能重大应用和产业化问题，全面提升人工智能发展质量和水平。**三是2024年《政府工作报告》中提出开展“人工智能+”行动。**政府工作报告第一次提出“人工智能+”，为人工智能新技术新应用创造更好的发展机遇，加速推动了数字技术和实体经济深度融合，促进了社会生产力实现新的跃升。

大模型助力千行百业提升服务效率。虽然通用大模型在人工智能领域扮演着重要角色，但它们在企业级场景中的应用常常存在缺乏行业深度、与业务结合不足等局限性。相比之下，行业大模型通过针对性地训练特定行业数据，深刻理解该领域的内在逻辑与细微

差别，能够提供更为精确、贴合实际业务场景的解决方案，展现出与行业深度融合的巨大潜力。如今，大模型已经渗透到各行各业，如金融、教育、医疗、网络安全、政务、互联网等领域，被用于智能客服、智能写作、自动摘要、文本生成、知识问答、个性化推荐等多个应用场景，改变传统生产方式，有效提升行业服务效率和服务质量，创造更高经济价值。

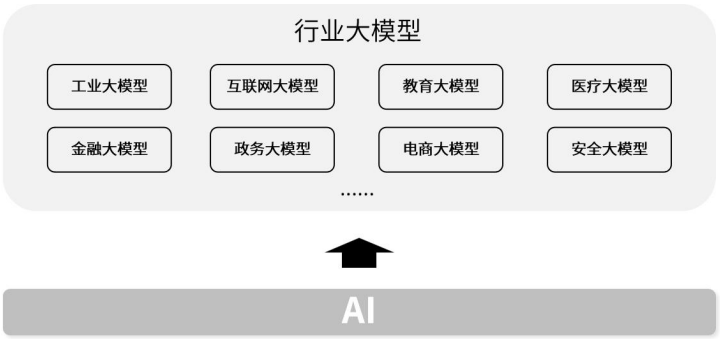


图 3 大模型行业应用

安全大模型赋能网络安全的创新变革。网络安全作为互联网发展的基石，是保障现代社会基础设施政策运作的基础。安全大模型通过整合网络安全领域的知识、技术和数据，形成统一、具有高度智能化和自适应能力的安全管理与防护系统。安全大模型凭借其庞大的参数量与深度学习能力，在深度理解行业特性和业务流程的基础上，实现对复杂攻击模式的精准判断和预警，促进安全资源的精准配置与合规性管理的自动化，为网络安全防护体系带来创新变革的智能化水平提升，构建更加智能、高效、全面的安全防护生态系统，以适应数字化时代复杂多变的安全挑战。

二、大模型技术安全行业应用现状

大模型技术的迅速发展给安全行业的核心业务场景带来深远的影响。基于对大规模数据的深度学习和复杂模式识别能力，大模型在多个关键领域展现出显著的潜力。随着技术的持续演进，大模型正逐渐成为驱动安全行业创新和转型的关键力量，推动着安全防御从被动响应向主动预防的范式转变。

（一）大模型赋能威胁检测，全面提升场景效能

1. 传统检测方法在面对新型威胁时效率不足

在当前复杂多变的威胁环境下，如图 4 所示，传统检测方法的效率和精度已无法满足不断更迭的安全需求，难以应对新型威胁。

图 4 传统威胁检测方式弊端

一是传统方法高度依赖人力投入，传统的威胁检测方法主要依靠安全专家手工定义规则，通过特征匹配的方式来识别已知威胁。这种方法对安全专家的经验和技能要求较高，且规则制定的周期较

长。一旦出现新型威胁，现有规则难以及时更新，导致检测存在盲区。攻击者可利用这一缺陷，通过变换攻击策略，规避检测。

二是传统方法难以应对未知威胁，当前网络环境日益复杂、新型攻击手段层出不穷，零日漏洞、APT 等新型威胁不断涌现，而传统方法只能检测已知威胁，对未知威胁难以奏效。攻击者不断变换策略，采用多种隐蔽技术，使攻击行为更加难以捕捉和识别。传统威胁检测因其规则更新滞后，导致检测效率低下，无法及时发现潜在威胁。

三是传统威胁检测数据处理量级受限，难以处理海量安全数据。随着网络规模的不断扩大，安全设备、终端用户数量激增，每天产生的安全日志、网络流量等数据呈爆炸式增长，达到了 PB 级别。传统方法采用关系型数据库进行存储和查询，其性能难以满足大数据处理的需求。此外，安全数据来源多样，格式各异，传统方法难以实现多源异构数据的融合分析，无法全面挖掘数据价值。

四是传统方法缺乏多维度关联分析能力，很难精准识别威胁的行为特征。攻击者往往采用多种手段，利用不同层面的漏洞，形成攻击链。但传统方法主要关注单一维度，如 IP、域名等，忽视了威胁在网络环境、时间序列上的关联性。缺乏立体化的分析视角，难以准确把握威胁全貌。传统方法分析也多采用规则和阈值的方式，泛化能力不足，难以应对威胁的多样性和变化性。综上所述，面对日益严峻的网络安全形势，传统威胁检测方法暴露出效率低、精度

差、分析能力弱等诸多不足，亟需创新的智能化手段来弥补短板，增强威胁检测的针对性和有效性。

2. 大模型提升传统威胁检测准确性和全面性

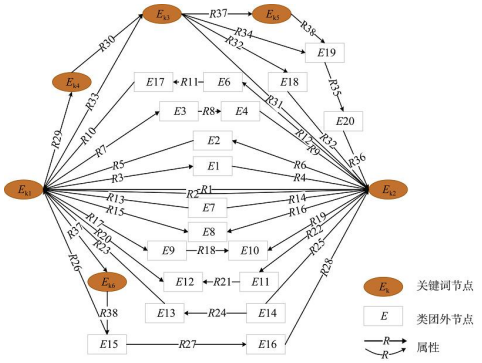
一方面，大模型能够快速处理量级较大的安全数据，利用并行计算架构实现数据的高效存储和查询，为威胁检测提供了坚实的数据基础。传统方法受制于数据处理能力，往往只能对采样数据进行分析，而大模型可以处理大量安全数据，最大限度地挖掘数据价值。通过对流量、日志、文件等数据的深入分析，大模型能够发现更多隐藏在数据中的威胁蛛丝马迹。IBM 的一项研究表明，使用人工智能和机器学习的组织在检测和响应网络安全事件方面的效率提高了 3 倍[1]。

另一方面，大模型可以从多个维度挖掘威胁的关联特征，采用深度学习、图神经网络等先进算法，从网络流量、日志、文件、行为中自动生成检测模型，显著提升检测有效性。如图 5 所示，从全局视角关联分析多源异构信息，可有效加强关键节点和类团外节点的挖掘，通过关联分析减少威胁情报的漏报和误报，显著提升检测准确率，且检测的威胁类型更加全面。麦肯锡的一份报告表明，采用人工智能和机器学习技术的网络安全解决方案可以将威胁检测的准确率提高 3 倍，同时将误报率降低 5 倍[2]。相比传统的特征工程，

[1] IBM《The What Why and How of AI and Threat Detection（2024）》

[2] 《The Four Types of Threat Detection（2018）》

大模型能够自动学习提取威胁的本质特征，捕捉威胁在不同环节、不同时间的活动规律。通过端到端的特征表示学习，可最小化人工设计的局限性，全面刻画威胁行为的异常模式。



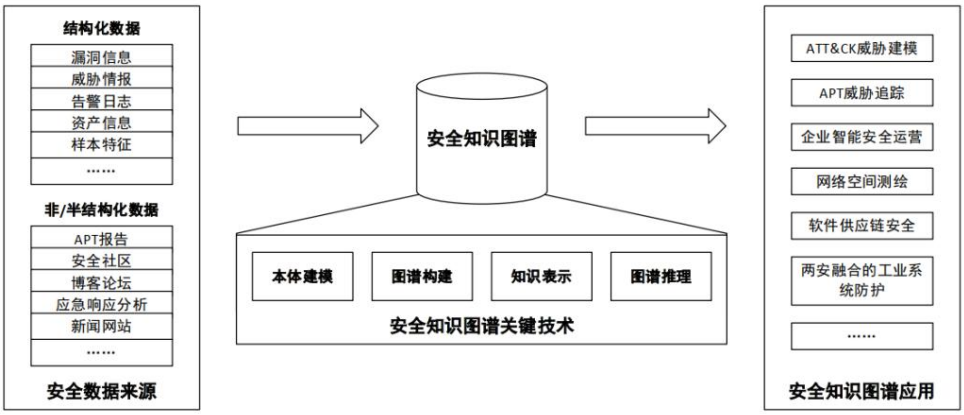
数据来源：《基于关联数据的类簇语义揭示模型研究》

图 5 大模型关联分析示意图

3. 大模型为威胁检测提供新技术和新范式

首先，传统的威胁预测主要依赖专家经验和统计模型，难以全面把握威胁的复杂演变规律。大模型则可以从海量的历史攻击事件数据中自动学习提取威胁的关键特征和演变模式，建立起威胁知识图谱[3]，从数据来源、本体设计、图谱构建等环节，为威胁检测提供技术支撑，如图 6 所示。

[3] 绿盟科技《2022 年中国威胁情报市场报告》



数据来源：公开数据整理

图 6 典型安全知识图谱构建过程

其次，基于图神经网络等技术，大模型能够深入理解攻击事件之间的内在联系和时序规律，从而对未来一段时间内可能出现的攻击类型、攻击目标、攻击手法等进行预判。例如，一些研究机构利用门控循环单元(GRU)等深度学习模型，对 APT 攻击的生命周期进行建模，通过对攻击事件的时间序列分析，预测 APT 组织未来的攻击活动，准确率可达 80%以上[4]。

最后，大模型生成的威胁预测情报可以帮助安全团队更加主动、精准地制定防御策略，提前采取针对性的防护措施，最小化安全风险。一方面，模型预警的高危攻击事件可以作为输入，触发应急响应流程，调整安全策略配置，对潜在目标进行加固。另一方面，研判攻击者的意图、偏好、未来行动，有助于围绕攻击者展开态势感知，进行威慑、诱捕、溯源等主动防御行动。

[4] IEEE《Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection（2023）》

4. 大模型强化威胁检测自适应性

一方面，大模型可通过无监督和自监督学习等技术，挖掘安全数据内在的统计规律和行为模式，构建正常基线。大模型通过学习历史未知威胁的演化规律，利用 Few-shot Learning 等算法，有望在少量样本的情况下及时识别出新型威胁，提高未知威胁的检出率。

另一方面，依托预训练模型强大的语义理解和泛化能力，可利用少量样本快速进行自适应学习，生成面向新型威胁的检测模型。《基于深度迁移学习的恶意软件检测系统研究与实现（2023）》[5]中提到，预训练语言模型强大的语义理解和表示能力可以迁移到恶意软件检测领域，并且在面对新的恶意软件时，只需少量特定领域数据进行微调，就可以适应不断变化的恶意软件。通过迁移学习技术，还能实现跨场景、跨领域的威胁知识复用，有效应对“未知未见”的安全威胁。

（二）大模型改写传统运营方式，推动全局智能化

1. 传统安全运营缺乏全局性、效率低下

一是安全运营数据割裂，缺乏统一管理和关联分析能力。企业内部不同安全设备和系统各自为政，数据散落在网络、主机、应用等不同层面，难以形成全局视图。手工梳理、关联海量异构数据，

[5] 江苏大学《基于深度迁移学习的恶意软件检测系统研究与实现（2023）》

效率低下且容易遗漏关键线索[6]。

二是规则知识库更新滞后，对未知威胁检测识别能力不足。传统安全产品主要依靠特征库、规则库等进行威胁检测，但面对快速迭代的攻击手法，知识库更新往往滞后，导致检测时效性和覆盖率不高。对零日漏洞、变种木马等未知威胁缺乏有效的发现手段[7]。

三是缺乏威胁情报的自动化生成和共享机制，企业内外部情报获取渠道有限，通常依赖人工收集整理，且情报质量参差不齐，缺乏可信度评估。情报的加工、分析、呈现流程尚未实现自动化，难以做到实时共享和快速响应[5]。

四是安全专家经验难以沉淀积累和复用，资深安全专家凭借多年经验，可快速分析事件上下文，判断威胁影响范围，给出缓解方案。但专家经验通常以隐性知识形式存在，缺乏有效的提炼方法，导致经验难以在团队内传承复用，专家离职易造成知识流失。

五是缺乏全局安全视图和威胁演化模型，企业难以精准刻画自身的安全状态，缺乏可量化、可溯源的风险评估指标。安全部门常被动应对已发生的攻击，难以未雨绸缪，对未来一段时间内的安全态势缺乏预判。

六是运营高度依赖人力，自动化、智能化水平不高，从海量告

[6] 《资产测绘与攻击面管理助力构建数字化安全运营体系（2022）》

[7] IEEE《A Survey on Deep Learning Techniques for Malware Detection and Classification（2022）》

警中甄别失陷主机、分析攻击路径、处置安全事件，需要大量安全运营人员投入，且不同层次人员职责界定不清，工作流程缺乏统一协作平台支撑，整体人效不高。综上所述，传统的安全运营方法在愈发隐秘的安全风险场景中变得不够全面，存在诸如内部数据散落、经验共享机制弱等风险，如图 7 所示。

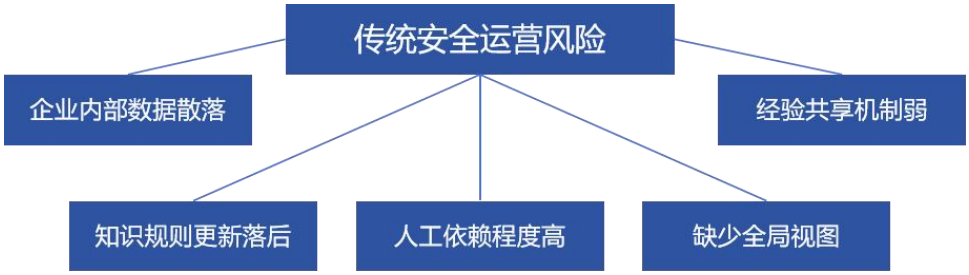


图 7 传统安全运营风险

2. 大模型实现智能关联分析，帮助挖掘安全隐患

传统的关联分析规则大多依赖人工配置，规则数量有限且容易产生误报、漏报。大模型可从海量异构数据中自动学习安全要素之间的关联模式，挖掘出难以被人力发现的隐性联系，揭示安全事件的来龙去脉。

一是大模型可对网络流量、主机日志、用户行为等多源异构数据进行融合建模，刻画各类数据之间的时间、空间、因果等多维关系。基于图神经网络等技术，大模型能够在复杂的异构图中学习顶点语义信息和边关联模式，揭示不同安全事件的内在联系，如识别出同一攻击者利用不同 IP 对多个目标发起的攻击行为[8]。

8 《2023 年中国 AI 大模型应用研究报告》

二是大模型可将原始日志数据映射到低维语义空间，实现精细化、结构化的特征提取。相比传统的规则匹配方法，大模型可以将将高维、稀疏的词向量空间映射到一个低维、稠密的语义空间，数据的维度大大降低，从通常成千上万的词汇量大小降到预设的如256或512的低维度。这样更容易进行可视化、聚类、异常检测等下游任务，如图8所示。

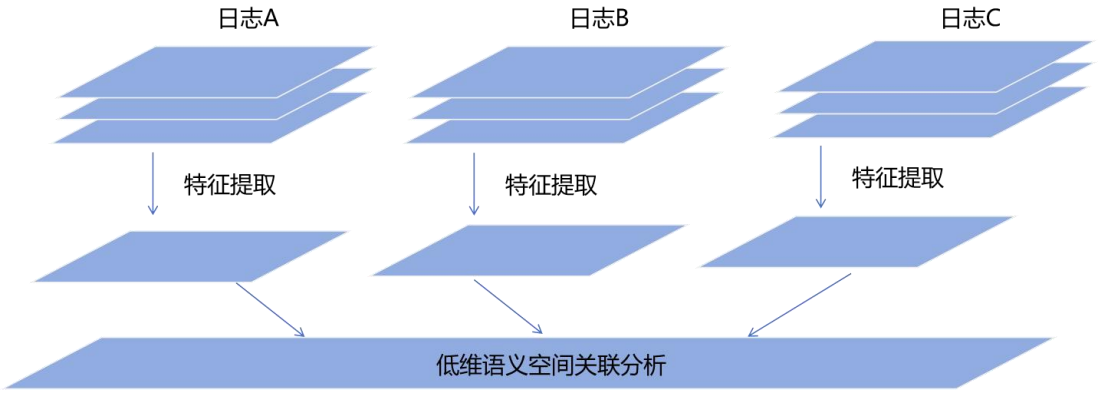


图 8 日志低维语义空间分析示意图

3. 大模型技术强化人机协同，改写传统运营分析方式

传统的安全分析主要依赖人力，专家需要花大量时间处理数据、提取特征、验证假设，分析效率难以满足日益增长的安全需求。大模型可作为安全专家的得力助手，在人机交互中发挥各自所长，形成互补协作[9]。

一是精细化运营专家风险关注点，在人机协同分析模式下，安全专家主要负责提出分析思路和假设，提供安全知识和经验见解。大模型则承担数据工程、特征工程等繁琐的任务，快速验证专家的

[9] 《自然语言交互：大语言模型带来的交互方式变革（2023）》

假设。例如，专家基于经验判断某 IP 可疑，大模型就可以快速搜索该 IP 的相关数据，提取关键特征，评估其异常程度。专家再对模型结果进行解释和优化，形成迭代优化的闭环。

二是帮助完成运营数据预处理，有效减轻人力负担。海量数据的收集、清洗、融合等前置工作都可交由大模型完成，让专家可以直接分析结构化的信息。深度学习模型可自动学习抽象出高维语义特征，揭示被数据所掩盖的关键模式，为专家的分析提供更丰富的信息。在人机协同分析中，大模型承担了类似“助理”的角色，为专家提供数据支持和分析帮助。专家则更多地充当“智囊”角色，负责结果解释、知识积累、分析决策等专业性工作。二者优势互补、协同作战，既发挥了人的探索创新能力，又利用了机器的海量计算优势[10]。

4. 大模型技术强化安全运营自动编排，构建全局防护体系

传统的自动化局限于既定场景下的处置动作，缺乏全局视野和策略思维。大模型则可通过持续学习，掌握安全策略制定、防护体系构建、响应流程优化等领域知识，具备顶层设计和纵深防御的能力。

一是，大模型可深度理解业务场景和安全目标，周期性评估组织面临的安全风险，结合外部威胁态势和内部资产状况，持续优化

[10] 《2023-2029 年中国人机交互行业现状分析与前景趋势报告》

整体安全策略。同时，大模型可实时感知网络安全态势变化，预测攻击事件的发生概率和可能路径，针对性地动态调整各领域的防护策略和响应预案[11]。

二是，大模型可自主编排端到端的安全响应流程。传统的响应流程通常是事先定义的，缺乏灵活性和适应性。大模型可结合安全事件的严重性、影响范围、攻击阶段等多种因素，实时生成最优的响应流程。例如，对于高危事件，模型会自动触发流量隔离、加固防护、通知下线等一系列联动措施；对于低风险事件，模型则会执行威胁监视、调查取证、记录跟踪等常规流程[12]。

三是，大模型可以通过平衡封堵成本和防护收益，动态调整策略阈值和流程参数，使安全编排越来越智能。大模型强化学习等技术，从海量的响应日志和效果反馈中自主学习和优化，可持续评估不同防护动作的时效性和有效性，实现告警、分析、响应、处置的全流程自动化，从而极大提升运营效率，如图 9 所示。

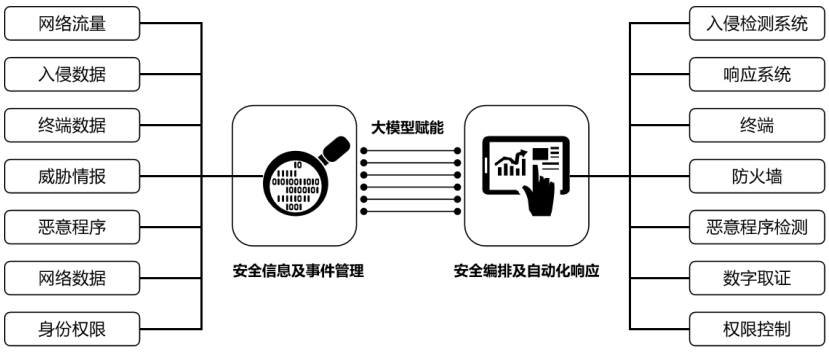


图 9 大模型安全编排自动化响应示意图

[11] 《基于安全态势感知的智能决策与联动防护方案研究（2022）》

[12] 《基于安全编排自动化与响应技术在网络安全应急响应中的应用探索（2022）》

（三）大模型推动行业信息互通，强化安全知识互联

1. 传统安全行业信息关联程度弱

传统安全行业的知识链接存在时效性差、覆盖不全、智能化程度低等问题，难以适应网络安全形势的快速演变，也很难满足从业者日益增长的需求。

一方面，传统方法通过查阅安全知识库或文档来获取信息，知识更新速度慢，难以跟上技术迭代节奏，许多安全厂商和研究机构会建立自己的知识库，囊括常见的攻防技术、漏洞分析、应对方案等，安全从业者通过关键词检索、目录浏览等方式查找所需知识，但知识库通常覆盖有限，更新滞后，检索体验欠佳[13]。

另一方面，传统搜索引擎对于安全行业知识检索效率低下，安全从业者会利用搜索引擎尝试检索互联网上的海量安全资讯，但搜索结果良莠不齐，需要人工甄别和整理，效率低下，同时搜索引擎很难无法理解专业语义，无法给出针对性的解答。

2. 大模型革新安全行业知识应用

一是大模型知识库的构建帮助夯实安全行业知识储备，一些安全企业利用大模型技术，从安全资讯、技术文档、漏洞情报等海量非结构化数据中自动提炼知识要素，构建起规模化、常更新的安全知识库。

[13] 《安全知识图谱技术白皮书（2021）》

二是大模型技术带来智能问答交互新方式，大模型智能问答机器人为安全行业带来了全新的交互方式，传统的安全社区论坛通常依赖用户之间的互动来解决问题和分享知识，但这种交互方式受到时间和人员的限制，而基于大模型的智能客服系统可以提供全时段的安全知识问答服务，用户可以随时提出问题并获得即时响应。智能客服系统利用自然语言处理和机器学习技术，理解用户的问题意图，并从海量的安全知识库中检索出最相关的答案。它们可以解答常见的逆向分析、漏洞挖掘、安全工具使用等方面的问题，为安全从业者提供便捷、高效的知识获取渠道。同时，智能客服系统还可以通过与用户的持续互动，不断学习和优化知识库，提升问答的准确性和覆盖面。这种智能化的交互方式极大地提高了安全社区的服务质量，使得安全从业者能够更加便捷地获取所需的专业知识，提升技能水平。

三是大模型技术强化安全行业的知识检索与挖掘。传统的知识检索方式通常依赖关键词匹配，难以处理复杂的语义关系和领域特征，导致检索结果的相关性和准确性不高。**一方面，**大模型技术利用深度学习算法，通过对海量安全文献、代码库、社区讨论等数据进行训练，学习到了丰富的语义表示和领域知识。这使得基于大模型的知识检索系统能够真正“理解”用户的查询意图，即使查询中包含行业专业词汇、缩写、上下文依赖等复杂因素，也能准确地推

断出用户的真实需求，返回最相关的结果[14]。**另一方面，**大模型技术还为安全知识挖掘开辟了新的途径。通过对安全领域的非结构化数据如博客文章、研究报告、论坛等进行深入分析，大模型可以自动提取其中蕴含的安全概念、漏洞特征、攻击手法等关键信息，发现隐藏的关联关系和新颖的威胁模式。这些自动化的知识挖掘功能可以帮助安全从业者及时洞察最新的攻防技术动向，掌握行业内的热点话题和前沿进展，从海量数据中汲取有价值的安全情报，提升威胁发现和响应的时效性。

[14] 《大模型在威胁情报中应用可行性研究报告（2024）》

三、安全行业大模型技术应用落地关键点

（一）发挥大模型技术优势，深入应用创新

1. 充分发挥大模型技术优势，促进安全行业智能化升级

大模型技术的持续发展正在为网络安全领域带来革命性的变革，推动整个行业迈向一个全新的防护时代。如图 10 所示，充分发挥大模型的技术优势，不仅可以显著提升威胁检测的响应效率、加强对复杂威胁的适应能力，还能够实现更加主动和智能的安全防御策略，为构建更加安全、可靠的网络环境奠定坚实基础。

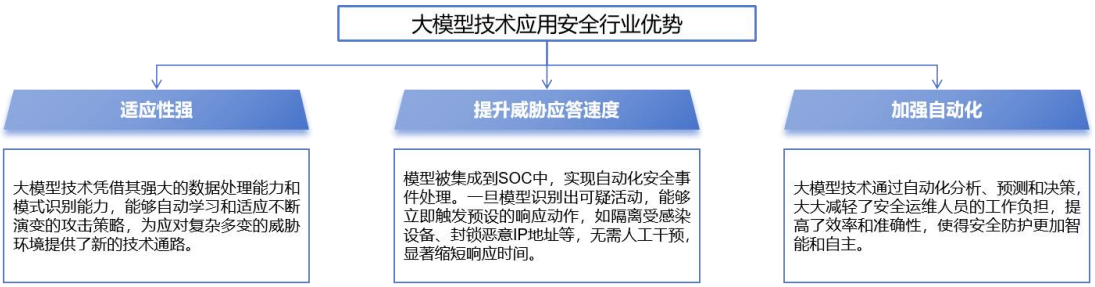


图 10 大模型技术应用安全行业优势

一是，大模型技术可以帮助安全行业应对更加复杂多变的威胁环境，随着网络空间的不断扩张和信息技术的快速发展，网络安全威胁呈现出多元化、隐蔽化和智能化的特点。传统的基于规则的安全解决方案在处理海量数据、识别新型攻击模式方面显得力不从心。大模型技术凭借其强大的数据处理能力和模式识别能力，能够自动学习和适应不断演变的攻击策略，为应对复杂多变的威胁环境提供了新的技术通路。诸如在对大数据的高效处理和分析上，通过整合

和分析来自不同来源的数据，如网络流量、系统日志、社交媒体信息等，大模型能够自动挖掘潜在的威胁信号，形成实时的威胁情报。Gartner 在 2022 年的网络安全预测报告中指出，到 2025 年，将有 40%的组织采用基于 AI 和机器学习的网络安全产品，以应对不断演变的威胁形势[15]。报告强调，大模型技术将在智能威胁检测、自动化响应和安全决策优化方面发挥关键作用。

二是，大模型技术持续提升安全行业威胁检测与响应速度，在高速发展的网络环境中，安全事件的响应速度至关重要。大模型以其大数据实时处理能力，将能够快速识别异常行为和潜在威胁，显著缩短从发现到响应的时间间隔，为组织赢得宝贵的时间窗口，有效遏制攻击扩散。大模型被集成到 SOC 中，实现自动化安全事件处理。一旦模型识别出可疑活动，能够立即触发预设的响应动作，如隔离受感染设备、封锁恶意 IP 地址等，无需人工干预，显著缩短响应时间。

三是，大模型技术强化安全行业自动化与智能化水平，随着网络攻击的频次和复杂度增加，手动分析和响应变得越来越不可行。大模型技术通过自动化分析、预测和决策，大大减轻了安全运维人员的工作负担，提高了效率和准确性，使得安全防护更加智能和自主。诸如大模型常被用于构建安全智能决策支持系统，使得这类系统能够基于历史数据和当前网络状态，预测安全风险趋势，为安全

[15] Gartner《网络安全预测报告（2022）》

策略的制定和调整提供科学依据。

2. 坚持安全行业应用创新，持续推动大模型技术落地

一是坚持基于业务场景进行创新，安全行业涉及网络安全、数据安全、终端安全等多个领域，每个领域都有其特定的技术挑战和业务痛点。如图 11 所示，大模型应用须深入一线，洞察行业需求，针对具体问题设计解决方案，只有紧贴实际，才能研发出契合行业特点，满足用户需要的应用产品，唯有以需求为牵引，以问题为导向，在跨界融通中探索突破，大模型才能在安全行业内深入应用。

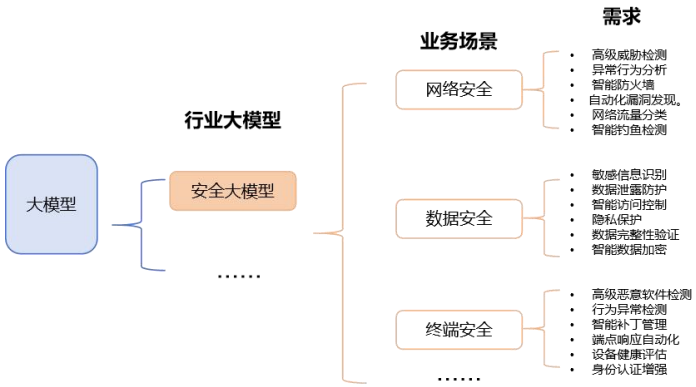


图 11 大模型应用须深入一线需求

二是勇于探索新的应用模式，传统安全解决方案往往基于规则和特征码，而大模型则可通过强大的数据挖掘和关联分析能力，实现更智能、全面的安全防护。例如基于大模型构建智能安全运营中心，通过多源异构数据关联分析，实时洞察全网安全态势；也可以利用大模型实现自适应安全策略，根据环境变化动态调整防护措施。这些创新的应用模式，有望极大提升安全防护的智能化水平。

三是建立有效的评估机制，由于安全领域的特殊性，对大模型

应用的有效性、可靠性、安全性等提出了极高要求，须建立科学的应用评估体系，全方位评估大模型的实际应用效果，包括准确率、响应时间、资源消耗等多项指标，通过客观评估，识别短板不足，持续优化迭代，以确保应用产品真正满足行业需求。

四是坚持应用创新与技术创新齐头并进，面对日新月异的网络威胁，需要在底层技术上不断精进，针对安全领域的特定问题，如对抗样本、小样本学习等，展开有效攻关，通过技术创新，不断增强模型性能，提升行业任务处理的精准度和效率，让大模型在安全领域发挥更大价值。

（二）防范大模型应用风险，加强规范建设

1. 加强数据安全与隐私保护

安全行业大模型在训练和应用过程中涉及大量高度敏感的安全数据，包括网络流量日志、威胁情报、漏洞信息、事件响应记录等。这些数据不仅包含个人身份信息、IP 地址，还可能涉及组织的核心安全策略和防御机制。

一是在大模型的整个生命周期中，必须实施严格的数据安全措施。首先，需要建立数据脱敏机制，确保在不损害模型性能的前提下，有效去除或模糊化敏感信息。可能包括使用先进的匿名化技术，如差分隐私或同态加密，以保护个人身份信息和其他敏感数据。**其次，**需要采用高强度的加密算法进行数据传输和存储，防止数据在

传输或存储过程中被截获或窃取。**再次，**还应建立全面的数据生命周期管理策略，从数据收集、处理、使用到最终的销毁，每个环节都要有相应的安全控制措施。**最后，**定期进行安全评估和渗透测试，及时发现和修复潜在的安全漏洞。通过这些全方位、多层次的安全措施，可以最大限度地保护敏感数据的安全，确保大模型能够有效利用这些数据进行训练和优化。

二是需实施细粒度的访问控制策略，严格限制对原始安全数据的访问权限，只有经过授权的安全专业人员才能接触这些敏感信息。这要求建立一个多层次、严密的权限管理体系。**一是，**应该实施强大的身份认证机制，如多因素认证，确保只有合法用户能够登录系统。**二是，**采用基于角色的访问控制模型，根据人员的职责和需求分配最小必要权限。**三是，**对安全数据进行精细化分类和标记，并为不同级别的数据制定相应的访问和处理规则。**四是，**应该实行数据访问审计，记录所有对敏感数据的访问操作，以便追踪和回溯。**五是，**定期进行权限审核和更新，确保离职员工或角色变更人员的权限及时撤销。通过这些措施，可以在保证数据安全的同时，也为合法的安全分析和模型训练提供必要的数据支持。

三是须严格遵循相关的网络安全法规和数据保护条例，需要建立完善的合规框架，确保大模型的训练和使用符合法律要求，并能够应对潜在的审计和监管检查，包括实施数据主体权利管理机制，如允许用户访问、更正、删除其个人数据的权利；建立数据处理活

动的详细记录，包括数据的来源、使用目的、处理方式；进行定期的数据保护影响评估，识别和降低数据处理活动中的风险、制定和实施数据泄露响应计划，确保在发生数据泄露时能够及时通知相关方并采取必要措施。

2. 保障安全大模型应用可信

在网络安全、数据安全和终端安全等关键领域，大模型的判断直接影响到组织的安全态势，因此安全行业的大模型应用必须具备极高的可靠性和准确性,能够经受住最严格的安全测试和全面验证。

一是安全行业大模型不仅要保证高准确率，还必须具备极低的误报率以及抗干扰性，在威胁检测、漏洞分析和事件响应等核心安全任务中，模型的每一个决策都可能影响到防御体系的整体效能，模型还需要具备强大的抗干扰能力，能够应对可能的对抗性攻击，保持在复杂多变的网络环境中的稳定性。

二是安全行业大模型的决策过程须保持高度透明和可解释性，这意味着模型不能成为“黑盒”，而是要能够清晰地展示其推理过程和依据。这对于安全分析师理解和验证模型的判断至关重要，也是满足合规要求和接受外部审计的必要条件。例如，在遭遇 APT 攻击时，模型应能够详细说明其判断依据，包括观察到的可疑行为、匹配的威胁特征等，以便安全团队能够迅速验证和采取相应措施。

三是安全行业大模型需具备完善的监控和反馈机制，一旦模型

出现异常或做出错误判断，系统应能迅速发现并报告，允许安全专家快速介入，查明原因并及时纠正。需要建立一套全面的模型性能监测系统，持续跟踪模型在实际环境中的表现，并能够快速响应和调整。

四、安全行业大模型技术应用发展趋势与展望

（一）技术日益成熟，持续赋能安全行业

当前，安全行业大模型技术正处于快速发展阶段，需要各界协同努力，共同推动这一前沿技术在安全领域的长足发展。

一是在技术创新方面，安全行业大模型将朝着更专业化和垂直化的方向发展。未来可能会出现针对网络攻击检测、恶意软件分析、威胁情报分析等特定安全领域的垂直大模型，这些模型将具备更强的领域专业知识和更高的处理效率。同时，实时响应能力和自适应学习能力也将大幅提升,使得模型能够更快速地处理海量安全数据，并从新出现的威胁模式中不断学习和进化。

二是在数据融合和分析方面，多模态融合将成为重要趋势。安全大模型将更多地整合文本、图像、网络流量、系统日志等多种数据模态，实现更全面的安全分析。这种多模态融合不仅能提高威胁检测的准确性和全面性,还能为安全分析师提供更丰富的决策依据。此外，可解释性也将得到增强，帮助用户更好地理解模型的决策过程，提高对 AI 辅助决策的信任度。

三是在隐私保护与产业协同方面，未来的安全大模型将更多地整合隐私保护技术，如联邦学习、同态加密等，基于隐私保护技术，安全大模型将突破传统的数据孤岛限制，实现云端、终端设备、物

联网等多元平台的无缝协作。这种协同不仅提高了模型的适应性和普适性，还实现了不同技术、不同产业间的有机融合，为安全应用开辟了更为广阔的发展空间。

四是在国际合作与标准化方面，面对日益全球化的网络安全威胁，未来可能会出现更多的国际合作，共同开发和使用安全大模型，并推动相关标准的制定。这将有助于应对跨国网络犯罪和高级持续性威胁，同时也为全球安全行业的发展提供统一的参考框架。

（二）产业逐渐完善，推动生态优化升级

随着人工智能技术的深入发展，安全大模型生态日益成熟，为安全产业的优化升级提供了强大动力。

1.供给侧：技术与产品持续创新，推动安全大模型应用普及

一是安全大模型技术投入持续增加，研究机构和企业有望加大对安全大模型的研发投入，在模型架构、训练方法、推理效率等技术领域不断深入，开发出更具有针对性和实用性的行业大模型，提升大模型在安全领域的适用性和性能。

二是安全大模型产品体系逐步完善，安全企业积极将大模型技术融入现有产品，推出智能化程度、业务场景贴合度更高的安全解决方案。从态势感知、威胁检测到风险评估、应急响应等，大模型有望赋能安全治理的各个环节，为用户提供更全面、更智能的安全

保障。

三是安全大模型基础设施建设加快，云服务商有望和专业安全厂商合作，推出针对安全场景优化的大模型云服务，使中小企业也能便捷地接入和使用安全大模型能力。同时，开源社区的蓬勃发展也为安全大模型的广泛应用奠定了基础。

2.需求侧：用户需求增长，驱动安全大模型市场扩张

一是用户对大模型价值认知提升，随着大模型在安全领域的成功案例不断涌现，越来越多的企业和机构将意识到大模型对提升安全管理效率、增强威胁应对能力的重要作用。安全决策者将更倾向于在预算中增加对大模型相关产品和服务的投入，从而带动安全大模型市场的持续发展。

二是安全智能化需求推动大模型技术深入应用，面对日益复杂的网络安全威胁，传统的人工分析方法难以应对，安全团队需要借助大模型的智能分析能力，通过海量数据的快速处理、复杂威胁的精准识别和高效的决策支持等能力，实现“检测、响应、处置”的智能化运营闭环，通过深入应用大模型技术提高安全运营整体自动化水平，强化整体安全运营效率。

（三）标准体系愈发清晰，行业应用更趋规范

随着大模型技术在安全领域的深入应用，标准化体系的建立和完善将成为推动行业规范化发展的关键因素。

一是在技术标准制定方面，安全大模型的相关核心技术标准有望加快制定周期，业内产、学、研各组织将积极参与制定模型架构、训练方法、数据处理等核心技术环节的标准化规范。这些标准的落地将为安全行业厂商的产品提供良好的基础，推动整个安全行业大模型技术的进步和产品成熟度的提升。

二是在安全准则落地方面，安全行业的场景应用对安全准则提出更高的落地要求，随着实践探索的日益深入，安全大模型应用将更加规范。行业领军企业和研究机构将联合制定详细的安全准则实施指南，为业内开发者提供可操作的安全措施。同时，用于评价安全大模型应用的自动化测试工具将被更为广泛的应用，以提高业内应用的整体合规水平，通过覆盖开发、测试的实施指南实现产品持续迭代和质量提升的闭环，进一步推动安全准则与具体安全行业场景的深度融合。

附录 大模型技术在安全行业的创新应用案例

案例 1 基于 AI 大语言模型的 SAST 安全治理

一、需求背景

某安全厂商在进行软件供应链安全治理过程中，使用静态应用程序安全测试(SAST)工具通过分析源代码、字节码或二进制代码来发现安全漏洞。但是，传统的 SAST 工具可能产生大量的误报，需要人工确认，并且针对漏洞的修复，安全人员与开发人员存在较大的沟通成本，从而带来更高的运营成本。

二、案例介绍

为了应对传统静态应用程序安全测试（SAST）工具在误报率和运营成本上的挑战，采用了大模型（LLM）与 SAST 工具相结合的解决方案。

1.技术架构

- 代码扫描模块：SAST 工具负责初步的源代码、字节码或二进制代码扫描，生成初步的漏洞检测报告。
- 大模型（LLM）验证模块：利用训练有素的大型语言模型对初步检测结果进行二次验证。大模型可以对静态分析结果进行语义理解，精准判断漏洞的真实性。
- 修复建议生成模块：基于大模型对每个实际存在的漏洞生成

具体的修复建议代码片段，供开发人员直接使用。

2.主要能力

- 误报率降低：通过使用垂直领域大模型对数据流污点传播过程构建检测工作流进行二次验证，发现检测的误报率显著降低至 20%以下，相较于传统方法减少了 20 个百分点。
- 高效沟通与协同：解决方案能自动生成修复建议，减少安全人员与开发人员之间的沟通成本，从而提高整体运营效率。
- 高修复成功率：针对代码中的漏洞部分，生成的修复代码可以直接修复漏洞的成功率超过 80%，开发人员能够高效地将其应用于实际项目中。

三、用户价值

项目通过创新性地将大模型（LLM）与静态应用程序安全测试（SAST）工具相结合，解决了传统 SAST 工具的高误报率和高运营成本问题。采用先进的污点分析技术和大模型集成，提高了漏洞检测的准确性，将误报率降低到 20%以下。同时，通过大模型生成自动修复建议，提升了开发人员的修复效率，修复成功率超过 80%。实现了从检测到修复的高效闭环，推动了安全监测技术的智能化发展。

案例 2 某企业安全运营智能体平台

一、需求背景

某企业在日常运维中面临着海量的安全日志分析带来大量误报和漏报的问题,安全运维人员每天都要处理大量的监控日志和警报，如何从中筛选出真正的威胁信息成为一大挑战。误报和漏报的问题也可能导致团队浪费大量时间在无关紧要的警报上，进而错过实际威胁。

二、案例介绍

安全运营智能体平台提供了全流程的安全运营智能化解决方案，采用分层的技术架构来不断优化智能运营场景的业务效果。自底向上分成模型层、框架层和应用层。

图 12 安全运营智能体平台技术架构

模型层采用 Post-Pretrain、SFT、Dagger 等机制，通过安全专业领域语料提升基础模型安全知识水平；利用安全专家经验提取的思

维链、任务规划、工具使用等运营数据，构建 TTP 框架映射，形成数据飞轮，助力安全大模型在安全攻防方面推理能力不断提升。

框架层设计实现满足专家与智能体协作的机制，使得基于大模型的智能体落地到实际安全运营场景更可控、更可被观测。通过动静态结合的工作流编排，在运营流程的可控性和发挥大模型的脑洞思维方面取得平衡。构建安全领域知识管理能力，沉淀私有的专家运营经验，结合外部安全情报和内部资产数据，通过检索增强技术（RAG）等方式提供给智能体使用，使得大模型更了解企业内部的运营偏好及当前的安全态势，使模型决策和推理更有依据。搭建安全领域的工具集，减少安全产品孤岛效应、充分解放工具的效率。

应用层采用 LUI、GUI 相结合的设计使得安全专家更方便的控制整个安全运营流程及跟 Agent 交互。静态编排的工作流和动态生成的工作流，满足智能化安全运营全场景需求。构建了安全运营关键环节 Agent，结合到安全运营平台线上化、自动化、智能化的处置安全告警和安全事件。

三、用户价值

有效提升了入侵检测、告警研判、事件分诊、溯源、报告等环节的效率和效果。特别是在告警运营、事件处置等关键环节，大幅提升了安全运营的效率。在关键场景的指标达到 90%或以上的准确率。

案例 3 基于 AI 大模型的新一代安全运营平台

一、需求背景

某集团在信息安全运营中引入“安全垂直领域 AI 大模型”，逐步实现从“人工运营”为主到“机器运营”为主的智能转变，用安全大模型理解利用 AI 生成的新型攻击，协助集团信息安全运营人员进行主动安全运营，构建“AI+人工”共治的新一代安全运营平台。帮助某集团保护核心资产，建立完善的信息安全运营体系，提升信息安全运营效率，提高安全威胁防御能力，缓解专业运营人员不足的难点问题。

二、案例介绍

基于 AI 大模型的新一代安全运营平台从下至上分为四层：模型设施层、模型训练层、业务应用层、产品联动层。

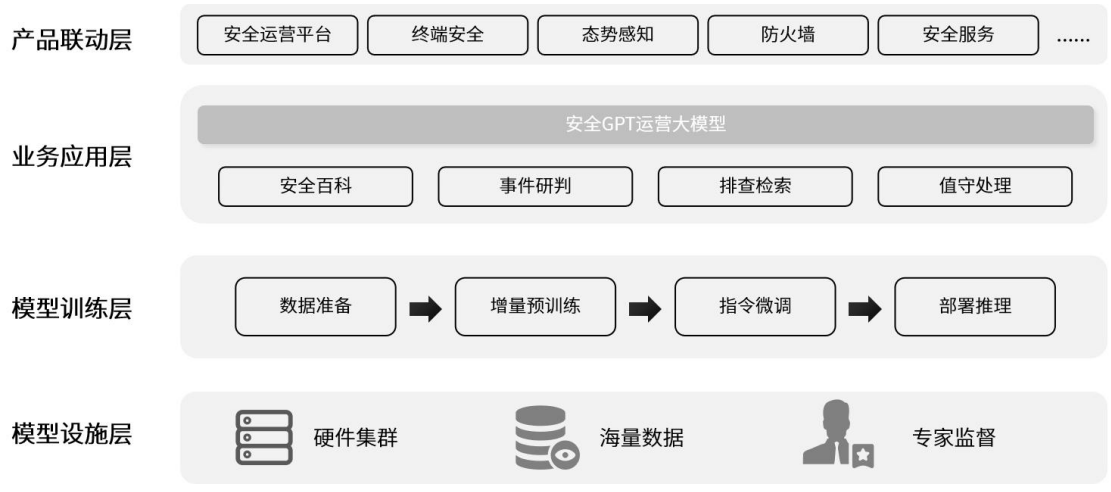


图 13 安全运营平台技术架构

安全 GPT 利用海量语料和安全领域高质量安全知识，经预训练

和指令精调等阶段，学习大量安全语言范式和思考逻辑，实现网络安全领域高质量内容的生成、推理、摘要、总结等核心能力。通过与集团现网的终端侧安全组件、网络层安全组件、运营平台、以及相关的资产数据对接，上传 100 多台安全设备日志辅助训练和微调，让安全 GPT 运营大模型更懂集团的实际业务。并通过安全分析人员与大模型进行自然语言交互的方式，辅助分析师快速看懂问题、调用工具和流程，完成数据的分析统计等工作，提高安全运营的效率。打造出基于安全 GPT 为内核的“虚拟安全专家”，实现全天 24 小时主动值守。

三、用户价值

从最开始的人工运维阶段成功迈向了智能运营时代，实现了企业互联网安全的实战化、体系化、常态化和智能化运营。全集团范围内日均产生安全告警从之前的 40 多万条降低至 2000 多条，日均自动化封禁扫描恶意超过 200 个 IP，封禁时间从之前的 30 分钟降低至 10 秒以内，阻断外网扫描次数超过 10 万次。在 GPT 的运营辅助下，集团信息安全运营效率直线提升。

案例 4 智慧家庭“小微”安全运维体系

一、需求背景

随着家庭业务数据规模的不断提升，安全场景的专业性复杂性指数级增加，智慧家庭运营中心的安全管控能力要求也越来越高。在安全管理过程中，用户往往面临着各种问题和挑战，需要快速获取相关的知识和解决方案,传统的安全管理方式已经难以满足高效、精准的需求。

二、案例介绍

该项目基于检索增强生成技术（RAG）注入多维公域、私域安全知识，是安全领域的专业私人助理。能够精准解答安全领域的各类专业问题，迅速协助发起流程、搜寻所需的文件材料。通过将公域和私域的安全知识收集整理后，切割为文字片段。经向量化后存入向量数据库。在安全工作者进行提问时，将历史对话和问题组合起来，进行向量化后，在向量数据库中进行检索。将检索到的知识内容，通过 prompt 模板渲染后，整体输入大模型中，利用大模型的文本能力进行个性化回答。

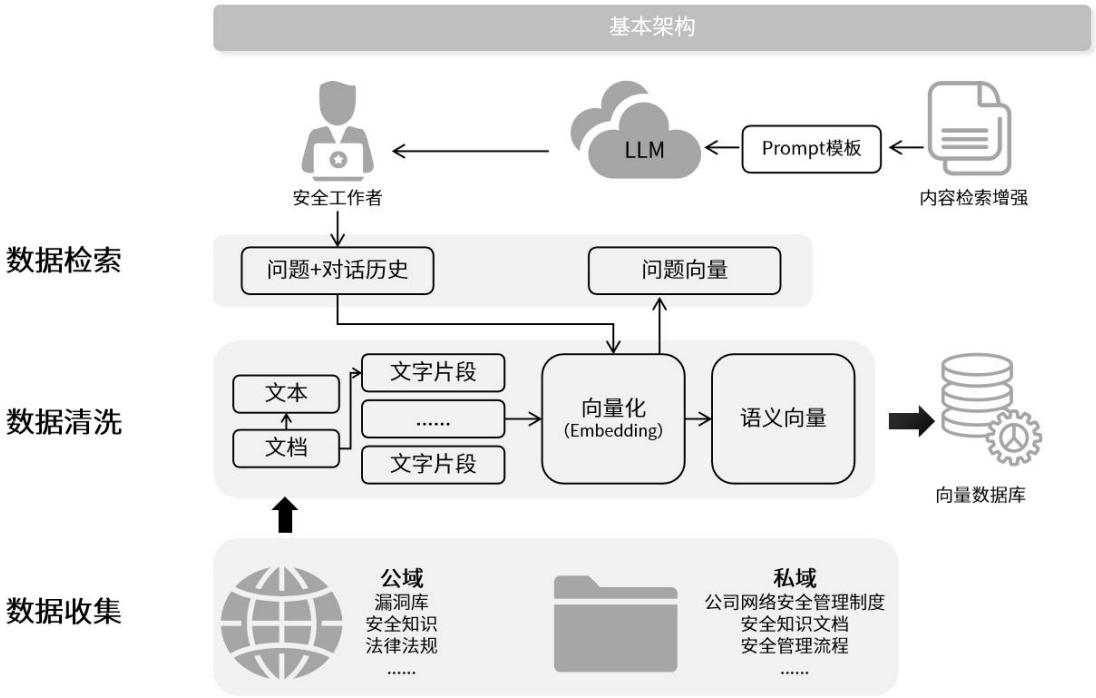


图 14 “小微”基本架构图

三、用户价值

案例服务全网 2 亿高清用户、4000 万安防用户及 2000 万户智慧社区、数字乡村等新场景解决方案后台安全服务场景，快速复用安全知识实现安全能力快速响应。

案例 5 微调大模型智能修复漏洞

一、需求背景

在传统的代码漏洞修复流程中，首先通过配置 DevOps 流水线来识别代码库中的漏洞，随后发出修复工单指派给原开发工程师。工程师根据提供的漏洞修复建议进行问题解决，修复完成后的代码被提交回代码库，并再次进入 DevOps 流水线进行复测。整体流程中存在漏洞检出和修复的时间延迟、修复质量不一致、漏洞复测的效率难把控、漏洞修复反馈循环长等问题。

二、案例介绍

该解决方案采用了基于开源的 CodeLlama 构建的漏洞修复大模型技术架构，通过对历史修复代码样本进行训练，从而继承并发扬过往的优秀修复实践打造出在漏洞修复任务上表现卓越的新模型。此外，利用这一新模型，开发了一个集成开发环境（IDE）插件，该插件自动产生修复后的代码，并将其推送至 IDE 界面。使得开发工程师能够在代码开发阶段便完成代码的合并，从而高效完成漏洞修复工作。

三、用户价值

在某企业开发团队中应用，平均每周生成修复代码超过 1w 行，修复通用漏洞超过 300 个，提高了修复过程的效率和一致性，加速了开发流程，推动了安全开发理念的进步。

案例 6 某集团安全 GPT 应用实践

一、需求背景

在外部层面，某集团业务广泛，涉及对消费者的 IoT 业务，ToB 的楼宇、机器人自动化等业务，还涉及到跨国海外业务合规等问题，需要满足监管合规要求。在内部层面，数字化程度较高，数智化业务系统涵盖广，员工安全意识参差不齐，使得企业面临整体安全情况更加复杂。

二、案例介绍

智能化化安全运营体系分为三层：安全大脑（安全 GPT），联动层（XDR-SOAR），执行层（安全组件）。

通过安全 XDR 整合原有安全组件和业务系统集成的基础上，将海量安全日志清洗后，上传至云端安全 GPT 大模型；采用 SaaS 化方式部署安全运营大模型，将 XDR 清洗后的数据加密脱敏上传至云端安全大模型，通过云端强大算力支持，实现安全运营智能化。

通过大模型的预先训练和基于场景化的工作实践经验，将资产梳理、加固预防、监测研判、调查处置、联动处置、情报查询及溯源总结等方面的工作能力流程化，使安全运营人员可以通过自然语言与安全大模型进行交互，以快速问答的方式，调动对应的工具、人员和流程，提升安全运营效率。

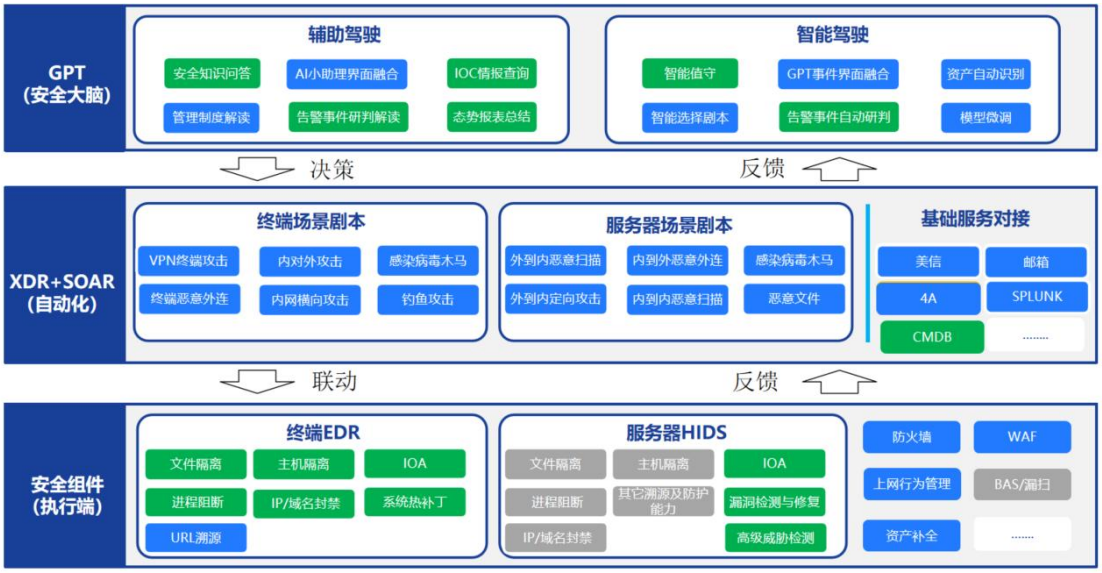


图 15 智能安全运营技术架构图

三、用户价值

安全 GPT 落地后让运营人员在广度和深度上都能做全局把控。在广度上，少量运营人员即可守护数十万资产，每天只需关注安全 GPT 逐一研判后的 100 余条高危告警，准确度超过 95%。在深度上，安全 GPT 对任意一条告警都可解读，直观呈现完整分析过程，帮助运营人员更好理解攻击意图，快速完成研判决策。事件研判平均用时从 30 分钟以上减少到 5 分钟以内，从依赖“个别安全专家”模式转换成“安全 GPT 辅助+专业安全服务团队”模式，补齐了人才短板。通过安全 GPT 落地，增强了安全威胁检测及响应能力，综合提升了安全运营整体效率，实现安全运营的提质增效。

致谢

本报告在撰写过程中得到了以下单位的支持和帮助，在此表示感谢。由于撰稿时间有限，行业态势变化快，如有疏忽和纰漏，欢迎批评指正。

中国信息通信研究院、天翼云科技有限公司、浪潮云信息技术股份公司、华为云计算技术有限公司、腾讯云计算（北京）责任有限公司、三六零安全科技股份有限公司、中移（杭州）信息技术有限公司、天融信科技集团股份有限公司、北京百度网讯科技有限公司、蚂蚁科技集团股份有限公司、天翼安全科技有限公司、北京火山引擎科技有限公司、启明星辰信息技术集团股份有限公司、奇安信科技集团股份有限公司、中移动信息技术有限公司、五色石（杭州）数据技术有限公司。

（排名不分先后）

可信安全

CONTACT US

若您对本报告有任何建议，请与我们联系：
mamingyang@caict.ac.cn