网络安全2026

启航·十五五

CYBERSECURITY 2026-REGULATION POSTURE AND TECHNOLOGIES

NSFOCUS

绿盟科技集团股份有限公司（以下简称绿盟科技），成立于2000年4月，总部位于北京。公司于2014年1月29日在深圳证券交易所创业板上市，证券代码：300369。绿盟科技在国内设有50余个分支机构，为政府、金融、运营商、能源、交通、科教文卫等行业用户与各类型企业用户，提供全线网络安全产品、全方位安全解决方案和体系化安全运营服务。公司在美国硅谷、日本东京、英国伦敦、新加坡及巴西圣保罗设立海外子公司和办事处，深入开展全球业务，打造全球网络安全行业的中国品牌。

目录

CONTENTS

卷首语

2025年，国家“十四五”规划全面收官，“十五五”规划统筹部署、蓄势待发。一年中，国家持续深化网络安全能力和体系建设，锚定“扎实推动高质量发展”全局目标，围绕战略规划、法规标准、市场推进、产业赋能等要素综合施策，切实推进着我国网络安全保障体系持续完善发展。

回顾过去的一年，AI安全、数据安全、AI攻防博弈、产业链供应链韧性与安全等等，成为国内外网络安全领域热议的高频词汇，也同样成为引领网络安全行业发展新的市场“风口”。探究其背后的规律，则是新质生产力、人工智能+、全国统一大市场、产业转型升级等经济社会发展内在需求在网络安全领域的体现。

绿盟科技作为深耕网络安全产业前沿的一份子，密切关注国内外网络安全发展态势，并积极赋能网络安全供给侧创新。为此，我们依托自身研究队伍积淀，结合持续热点跟踪，将核心研究成果集结成册，形成本报告。

本报告包括三个篇章，即：宏观观察篇、安全态势篇、技术发展篇，筛选汇聚了绿盟科技本年度在网络安全跟踪研究中的重点研究成果。其中，宏观观察篇综述了年度法规政策要点，并筛选分析了我国和国外重点国家年度内发布的10部网络安全热点政策法规；安全态势篇重点梳理分析了我国网络安全年度发展的5大领域态势；技术发展篇重点梳理并分析了网络安全年度发展的9项代表性技术。

辞旧迎新之际，寄望本报告能为支撑国家网络安全决策、服务行业网络安全建设略尽绵薄。并期待继续依托我司技术产品和服务，秉承“专攻术业，成就所托"的宗旨，助力客户"十五五"平稳开局、共同推动国家高质量发展，并为全面加强国家网络安全保障体系和能力持续贡献力量。

胡忠华

2026年1月

01

宏观观察篇

1.1 “十五·五”开局

《中共中央关于制定国民经济和社会发展第十五个五年规划的建议》（以下简称《建议》）从国家安全体系和安全能力建设的视角，明确了未来五年我国网络和数据安全发展的目标和任务体系，可以从战略和战术两个层面进行分析和理解。

首先从战略层面来看。《建议》遵循“统筹发展和安全”的总体国家安全观基本纲领，并着重明确将“防范化解各类风险，增强经济和社会韧性”作为未来五年安全发展的主线。作为国家安全重要构成的网络和数据安全，其发展也必然服务于这一主线，协同致力于以新安全格局保障新发展格局。

其次从战术层面来看。《建议》对于网络和数据安全的规划部署主要涉及以下四个细分领域。

一是加快科技自立。技术的自主和自立是网络和数据安全的重要动力，《建议》提出强化关键核心技术攻关，其中的集成电路、基础软件等都是网络和数据安全的重要基础技术领域。《建议》结合数字中国建设，提出构建开放共享安全的全国一体化数据市场，这无疑将成为数据要素基础制度建设的重点任务之一。另外，人工智能等数智技术创新也是《建议》明确的赋能安全重要方向。

二是健全安全体系。《建议》强调了国家安全体系的构成，即法治体系、战略体系、政策体系、风险防控体系，这为网络和数据安全体系化发展给出了明确对标框架。《建议》同时也强调了国家安全重点领域和重要专项协调机制的建设，具体到网络和数据安全领域而言，包括但不限于管理协调机制、安全审查协调机制、数据保护协调机制等等。

三是加强安全能力。《建议》对于十五·五时期网络数据安全能力的任务，明确了两个方面。包括：安全基础保障方面的重要产业链供应链安全、重大基础设施安全；以及新兴领域的网络、数据、人工智能、低空等安全能力建设。

四是强化安全治理。《建议》从网络和数据安全综合施策协同治理的角度，明确了重点治理任务包括：加大预防和打击电信网络诈骗、深化网络空间安全综合治理，加强个人信息保护等。

1.2 国内网络安全政策发展

1.2.1 年度综述

首先，从行业分布来看，发布数量占比较多的包括：网信 (36.6%) 、数据 (15.3%)、工信（13.8%）、金融（5.7%）、密码（2.1%）、能源（1.8%）等行业。其中尤为值得关注的是，随着数据要素管理机制的逐步健全，我国数据行业法规制度体系在今年呈现了迅猛的增长势头，实现了从数据要素理论体系向现实管理制度体系的快速转化。目前我国数据制度体系已涵盖数据权属、数据流通、数据交易、数据基础设施等多个方面。

其次，从内容领域来看，发布数量占比靠前的包括：网络安全（19.6%）、数据安全（18.7%）、数据要素（16.5%）、人工智能（11.7%）、技术产品管理（11.4%）、个人信息保护（8.1%）等领域。

在这些内容板块中，各自又包含了许多不同的细分子领域，一定程度上可反映不同领域在网络安全建设中所处的不同阶段，以及其对于安全关注的不同侧重。

1. 网络安全。侧重点包括网络安全事件管理、网络安全等级保护、关键基础设施网络安全、网络安全分类分级、网络安全人才和素养、卫星及无线电网络安全等。涉及的行业主要包括工业、通信互联网、金融、应急、安全等。

2. 数据安全。侧重点包括数据流通安全、数据出境安全、数据隐私技术、数据内容安全、数据合规监管、数据共享安全等。涉及的行业主要包括金融、能源、政务、卫健、自然资源、商贸等。

3. 个人信息保护。侧重点包括个人信息出境认证保护、个人信息保护合规审计、人脸识别技术应用管理、敏感个人信息安全管理、未成年人个人信息保护等。涉及的主要行业包括网信、市监、通信、医疗、民政、公安等。

4. 数据要素。侧重点包括可信数据空间、高质量数据集、一体化算力网、数据基础设施体系、数据资产管理、公共数据资源开发运营、数据开放共享、数据产权和流通交易等。涉及的行业主要包括财政、物流、数据、工信等。

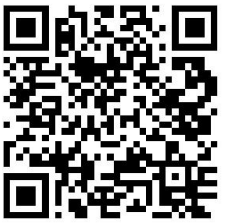
5. 人工智能。侧重点包括人工智能标识管理、人工智能安全治理框架、人工智能备案管理、人工智能技术测评、政务大模型应用、人工智能科技伦理管理等。涉及的行业主要包括政务、交通、教育、能源、知识产权、气象等。

此外，网络空间治理、数字经济、网络安全专项资金管理、网络安全国际合作等也是我国 2025 年度网络安全法规政策的重要关注领域。

1.2.2 热点政策法规分析解读

1.2.2.1 【网络安全】《关于对网络安全等级保护有关工作事项进一步说明的函》（公网安〔2025〕1846 号）

2025 年 4 月 27 日，公安部发布《关于对网络安全等级保护有关工作事项进一步说明的函》（公网安〔2025〕1846 号）。该文件是对 2025 年 3 月 8 日发布的《关于进一步做好网络安全等级保护有关工作的函》（公网安〔2025〕1001 号）的进一步补充说明，文件内容涉及等保业务的重大更新调整及说明，要求各级单位深化系统备案更新、数据资源摸底及风险整改，并对部分关键问题进行了指导说明。



绿盟简评：

（一）主要内容

1846 号文件的主要内容是以问答形式对 1001 号文件的要求给出细化和解释说明。文件附件《关于进一步做好网络安全等级保护有关工作的问题释疑》包含 24 条问答，其内容可以概括为五个方面。一是系统备案动态更新工作要求，包括备案管辖、备案证明有效期等；二是第五级信息系统定级工作要求，包括定级范围、测评依据、测评频次等；三是数据摸底调查工作要求，包括数据摸底的必要性、《数据摸底调查表》如何填写和报送等；四是风险隐患排查工作要求，包括高风险判定依据、与重点风险隐患的关系等；五是制定保护方案工作要求，包括方案的范围和内容、方案涉及资产的上报时间等。

（二）主要变化

与此前的等保备案工作相比，1846 号文件主要呈现以下 5 个方面的不同。

1. 备案工作增加了数据摸底调查专项内容。
2. 将等保测评结论由此前的百分制评分改为三类结论模式（符合、基本符合、不符合）。
3. 测评报告模板新增“重大风险隐患及整改情况”附录，体现从“合规达标”向“动态防护”目标思路的转变。
4. 测评报告模板部分具体要求的变化，如渗透测试细化、被测系统内部安全域双维度拓扑展示、风险评估标准更新等等。
5. 测评报告模板对部分章节的问题描述和风险分析内容进行了更细化的阐述。

（三）重要影响

一是进一步强化了关基和等保的衔接。《网安法》中明确规定了“对关键信息基础设施，在网络安全等级保护制度的基础上，实行重点保护”（第 31 条）；2020 年公安部发布的《贯彻落实网络安全等级保护制度和关键信息基础设施安全保护制度的指导意见》，则进一步强化了二者关联的协同保护要求。

二是进一步强调和明确了关基认定的依据。《网安法》规定“关键信息基础设施的具体范围和安全保护办法由国务院制定”（第 31 条）；而落实层面的《关基条例》对关基的认定则仅给出了“范围列举+授权认定”的框架方法（第 2 条、第 9 条）。而此次发布的 1846 号文件及此前发布的 1001 号文件，明确指出“将第五级信息系统作为关键信息基础设施认定的重要因素”。虽然文件也明确二者“并不是等同关系”，但这无疑为关基认定提供了一个重要且具有很强实践基础和操作性的认定依据。

三是等保备案更新工作的启动，对于相关的数据安全保护、关键信息基础设施保护工作或将产生综合影响。等保工作作为我国网络安全领域跨行业、跨领域的最重要

网络安全运行制度之一，自 2007 年实施已近二十年，且该制度自身也经历了持续的更新发展。从网络安全保障要素的角度看，等保制度已基本囊括了我国网络安全保障的主要合规要素；从工作机制来看，等保积累形成了较为体系化的工作和管理机制。因此，在等保备案更新工作推进的同时，进一步加强等保机制与网络安全其他工作机制之间的协同配合，无论对于促进国家网络安全保障能力和体系提升、还是对于推进网络安全产业的纵深发展，都会产生重要影响。

1.2.2.2 【数据安全】《数据出境安全评估申报指南（第三版）》

2025 年 6 月 27 日国家网信办发布《数据出境安全评估申报指南（第三版）》。《指南》对数据处理者申报数据出境安全评估需要提交的相关材料进行了优化简化，明确数据处理者申请延长数据出境安全评估结果有效期的条件、流程、材料等内容。《指南》规定，数据处理者因业务需要向境外提供重要数据和个人信息，符合数据出境安全评估适用情形的，应当根据《数据出境安全评估办法》和《促进和规范数据跨境流动规定》，按照申报指南申报数据出境安全评估。评估结果有效期届满，符合申请延长评估结果有效期条件的，数据处理者可以在有效期届满前 60 个工作日内提出延长评估结果有效期申请。

绿盟简评：

（一）法规政策脉络简要回顾

回顾我国数据出境安全评估制度的法律法规脉络，可以大致归纳为三个层面。

一是法律层面。数据出境安全评估制度的上位法依据，主要包括《网络安全法》第 37 条、《数据安全法》第 31 条、《个人信息保护法》第 40 条三部分。这些上位法对于重要数据和个人信息的出境安全评估做出了制度授权并预留了法规细化空间。

二是规章政策层面。数据出境安全评估制度的细化和统一，历经多次调整演变。先后经历了《个人信息和重要数据出境安全评估办法（征求意见稿）》（2017）、《个人信息出境安全评估办法（征求意见稿）》（2019）、《数据出境安全评估办法（征求意见稿）》（2021）三个主要文件。2022 年正式发布的《数据出境安全评估办法》，可谓是该制度在规章层面的集大成者。它实现了数据出境安全管理制度的两个统一：一是将个人信息、重要信息出境纳入统一管理，明确数据出境评估的范畴；二是将“关键基础设施重要数据”和其他重要数据纳入统一管理，不再区分适用不同规则。

2024 年《规范和促进数据跨境流动规定》的发布将数据出境安全评估制度做了进一步优化完善。也奠定了现行数据出境安全评估依据“1 办法+1 规定”的基本格局。

三是在管理规范层面。国家网信办发布并持续更新了三版《数据出境安全评估申报指南》，从实际操作角度确保数据出境安全评估制度的切实落实和实施。

（二）内容要点和变化

从内容上看，《数据出境安全评估申报指南（第三版）》有以下几处主要变化。

一是框架调整。新版指南保留了第二版的“适用范围”、“申报方式及流程”、“咨询、举报联系方式”三个章节；增加了“申请延长评估结果有效期”一个章节。此外，新版指南还将第二版的“申报材料”章节合并到申报方式和流程中，作为一个独立条款。

二是内容优化。新版指南结合申报方式和流程要求，重点对在线申报系统的使用方法和流程做了较大篇幅的细化和明确。此外，还对数据处理者申请延长数据出境安全评估结果有效期的条件、流程、材料等内容用专门章节进行了明确。

另外，新版指南还对正文及附件材料的部分内容的表述进行了优化简化调整。

（三）影响思考

首先，新版指南的出台，进一步提升了申报数据出境安全评估的便利性，其对相关内容的简化明确，也有助于减轻数据处理者相关申报负担。

其次，在牵引行业发展方面，新版指南也将起到积极的市场推动作用。

一是促进数据出境安全评估等服务供给。一方面可促进面向数据处理者的专业化数据出境安全服务发展，如数据出境自评估咨询、出境数据资产审计等。另一方面，可促进面向数据评估专业机构的支撑服务供给，如出境重要数据识别服务、数据出境安全事件处置等方案和产品研发等。此外，还能促进数据安全出境相关培训服务的发展。

二是拉动数据出境安全监管支撑需求。数据出境安全评估市场的稳定发展，离不开国家相关职能部门监管机制和手段的健全完善，这无疑会扩大对监管支撑相关服务的需求。主要包括：数据出境风险监测、敏感数据威胁情报支持、数据出境安全协同管理平台等等。

1.2.2.3 【个人信息保护】《个人信息出境认证办法》

2025 年 10 月 14 日，国家互联网信息办公室与国家市场监督管理总局联合发布《个人信息出境认证办法》，自 2026 年 1 月 1 日起正式实施。该办法作为《个人信息保护法》的重要配套规定，明确个人信息出境认证的适用情形、申请流程及监管要求。此举将规范个人信息出境活动，在保障个人信息权益的前提下促进数据跨境安全有序流动。

绿盟简评：

（一）背景简析

个人信息出境管理是我国个人信息保护的重要领域之一。此前，国家市场监督管理总局、国家互联网信息办公室于 2022 年 11 月联合发布《关于实施个人信息保护认证的公告》、《个人信息保护认证实施规则》，标志着个人信息出境保护认证这一重要管理机制正式启动。

国家网信办又于 2025 年 1 月发布了《个人信息出境个人信息保护认证办法（征求意见稿）》，旨在进一步明确和规范相关管理要求。此次发布的《个人信息出境认证办法》（以下简称《办法》）即是在《征求意见稿》基础上发展完善而来。

（二）变化和看点

《办法》除了进一步规范个人信息处理者、认证机构两类主体应承担的合规义务之外，其更在管理机制、监管机制、制度衔接等多个方面呈现诸多重要变化。

看点一：升级备案管理。

《办法》进一步增加了备案管理的要求，设置“不予备案”的选项。反映了管理部门对认证机构趋严的管理态度，即：即使认证机构已经取得“个人信息保护认证资质”，但若其未通过备案审核，则仍无法承接相关认证业务工作。

看点二：完善监管体系。

将“国家数据管理部门”纳入监管体系。从第四条、第十二条的规定看，国家数据管理部门主要参与“制定有关个人信息出境认证相关标准、技术规范”，并参与对认证机构的备案审核等两项重要工作。

看点三：增加评估环节。

《办法》将“个人信息保护影响评估”明确为个人信息处理者申请个人信息出境认证的前置条件，重点考察相关个人信息的转出者、境外接收者是否具备相应保护条件，以及个人信息是否面临风险等。该项规定也是落实《个人信息保护法》对个人信息出境的共性要求。

看点四：补充合规义务。

《办法》对于数据处理者和认证机构的义务规定，在基本保持《征求意见稿》框架的基础上，进行了补充和完善。如：个人信息处理者“不得采取数量拆分等手段”规避个人信息出境保护认证；认证机构备案时需提交“人员安全背景审查材料”等等。

（三）影响和后续关注思考

《办法》进一步理顺了个人信息出境认证的管理体系和管理要求，对网络安全行业而言既有切实影响，也有值得继续关注的后续工作。

从影响方面来看，《办法》对合规义务的完善补充，在强化个人信息保护的同时，无疑将增加相关个人信息处理者、认证机构的合规成本；而有关个人信息保护影响评估的要求，则同样会对个人信息保护行业带来新的市场拉动效应，尤其对个人信息保护供给侧厂商而言。

从后续关注来看，《办法》的全面落地实施，还有某些重要因素需要进一步确认。例如在个人信息保护影响评估实施方面，当前可作为参照的标准有《信息安全技术 个人信息安全影响评估指南 GB/T39335—2020》和《数据安全技术 个人信息跨境处理活动安全认证要求 GB/T46068—2025》，二者对于个人信息保护影响评估的规定相对分散；且对于评估的机构、流程等也没有统一规定。后续是否出台统一标准，值得期待。

1.2.2.4 【数据要素】《国家数据基础设施建设指引》

2025 年 1 月 6 日国家发展改革委、国家数据局、工业和信息化部联合发布《国家数据基础设施建设指引》。《指引》阐述了国家数据基础设施概念内涵、发展愿景、总体功能、总体架构，从数据流通利用、算力底座、网络支撑、安全防护等四个方面指明具体建设方向。其中，数据流通利用设施方面，分别建设数据流通利用设施底座、数据高效供给体系、数据可信流通体系、数据便捷交付体系等。算力底座方面，将重点推进算力资源科学布局；东中西部算力协同；算力与数据、算法创新融合；算力发展与安全保障协同。网络支撑方面，主要围绕建设高速数据传输网和推动传统网络设施优化升级展开。安全防护方面，分别面向国家数据基础设施安全保障和数据流通利用安全保障，提出了方向指引。

绿盟简评：

新年伊始，国家数据局继续加速出台专项政策文件。去年年底以来，国家数据局已先后发布了《关于促进数据产业高质量发展的指导意见》、《关于促进企业数据资

源开发利用的意见》、《可信数据空间发展行动计划（2024—2028 年）》等文件。这些专项政策的密集出台，是国家进一步加速构建数据基础制度体系的明确信号。

整体来看，《建设指引》有四大重要看点。

一是对数据基础设施的建设目标进行了细化。可简称为“五年规划，三步走目标”，即第一步到 2026 年，确定建设的技术路线和实践路径；第二步是到 2028 年，建成覆盖大中城市的数据基础设施；第三步是到 2029 年，建成数据基础设施主体结构、基本格局，及配套服务体系和管理机制。

二是明确了数据流通基础设施的重要地位。一方面，《建设指引》重申了数据基础设施的四种能力，分别对应流通、算力、网络、安全四类基础设施类型。但表述的顺序发生调整，将流通置于首位。也反映了促进数据流通对于整个数据要素工作的重要意义。另一方面，从《建设指引》所提出的技术架构和构成来看，“数据流通利用基础设施”发挥承上启下的中台作用，处于架构的核心；且从该文件附件定义的关键新名词来看，也都与数据流通设施密切关联。

三是首提并强调数据基础设施的主体属性。此前的政策文件对于数据基础设施的提法，基本都是从功能角度界定为算力、网络、流通、安全四类；而此次则是从主体角度，界定了企业、行业、区域、国家数据基础设施四类，并对各类基础设施的类型、建设内容做出了部署。这种界定，也从一个侧面体现出《建设指引》从实施角度更强调可操作性的思路演变。

四是一个重要变化。《建设指引》的一个重要变化就是将此前对数据安全基础设施的具体构成要素进行了调整，将隐私保护计算调整为数据流通基础设施的范畴。这种理论体系上的划分调整，或许对后续建设实施、服务保障、监管职责等都带来某些重要影响。

1.2.2.5 【人工智能】《2025 年人工智能技术赋能网络安全应用测试公告》

5 月 5 日国家互联网应急中心发布。本次测试活动共设置了 7 个测试场景，包括基于智能体的网络安全自动化分析响应、网络安全告警日志降噪、基于互联网流量的漏洞利用攻击识别及 PoC 生成、基于局域网流量的漏洞利用攻击识别、大模型生成内容安全风险检测、重点车辆船舶监控系统资产脆弱性识别、信用卡异常业务行为检测。

绿盟简评：

（一）主要变化

这是该项测试工作开展的第二年，与上一年度有几个显著不同。

一是测试的启动方式不同。2025 年度测试通过公开发布公告方式启动，而 2024 年度的测试工作是通过线下的邀请方式进行。因此，这对于测试工作的知悉度、影响范围、推广效果等方面都可能有较大影响。

二是测试场景的差异。本年度设置了 7 个典型场景，与 2024 年度相比，除了“网络安全告警日志降噪”之外，其余 6 个场景均不相同。比较来看，本年度测试的通用场景更加侧重于动态网络安全防护，行业场景则在金融领域场景的基础上增加了交通领域场景。

三是组织方式更加完善。2025 年度测试工作在组织体系、测试要求和标准、测试流程等方面的规定更加细致、明确，有助于该项测试工作的常态化推进。

此外，公告还对“测试结果应用”进行了说明，从鼓励推广应用的角度明确了测试结果对相关科技奖励、人才评选等的参考价值。

（二）影响思考

近年来，强化网络安全技术产品遴选和推广应用，是有关主管部门一直在大力推进的一项重点工作，试点遴选、典型案例遴选等是此类工作常见的载体方式。与其相比，公开测试的方式更加侧重技术产品的实际效果，其评价结果的直观性、可量化性、实效性等方面往往具有更强的说服力。

随着该项测试工作规范化、常态化趋势的日益明确，其对于我国网络安全监管体系、行业发展都将产生一定影响。从监管体系看，伴随测试工作相关的协调推进，我国网络安全监管体系在相关工作中的部门协同、职能优化、分工合作等或将更加完善。从行业影响来看，因测试结果具有的潜在应用价值，该项测试无疑将成为人工智能大模型供应商打造品牌影响力的“竞技场”，也将可能成为人工智能大模型用户备货的“购物车”。

1.3 美国 2025 年网络安全法规政策分析

2025 年，我们持续关注美国网络安全政策法规发展动态，并从网络公开信息中重点梳理了年内发布的 107 项网络安全法规政策。

1.3.1 年度综述

2025 年，我们持续关注美国网络安全政策法规发展动态。2025 年美国发布了 107 项网络安全法规政策，从内容领域分布来看，数量占比靠前的包括：网络安全（33.6%）、

技术产品管理 (28.9%)、人工智能 (28%)、数据安全 (5.6%) 和个人信息保护 (3.7%) 等领域。

在这些内容板块中，各自包含了许多不同的细分领域，一定程度上可反映本年度内，其安全关注的不同侧重。

1. 网络安全。侧重点包括网络安全风险管理、网络资产管理、预算和重点领域保障、供应链安全管控、密码和加密算法、网络安全事件响应等。涉及的行业主要包括科技、国土安全、贸易、国防、应急等。

2. 技术产品管理。侧重点包括漏洞管理、开发安全、软件供应链安全、差分隐私、零信任、存储介质安全、威胁检测等。涉及的行业主要包括电信和通信、能源、卫生、国土安全、科技等。

3. 人工智能。侧重点包括人工智能发展规划、管理机制、人工智能基础设施、人工智能供应链和产业生态、人工智能产品风险、人工智能责任机制等。涉及的行业主要包括国土安全、贸易管理、教育、科技、知识产权等。

4. 数据安全。侧重点包括数据共享、数据流通和交易合规、数据语料安全、数字化战略发展等。涉及的行业主要包括卫生、司法、金融、国土安全等。

5. 个人信息保护。侧重点包括儿童在线隐私保护、个人隐私标准框架、个人信息删除权利保护、数据经纪人合规等。涉及的主要行业包括金融、互联网平台、贸易等。

此外，网络安全国际合作、数字化应用发展等也是美国 2025 年度网络安全法规政策关注较多的领域。

1.3.2 热点政策法规分析解读

1.3.2.1 特朗普签署《关于消除美国在人工智能领域领导地位的障碍的行政命令》

《行政令》撤销了拜登政府时期“有害的人工智能”政策，为巩固美国在人工智能领域的全球领导地位扫清了障碍。主要内容包括：一是阐明了美国人工智能政策目标。它旨在巩固美国在全球人工智能领域的领导地位。二是明确要制定一项人工智能行动计划。该行政令要求总统科技助理、人工智能与加密技术特别顾问、总统国家安全事务助理等机构负责人制定一项人工智能行动计划。三是关于行政令撤销的实施计划。要求各部门和机构审查并修订或废除与拜登政府人工智能政策相关的所有行动，以确保其与特朗普政府的政策目标一致。

绿盟简评：

长期以来，美国一直处于人工智能技术发展的前沿，以 ChatGPT 为代表的人工智能大模型引领了全球人工智能发展的新浪潮。近几年，随着全球范围内人工智能技术竞争日益激烈。在此背景下，特朗普重返白宫伊始就将人工智能作为其施政的重点方向之一，采取多项举措力图保持美国在人工智能全球竞争中的领导地位。

首先，撤销“有害的人工智能”政策。美国总统特朗普 1 月 20 日上任首日即签署了一项行政令，撤销了拜登政府时期的 78 项政策，其中包括拜登政府于 2023 年 10 月 30 日发布的《关于安全、可靠和可信的人工智能开发与使用的行政命令》（以下简称“第 14110 号”行政令）。“第 14110 号”行政令要求人工智能开发者在发布可能对国家安全、经济或公共卫生构成风险的人工智能系统前，必须向政府共享安全测试结果，并制定相关标准防范风险。此外，该行政令还包括保护用户隐私、对高风险人工智能

进行风险评估并设立防范措施等内容。特朗普政府认为“第 14110 号”行政令阻碍了美国人工智能创新发展。

其次，宣布实施“星际之门”（Stargate）计划。1 月 21 日，美国总统特朗普又宣布了一项名为“星际之门”的投资计划，总投资将高达 5000 亿美元，旨在建设美国新一代人工智能基础设施。这充分体现了特朗普政府希望通过大规模投资和技术创新巩固并保持美国人工智能全球领导地位的战略意图。

第三是发布新的人工智能政策框架，即本行政令。该行政令可视为特朗普政府全面推进美国人工智能发展的总纲领。从内容来看，该行政令对拜登政府时期的人工智能政策进行了四方面的重大调整和修正。

一是政策思路的转变：安全优先转为创新主导。拜登政府强调人工智能的安全、伦理和社会风险，而特朗普行政令明确提出“开发消除意识形态偏见或人为社会议程的 AI 系统”，主张通过减少监管和推动创新来巩固其全球领导地位。

二是撤销拜登政府人工智能安全监管框架。直接废除拜登政府第 14110 号行政令，该行政令是美国首个针对人工智能监管的综合性框架，旨在加强联邦政府对人工智能的安全监管，减少人工智能对消费者、工人和国家安全构成的安全风险。

三是调整既有人工智能政策与行动。审查基于拜登政府第 14110 号行政命令所制定的所有政策、指令、法规、命令等，并对不符合特朗普政策目标的行政措施进行暂停、修订或撤销。修订拜登政府通过管理与预算办公室（OMB）发布的备忘录《推进联邦机构使用人工智能的治理、创新和风险管理》（M-24-10）和《对负责任地采购 AI 的指导意见》（M-24-18），以确保其与新政策目标一致。

四是弱化人工智能方面的国际合作。拜登政府主张促进负责任的创新、竞争和合作，积极吸引世界各地的人工智能人才。而特朗普更倾向于通过单边行动来推动人工智能创新发展。

1.3.2.2 《特朗普总统 2026 财年可自由支配预算请求》

该预算案将非国防可自由支配资金削减 1630 亿美元，比 2025 年制定的水平下降约 23%，这是自 2017 年以来最低的非国防开支水平。同时，该预算案提议国防开支将增加 13%，国土安全部的拨款将增加近 65%，网络安全领域拟削减预算 4.91 亿元。

绿盟简评：

该预算案对网络安全领域的预算进行了明显调整，主要体现在对美国网络安全和基础设施安全局（CISA）的预算削减，总金额由 2025 财年的 30 亿美元削减 4.91 亿美元，降至约 25.09 亿美元，降幅约为 16%，这是近年来最大幅度的缩减。

特朗普政府 2026 财年网络安全预算调整反映了其“聚焦核心、削减冗余”财政优先级，主要体现在：使 CISA 重新关注其核心职责——联邦网络防御和加强关键基础设施的安全；取消被视为政府“武器化”和浪费性支出的项目；裁撤国际事务办公室；取消与州政府重复的网络安全项目等。与 2025 财年相比，CISA 预算大幅度缩减，而国防和国土安全开支则显著增加，显示出其战略重心向传统安全领域的倾斜。该预算调整或将引发关于国家安全漏洞和效率的争议。

目前，该提案已提交国会审议，最终落地情况可能因两党博弈而有所调整。

1.3.2.3 美国 CISA 推出促进软件供应链安全采购的工具

8 月 26 日美国网络安全和基础设施安全局（CISA）发布。该 Web 工具依据《政府企业消费者软件采购指南：网络供应链风险管理（C-SCRM）生命周期中的软件保障》发布，提供数字化的手段，简化用户评估软件保障和供应商风险的方式，使用户能够做出符合联邦网络安全指南和最佳实践的决策。

绿盟简评：

（一）发布背景

近年来，软件供应链安全已上升为事关国家安全与企业有序运行的重要议题。美国政府将软件供应链安全视为国家网络安全的核心支柱之一，并逐步强化软件供应链安全制度体系。CISA 此次发布的供应链网络工具，正是 2024 年 8 月政策链条的呼应，它将抽象的安全要求转化为可操作、可评估、可嵌入采购流程的数字化工具。

（二）重点内容梳理

这一交互式网络工具的核心价值，在于其将复杂的网络安全要求转化为用户友好型的决策流程。它不再是一份静态的文档，而是一个能够根据用户具体场景动态生成评估路径的“数字顾问”。工具通过自适应问答机制，将庞大的指南内容分解为与用户最相关的部分，帮助采购人员聚焦关键问题，例如供应商的安全开发实践、漏洞管理机制、合规证明真实性等。

工具支持生成可导出的评估摘要，这一功能将极大便利跨部门协作。采购人员可以利用其来描述、评估和衡量供应商与软件生命周期相关的安全实践，也可直接使用

系统生成的标准化报告，从而在组织内部形成基于统一语言和框架的风险讨论，减少因专业壁垒导致的理解偏差。

该工具可帮助不同规模的组织将网络安全整合到其采购流程中，降低使用门槛，其随附的电子表格已经覆盖了 10000 多名用户，下载量超过 4000 次，充分反映了联邦、州和地方政府以及中小型企业的强劲需求。

（三）影响分析

这一工具的直接作用是大幅降低供应链安全合规的成本。此前因供应链安全评估流程冗长，常导致采购周期长达数月甚至数年；而工具的标准化与自动化特性，既能减少重复性人工评估，又能通过数据驱动决策缩短采购周期。

从外部影响来看，随着该工具及相关标准应用范围的扩大，其对整个软件供应链或将产生长远的影响，波及供应链安全生态中的各个环节。同时，在法规政策层面的实践，也将产生一定的示范参考价值。

1.3.2.4 美国发布《CISA 战略重点：面向网络安全未来的 CVE 质量提升路线图》

网络安全和基础设施安全局（CISA）发布《CISA 战略重点：面向网络安全未来的 CVE 质量提升路线图》(以下简称《路线图》)。该路线图明确了 CVE 计划在未来发展中的优先事项，旨在进一步提升该计划对全球网络安全社群需求的支持能力。

绿盟简评：

（一）发布背景

美国将漏洞管理作为网络安全的重点工作之一，此前发布的一系列漏洞相关政策法规，系统性地构建了其国家层面的漏洞管理能力与责任制度框架。进入 2025 年以来，也在持续完善漏洞相关的法规政策体系。

1. 2025 年 3 月，美众议院通过《2025 年联邦承包商网络安全漏洞消减法案》。旨在通过管理和预算办公室（OMB）及国防部（DoD）的行动，防止因联邦承包商带来的网络安全漏洞。

2. 2025 年 5 月，美国网络安全与基础设施安全局、美国国家标准与技术研究院发布了《可能被利用的漏洞：漏洞利用可能性的拟议指标》。该文件提出了一种名为 LEV（Likelihood of Exploited Vulnerabilities）的新型安全指标，旨在通过数学模型量化漏洞被利用的累积概率，以优化漏洞修复优先级。

（二）重点内容梳理

1、战略转变

该路线图标志着美国在漏洞治理领域从以规模增长为主的阶段正式转向以质量提升为核心的新阶段。这一战略重点将有助于增强信任、提高响应能力并提升漏洞数据的质量。路线图及其优先事项的制定，将继续基于 CISA 从广泛的国内外合作伙伴处收集的反馈意见，并结合该机构多年来对 CVE 项目的支持经验。

2、核心原则

CISA 认为，CVE 计划必须致力于实现无冲突且供应商中立的管理机制，推动广泛的多方参与，确保流程透明，并落实负责任的项目领导。CISA 承诺将维护 CVE 计划的核心原则：CVE 数据必须保持免费、公开访问，作为一项公共产品服务于网络防御协调，促进安全工具创新，并为全球行业与政府的安全防御者提供支持。

3、重点方向

CISA 对 CVE 计划未来的愿景包括以下 5 个重点方向：

一是拓展社区合作伙伴关系：CISA 将借助合作伙伴网络，增强国际组织、政府机构、学术界、漏洞工具提供商、数据消费者、安全研究人员、运营技术及开源社区等多方代表的参与度。

二是政府持续支持：作为关键公共产品，CVE 计划的基础设施与核心服务需要 CISA 持续投入资源。CISA 正在评估多元化资金机制的可能性，以回应社区的相关建议。

三是推进现代化进程：CISA 将加快技术改进措施的实施步伐。

四是提升透明度与沟通效率：CISA 将积极吸纳社群反馈，将其纳入路线图决策，并与全球伙伴保持定期沟通与互动。

五是提高数据质量：CISA 将与行业及国际政府合作建立新的标准化机制，包括扩展漏洞数据丰富化（如漏洞信息增强）的联合机制，并扩展授权数据发布者（ADP）职能。

（三）影响分析

《路线图》的发布，将进一步明确美国漏洞管理政策目标走向，对其他国家和地区或将产生一定示范引领。CVE 计划被美国视为“关键性的全球网络防御框架和全球网络防御基石”，已成为全球广泛应用的网络安全公共产品之一。过去一段时期是 CVE 计划的快速发展阶段，其重要成果之一是成功组建了由 460 多个 CVE 编号机构（CNA）构成的广泛全球网络。

1.3.2.5 美国发布《DeepSeek 人工智能大模型评估报告》

该评估报告针对中国领先 AI 开发商 DeepSeek 进行评估并发布评估结果。按照报告结论，DeepSeek 人工智能大模型在性能、成本及安全性等核心维度上均大幅落后于美国同类产品，但可能对美国国家安全构成潜在风险。

绿盟简评：

发布背景

该项评估是为完成《赢得人工智能竞赛：美国人工智能行动计划》中提出的开展相关人工智能态势评估要求而展开。评估目标是对美国内外人工智能系统的能力、外国人工智能系统的采用情况以及国际人工智能竞争态势进行全面评估。

从执行此次评估的主体来看。具体评估工作由商务部委托美国国家标准与技术研究院（NIST）下属的人工智能标准与创新中心（CAISI）承担。该中心被定位为美国政府与产业在 AI 领域的核心联络点，主要负责通过制定标准、推动测试与合作研究等。

从评估的方法流程来看。CAISI 组建专门专家团队，选取了 DeepSeek 的三个核心模型（R1、R1-0528、V3.1）与美国四个前沿模型（包括 OpenAI 的 GPT-5 系列和 Anthropic 的 Opus 4）进行对比测试评估。测试涵盖了 19 个基准领域，不仅包括公开的测试，也包含 CAISI 团队与高校及联邦机构合作开发的内部测试。

评估报告主要结论

CAISI 通过一系列测试，对 DeepSeek 模型给出了 6 项评估结论：

1. 性能落后：报告指出，最佳美国模型在几乎所有基准测试中都优于最佳 DeepSeek 模型（V3.1）。在软件工程和网络任务中，差距尤为显著，美国模型多解决了 20%以上的任务。

2. 成本更高：在达到相似性能水平时，DeepSeek 模型的平均使用成本比美国参考模型高出 35%。

3. 代理劫持风险高：基于 DeepSeek 最安全模型（R1-0528）的智能代理，被恶意指令劫持的可能性是评估的美国前沿模型的 12 倍。在模拟环境中，这些被劫持的代理能够执行发送钓鱼邮件、下载运行恶意软件、窃取用户凭证等危险操作。

4. 易被“越狱”：在使用常见越狱技术时，DeepSeek 最安全模型对 94%的明显恶意请求做出了响应，而美国参考模型的这一比例仅为 8%，显示出其安全防护机制的脆弱性。

5. 输出内容存在政治倾向：报告指出 DeepSeek 模型在输出中呼应“不准确和误导性的政治叙事”，其数量是美国参考模型的 4 倍。

6. 全球采用率激增：尽管存在上述缺陷，报告承认自 DeepSeek R1 发布以来，中国模型的全球采用率急剧上升，在模型共享平台上的下载量增长了近 1000%。这种快速增长本身，被报告视为一种需要警惕的风险扩散。

表 1.1 评估结果一览表[①]

Domain	Evaluation	Model					
		OpenAI GPT-5	Anthropic Opus 4	OpenAI gpt-oss	DeepSeek V3.1	DeepSeek R1-0528	DeepSeek R1
Cyber	CVE-Bench	65.6	**66.7**	42.2	36.7	36.0	26.7
	Cybench	**73.5**	46.9	49.5	40	35.5	16.7
	CTF-Archive	**50.6**	34.1	34.3	28.2	26.5	8.5
Software Engineering	SWE-bench Verified	63.0	**66.7**	42.6	54.8	44.6	25.4
	Breakpoint	**98.0**	92.3	93.0	78.5	60.2	16.0
Science and Knowledge	MMLU-Pro	89.8	**90.2**	85.5	89.0	89.0	87.5
	MMMLU	**87.7**	83.8	77.7	82.2	81.9	82.7
	GPQA	**86.9**	78.8	71.2	79.3	81.3	72.6
	HealthBench	**63.0**	41.7	61.7	52.5	55.7	50.5
	HLE	**26.6**	11.6	11.3	13.0	13.6	9.0
Mathematical Reasoning	SMT 2025	**91.8**	82.2	82.3	86.2	87.6	75.0
	OTIS-AIME 2025	**91.9**	66.7	72.9	77.6	73.3	58.3
	PUMaC 2024	**85.9**	69.1	67.3	77.7	72.7	60.9

影响分析

评估报告的发布，或将产生两方面的影响。

一方面，从美国国内来看，试图通过报告增强其国内人工智能产业发展信心，为特朗普政府的各项人工智能政策的推行营造气氛。

另一方面，从国际影响来看，该报告或将对人工智能产业生态产生一定影响，波及人工智能应用程序开发者和各类用户，对 Deep seek 模型产品的上下游供应链、生态维护、市场等或造成不利影响。

总体来看，美国商务部的这份评估报告，将技术测评与国家战略深度捆绑，反映了当前各国围绕人工智能的竞争已被纳入国家战略层面。

[①] 美国《DeepSeek 人工智能大模型评估报告》

02

安全态势篇

2.1 APT 攻击态势

2.1.1 概述

2025 年，全球网络安全形势持续恶化，0Day 漏洞在高级持续性威胁（APT）攻击活动中的使用频率显著上升，成为推动威胁加速升级的关键因素之一。大量 0Day 漏洞被应用在操作系统、浏览器、网络设备、安全软件等核心基础设施中，使得攻击者能够在较长时间内绕过防护体系，对广泛分布的目标形成深远且长期的影响。同时，多个新旧活跃的 APT 组织在全球范围内发动密集攻击，攻击动机从间谍窃密延伸至经济获利、供应链渗透与政治破坏，推动威胁格局进入更加复杂的阶段。

拉美地区也在近期遭遇高强度的 0Day 攻击。南美 APT 组织“盲眼鹰”(BlindEagle) 于 2024 年 11 月至 2025 年 2 月间多次发动针对哥伦比亚司法及政府机构的攻击行动，使用带有 CVE-2024-43451 变种漏洞的攻击载荷。Windows 系统在处理含恶意 SMB 链接的快捷方式文件时存在漏洞，只要用户执行右键、拖动、删除等交互操作，即会使攻击者能够捕获 NTLMv2 哈希并进而实施哈希传递攻击或身份冒用。这种低交互触发、零点击性质的漏洞再次验证了系统逻辑类 0Day 的高危险性。

来自朝鲜的 Lazarus 组织在 2024 年底至 2025 年初开展“SyncHole”行动，瞄准韩国多个行业。该组织掌握了韩国广泛使用的网银安全软件 CrossEX 的一个 0Day 漏洞，依托该软件的庞大装机量实现了大规模入侵和窃密。尽管行动结束后 Lazarus 迅速撤下漏洞利用代码，使得该 0Day 的具体技术细节至今未知，但这类借助高装机量安全产品进行攻击的模式，对国家级关键信息系统构成严重威胁。

2024 年 4 月，疑似来自土耳其的 APT 组织 Marbled Dust（Sea Turtle）对伊拉克库尔德军队发动攻击，相关情报直到一年后的 2025 年 4 月才被公开。该组织利用

目录穿越 0Day 漏洞 CVE-2025-27920，通过窃取凭证登录 Output Messenger Server Manager 后，将恶意脚本写入服务器自启动目录并执行，继而运行 Golang 后门控制受害主机。这一案例体现了凭证滥用与 0Day 漏洞结合的危险性，使攻击者能够实现高度隐蔽的持久化控制。

2025 年 3 月，APT 组织 StealthFalcon 通过钓鱼邮件攻击土耳其大型国防企业。其诱导受害者执行带有 0Day 漏洞 CVE-2025-33053 的网络快捷方式文件，从而将特种后门植入目标系统。后门载荷进一步用于加载其他攻击组件，对受害者设备执行情报收集与远程操控，显示该组织在针对高价值个人与企业的定向攻击上持续投入。

思科则在 2025 年 9 月披露了与上一年 ArcaneDoor 间谍活动相关的新一轮全球性攻击行动。某国家级威胁组织利用 Cisco 设备的 3 个 0Day 漏洞（CVE-2025-20333、CVE-2025-20362、CVE-2025-20363）对全球关键基础设施实施持续攻击，其中两个漏洞为 RCE，CVSS 分别达到 9.9 与 9.0。此次攻击影响多个美国联邦机构，表明网络基础设施领域的 0Day 漏洞已成为国际冲突与网络情报战中的关键资源。

在经济驱动型黑客生态中，0Day 漏洞利用链也呈现快速武器化趋势。2025 年 7 月，互联网上出现针对微软 SharePoint 的漏洞利用链，实现远程代码执行并迅速被 mimo、Warlock 等黑客组织用于大规模攻击。7 月下旬，该利用链还被整合了两个新增漏洞，形成对微软最新补丁的绕过能力，成为此阶段危害最大的企业级漏洞利用工具之一。

作为以窃取信用卡和客户数据获利为主的组织，活跃近十余年的 APT 组织 XE 在 2024 年第四季度发起了针对仓储管理软件 VeraCore 的供应链攻击行动。该组织利用了两个 0Day 漏洞，对公网部署的 VeraCore 系统进行入侵，再借助已被攻陷的服务器对下游客户实施投毒攻击。XE 通过这两项漏洞在供应链层面建立持久性控制点，使相关企业长期暴露在数据泄露与被进一步渗透的风险之下。

移动终端领域同样暴露在高危 0Day 之下。2025 年 8 月，苹果公司发布安全公告，修复 ImageIO 模块中高危越界写入漏洞 CVE-2025-43300（CVSS 评分 8.8）。该漏洞可能允许攻击者在无需交互的情况下进行远程代码执行，并已被用于攻击特定目标用户。尽管苹果未披露攻击细节，但此类图像处理类 0Day 历来是移动平台攻击链的关键节点，对高价值用户构成重大风险。

2.1.2 热点安全事件

2.1.2.1 新型 APT 链锯鲨针对我国的攻击行动

2025 年，绿盟科技伏影实验室披露了一个针对我国科研领域的新型 APT 组织“链锯鲨”（ChainedShark，追踪编号：Actor240820）。该组织自 2024 年 5 月起持续活跃，其攻击行动具有高度的战略一致性与技术复杂性，攻击目标集中于我国高校及研究机构中从事国际关系、海洋技术等领域的专业人员，意图窃取我国在外交、海洋技术等领域的敏感数据与情报。

链锯鲨组织表现出明确的地缘政治驱动特征，其攻击目标高度聚焦于我国高校及科研机构中的国际关系、海洋科学等方向的专家学者。该组织具备极强的社会工程学能力，能够撰写流畅、自然的高水平中文诱饵，并熟练运用会议邀请函、学术征稿函等专业场景构建欺骗性攻击入口，有效降低了目标的警惕性。

在技术能力上，链锯鲨展现出国家级攻击团队的专业水准。其武器库整合了 N day 漏洞利用与高复杂性特种木马程序，攻击链路设计精巧，载荷具备高度的对抗性与隐蔽性，表明其拥有成熟的攻击基础设施和持续的武器开发能力。

该组织的攻击活动在保持战略目标一致的前提下，其战术与技术在过程中呈现出明显的演进轨迹。

首轮攻击（2024 年 5 月）：此轮攻击是目前已发现活动中最为复杂的一次。攻击链中使用了其自研的特种木马 LinkedShell，该木马具备高度的定制化与对抗能力，技术细节复杂，体现了组织强大的初始武器化能力。

后续攻击（2024 年 8 月至 11 月）：在后续行动中，攻击者战术发生调整。由于成功利用了 2024 年 6 月才公开的 GrimResource 漏洞，攻击流程的复杂度显著降低，反映出组织在积极融入公共漏洞以提升攻击效率与成本效益。

通过多维度的线索关联，将不同时间段的攻击事件串联起来，形成了完整的攻击者画像。

- **目标一致性**：2024 年 5 月与 11 月的攻击事件中，出现了同一目标人员遭受两次不同攻击的情况，强有力地证明了攻击行动的定向性与持续性。
- **诱饵同源性**：尽管攻击载荷存在差异，但不同攻击活动中使用的钓鱼邮件在主题选择、行文措辞与社交话术上高度相似，构成了行为层面的“指纹”特征。

上述关联性分析不仅为归因提供了关键依据，也揭示了链锯鲨组织在长期行动中，始终遵循着一套成熟的社会工程学剧本与攻击管理流程。

图 2.1 链锯鲨组织典型攻击行动

链锯鲨攻击事件是我国网络安全领域面临的新型 APT 威胁的典型缩影。它清晰地表明，具备明确地缘政治目的、拥有高水平技术能力、并能够持续演进攻击战术的专业化组织，正对我国关键领域的科研与数据安全构成现实且严峻的威胁。面对此类攻击，防御方必须构建起覆盖威胁情报、行为检测与纵深防御的完整体系，方能有效应对。

2.1.2.2 Lazarus 组织加密货币窃取行动

2025 年 2 月，朝鲜背景的 APT 组织 Lazarus 对知名加密货币交易所 Bybit 发起了一次高度复杂的供应链攻击，成功窃取超过 40 万枚 ETH 与 stETH，总价值高达约 15 亿美元，成为迄今为止全球加密货币领域涉案金额最大的单次安全事件。此次攻击暴露出即使在多重签名冷钱包这一被视为行业黄金标准的安全方案中，软件供应链与人为操作环节仍存在致命弱点。

Lazarus 组织成功入侵了 Bybit 所使用的 Safe{Wallet}智能合约钱包的开发环境。攻击者首先通过社会工程学手段控制了一名开发人员的设备，进而获取了对该公司网络及代码发布系统的访问权限。在 2 月 19 日的例行更新中，攻击者将恶意 JavaScript 文件植入并通过官方域名 app.safe.global 进行分发。该恶意脚本篡改了钱包的用户界面与交易构造逻辑——当 Bybit 操作人员执行从冷钱包向热钱包转账的标准流程时，界面虽显示为正常操作，背后却将收款地址替换为攻击者控制的地址，并诱使操作人员在不知情的情况下对恶意交易完成签名授权。

值得注意的是，此次攻击反映出国家级 APT 组织在战术上的显著升级：其攻击目标从交易所自身系统转向其依赖的底层基础设施，攻击手法从暴力破解转向对“人机信任关系”的精准利用。尽管 Bybit 在流程中设置了多重签名机制，但恶意代码在界面层掩盖了交易的真实内容，导致审核人员在硬件钱包上进行的最终确认行为实质上变成

了对攻击交易的“盲签”。这一事件为整个数字资产行业敲响警钟：在追求技术方案完备性的同时，必须建立覆盖软件开发、分发、操作执行全链路的“零信任”防护体系，并为核心操作建立独立于界面显示的交叉验证机制，方能应对日趋隐蔽的供应链与交互层攻击。

2.1.2.3 利用 Chrome 浏览器 0Day 漏洞的攻击行动

2025 年 3 月，网络安全研究机构披露了一起被命名为“Operation ForumTroll”的高度复杂的定向攻击行动。该行动由一个未知的国家级 APT 组织策划，其核心攻击武器是谷歌 Chrome 浏览器的一个 0Day 漏洞（CVE-2025-2783），此漏洞可以实现沙箱逃逸，从而在受害者的 Windows 操作系统上直接执行任意代码，完全控制目标计算机。

在攻击链的构建上，Operation ForumTroll 展现出极高的专业水准。整个攻击始于一次精准的鱼叉式网络钓鱼活动。攻击者深入研究了目标群体的背景，精心伪造了俄罗斯知名学术论坛“Primakov Readings”的官方会议邀请函，并通过邮件分发给特定的科学家与学者。邮件中包含的链接会将受害者引导至与真实论坛几乎一模一样的克隆网站。当目标用户访问该网站时，隐藏在页面中的漏洞利用代码会被触发。值得注意的是，攻击者采用了短寿命域名技术来隐藏其真实的命令与控制（C&C）服务器，并且在攻击得手后，恶意链接会自动将用户重定向至真实的论坛官网，以此掩盖攻击痕迹，延长其攻击基础设施的存活时间。

技术深度分析揭示，此次攻击绝非孤立事件。攻击载荷具备高度的模块化与对抗性。初始漏洞利用仅负责逃逸浏览器沙箱并获取系统权限，随后会从云端下载第二阶段的加载器。该加载器采用了一系列复杂的反分析技术，包括虚拟机检测、沙箱环境感知以及安全软件进程排查，以确保其仅在真实的目标环境中激活。最终部署的是一

款名为“Dante”的定制化间谍木马，该木马集成了键盘记录、屏幕截图、文件窃取和远程命令执行等多种监控与窃密功能，其通信流量均经过高强度加密，并伪装成正常的HTTPS 流量，以规避网络层面的检测。

Operation ForumTroll 事件为全球网络安全界敲响了新的警钟。它表明，国家级APT 组织正持续将攻击矛头指向基础软件生态，特别是像 Chrome 浏览器这样被广泛使用的应用程序。攻击者不仅具备挖掘和利用底层 0Day 漏洞的能力，更在社工欺诈、攻击链伪装和载荷持久化方面展现出娴熟的技巧。对于依赖此类软件的关键基础设施与敏感行业而言，必须从“假定已失陷”的零信任理念出发，构建覆盖终端、网络和云的多层次纵深防御体系，并加强对高级威胁行为的狩猎能力，方能有效应对此类隐蔽而强大的定向攻击。

2.1.3 态势观察

随着数字化进程的加速推进，2025 年的网络安全威胁态势呈现出前所未有的复杂性和多样性。各类攻击组织在技术手段、攻击方法和对抗能力等方面持续演进，展现出高度的专业化、产业化和隐匿化特征。

高级持久性威胁（APT）载荷形态的形态变异与无文件化趋势愈发明显。在 APT 攻击领域，攻击者不断突破传统技术框架，开发出更具隐蔽性和对抗性的新型攻击手段。链锯鲨组织采用的可执行文件重构技术就是其中的典型代表。该技术打破了传统的文件格式转换逻辑，通过将 PE 文件拆解重组为 shellcode 的方式，实现了攻击载荷的深度隐匿。与简单的格式转换不同，这种技术先将 PE 文件的代码、数据、导入表等核心组件进行提取和融合处理，再嵌入专门的初始化代码，最终生成结构特殊的shellcode。这种重构后的 shellcode 内部组件以碎片化形式存在，尺寸接近 3MB，远超常规 shellcode，显示出攻击者已经实现了该技术的工具化、自动化。

与此同时，Lazarus 组织在区块链领域展现出的多重签名劫持技术，同样代表了 APT 攻击技术的新高度。这种基于供应链污染的特殊攻击方式，通过恶意智能合约和劫持委托调用等复杂机制，成功欺骗了 ByBit 的高层管理人员，获得了多重签名授权。攻击者首先通过钓鱼攻击获取了对 Safe{Wallet}线上代码的操作权限，随后修改 JavaScript 代码，实现了对多重签名流程的劫持。这种攻击手法的精妙之处在于，它在不破坏原有安全机制的前提下，通过替换关键环节的执行逻辑，实现了对资金转移的完全控制。

漏洞利用更加精细化，沙箱逃逸能力持续提升。在漏洞利用方面，Operation ForumTroll 攻击行动中使用的 Chrome 沙箱逃逸 0Day 漏洞 CVE-2025-2783，展现了攻击者对底层系统机制的深刻理解。这个极高威胁的浏览器漏洞利用了 Chromium 进程间通信机制中的设计缺陷，通过违规获取浏览器进程中的线程句柄，成功实现了沙箱逃逸。具体而言，攻击者发现 Chromium 浏览器进程在检测伪句柄时，只检查进程伪句柄（-1）而忽略了线程伪句柄（-2）的验证，从而利用这个疏漏将恶意代码注入到浏览器进程中。

这种精细化的漏洞利用技术表明，现代网络攻击者已经不再满足于简单的内存破坏漏洞，而是转向对系统组件间交互逻辑的深入分析和利用。攻击者能够准确识别出复杂软件系统中那些看似微小但具有关键意义的设计缺陷，并将其转化为有效的攻击武器。这种趋势对软件安全和漏洞防护提出了新的挑战。

面对这些发展趋势，传统的安全防御体系已经显得力不从心。未来的网络安全防护需要建立更加智能、主动的防御机制，结合威胁情报、行为分析、人工智能等技术，构建覆盖攻击链各环节的纵深防御体系。同时，还需要加强跨组织、跨行业的安全协作，建立更加有效的威胁信息共享和协同响应机制，共同应对日益复杂和专业的网络安全威胁。

2.2 高风险主机态势

2025 年，高风险主机持续关注云服务架构资产和远程办公资产变化情况。对比去年来看，全国高风险主机的总量有 6.53%的增幅，地理分布格局基本稳定。

从端口开放情况来看，21、22、2375 等端口开放数量对比去年持续增长，而 3389 端口开放数量有所减少。从服务类型、运营商分布和应用场景分布等综合分析，云服务端口及自动化运维端口暴露情况呈逐年增长趋势，远程协议端口暴露情况有所收敛。

这一现状与近年来的发展趋势直接相关：**一是业务架构和业务模式的持续迭代。**企业上云进程由〝规模扩张〞逐步转向〝深度使用〞，云主机、容器集群及云原生应用规模持续扩大，资产部署形态由〝集中式机房〞向〝多云、混合云与边缘云〞加速演进。DevOps、自动化运维及智能体调度逐渐成为常态运行方式，在提升业务效率的同时，也进一步放大了云平台控制面和运维接口的安全风险。因此，以 22、2375 等为代表的云运维与控制类端口长期处于高位运行状态，并呈现持续增长态势。**二是国产化落地成效逐步显现。**随着国产化进程在关键行业和政务系统中进一步推进，传统 Windows 远程桌面使用场景逐步收缩，零信任、堡垒机及云桌面等新型远程访问方式不断普及，公网直连远程访问需求明显下降，远程协议端口暴露情况得到一定程度的收敛。因此，需在持续推进端口治理工作的基础上，进一步加强云原生架构和国产化应用模式下的资产识别能力，持续收敛暴露面，并提升精细化安全管控水平，以有效应对技术架构和业务模式革新所带来的安全挑战。

2.2.1 高风险端口开放情况

据绿盟网络空间测绘平台统计，全国开放高风险端口的资产数量共计 10,030,303 个。其中，22、21、3389、2375 端口服务占比分别为 52.60%、30.34%、13.16%和

3.91%。从数量分布来看，高风险端口资产总量比去年多 6.53%。如图 2.2 所示，各端口的数量对比去年增长趋势变化不同。其中，21、22 和 2375 端口对应的资产数量呈现持续上升趋势，而 3389 端口对应的资产数量有所减少。由此可见，随着云原生架构的深度应用，文件传输、运维管理及云平台控制等相关资产的暴露面持续扩大，相应的安全风险逐步上升，数据泄漏、接口权限滥用及云平台控制面安全问题日益凸显。同时，随着国产化进程的不断推进，系统架构和技术栈发生调整，相关应用在迁移、适配和运维过程中所面临的安全风险也需给予重点关注。

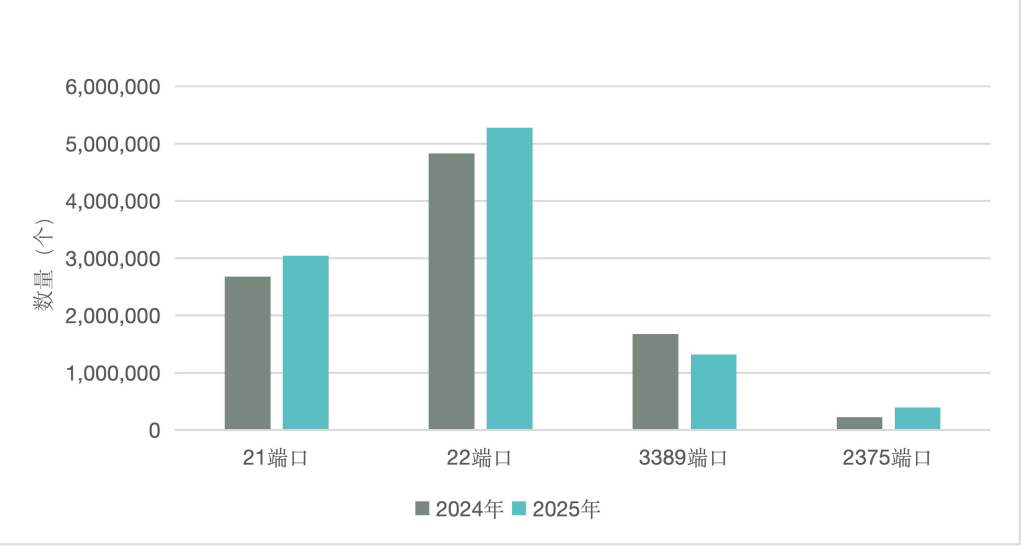
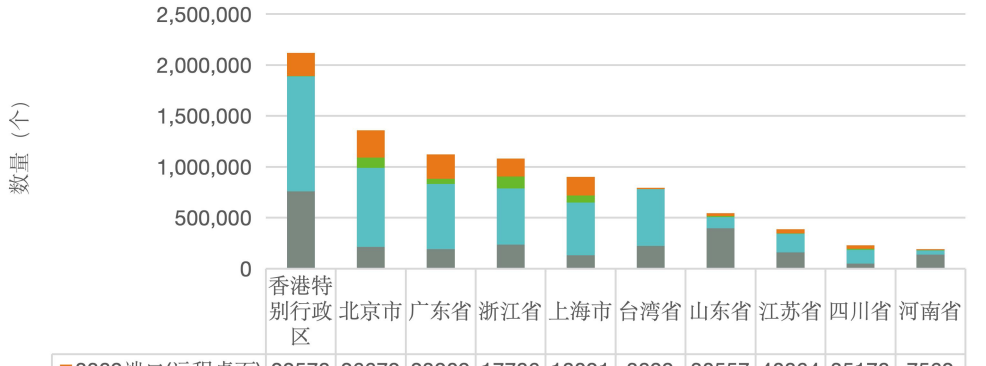


图 2.2 高风险端口分布

从地理分布来看，暴露在互联网上的高风险资产数量以香港特别行政区为最，其高风险端口服务数量达 2,118,715 个；其次是北京市和广东省，数量分别为 1,359,195 个和 1,123,400 个（如图 2.3 所示）。

一个有意思的发现是，高风险资产数量排名前五的地区，恰好也是网络资产总量排名前五的地区。从整体上看，一个地区的网络资产规模与其高风险资产数量大致呈

正比关系。然而，具体分析排名前五的地区可以发现：香港特别行政区的网络资产总量仅排名第五，但其高风险资产数量却高居第一；与之形成对比的是，大陆省份如北京、广东、浙江、上海，其网络资产总量位居前四，但高风险资产数量相对更低。这一差异在一定程度上反映出，大陆省份在网络安全治理的平均效果上表现更好。



	香港特别行政区	北京市	广东省	浙江省	上海市	台湾省	山东省	江苏省	四川省	河南省
3389端口(远程桌面)	22578	26672	23909	17786	18021	3899	30557	40864	35178	7562
2375端口(Docker)	4209	10289	50476	11581	71120	471	8620	3900	9226	2741
22端口(SSH)	11298	77636	64047	55201	51598	55950	10926	18401	13644	41061
21端口(FTP)	75885	21321	19335	23716	13321	22324	39838	15992	50071	13935

图 2.3 高风险端口地理分布 Top10 省市

2.2.2 高风险主机属性分布情况

从运营商分布上看，阿里云是 2025 年暴露在互联网上的高风险资产数量最多的运营商，总计开放高风险端口服务数量为 3,403,729 个；其次是中国电信和中国联通，开放高风险端口数量分别为 1,605,262 个和 1,179,907 个，如图 2.4 所示。从运营商的地理分布来看，今年香港特别行政区与台湾地区的运营商，其高风险资产数量仍在持续增长。由此可见，高风险端口的暴露情况，同时受到运营商自身的业务规模与所属地域的监管环境等综合因素的影响。

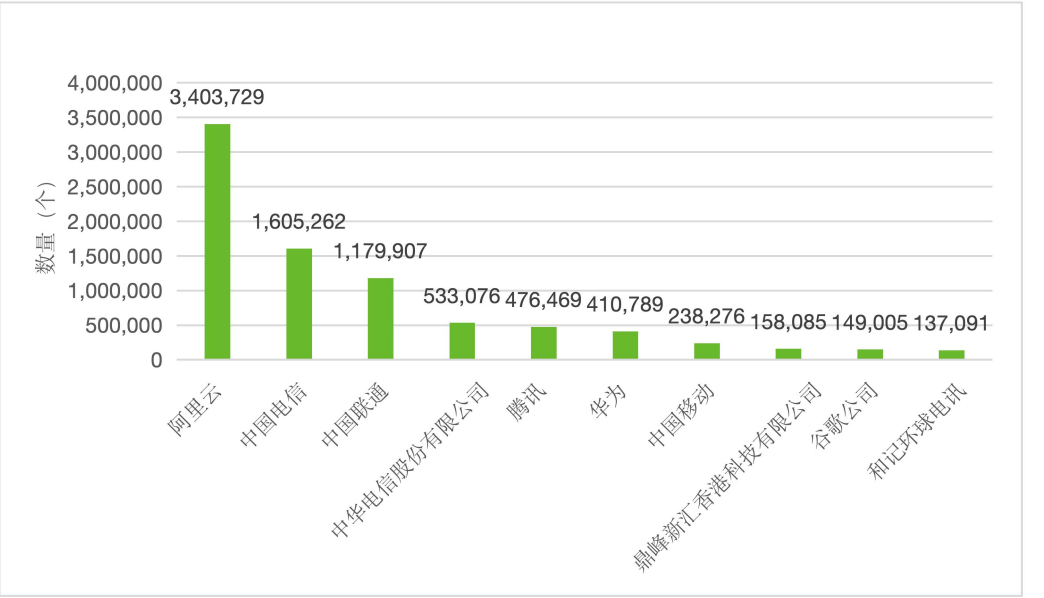


图 2.4 高风险资产所属运营商分布 TOP10

与 2024 年相比，2025 年国内主要运营商暴露在互联网上的高风险资产数量普遍增加，而国际运营商则总体呈下降趋势。从排名前五的运营商变化情况来看（见表 2.1 所示），2025 年阿里云、中国电信、中国联通以及中华电信股份有限公司的高风险资产数量排名有所上升，其余运营商则排名下降。 其中，排名前三位的依次为阿里云、中国电信和中国联通。造成头部运营商资产数量变化的原因主要有两方面：一是国内云架构部署加速，服务上云趋势导致大量云服务端口开放；二是不同运营商在网络治理过程中采取的防御策略存在差异，这影响了各自的风险暴露程度。 值得注意的是，尽管排名存在浮动，但各运营商的高风险资产数量在监管治理下整体呈下降趋势，这凸显了持续监管与治理的重要性。

表 2.1 运营商 Top5 排行情况对比

运营商	2024 年排行情况	2025 年排行情况	排名变化
阿里云	1	1	排名持平
中国电信	2	2	排名持平
中国联通	4	3	排名上升
中华电信股份有限公司	5	4	排名上升
腾讯	3	5	排名下降
华为	6	6	排名持平

从应用场景分布上看，高风险资产所属应用场景分布情况与去年相比没有显著变化。数据中心应用场景在互联网上暴露的高风险资产依然位居首位，总计开放端口服务数量为 7,065,328 个；其次是家庭宽带和企业专线，开放的高风险端口数量分别为 2,034,085 个和 619,221 个，如图 2.5 所示。后续排名依次为学校单位、移动网络、基础设施、专用出口、组织机构等。由此可见，高风险端口的暴露情况基本与其所属应用场景的实际业务需求相符。

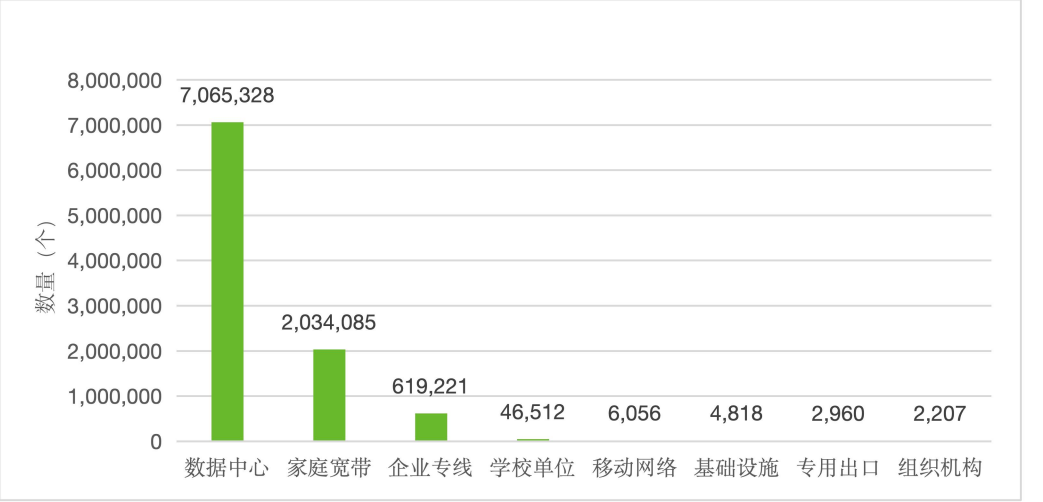


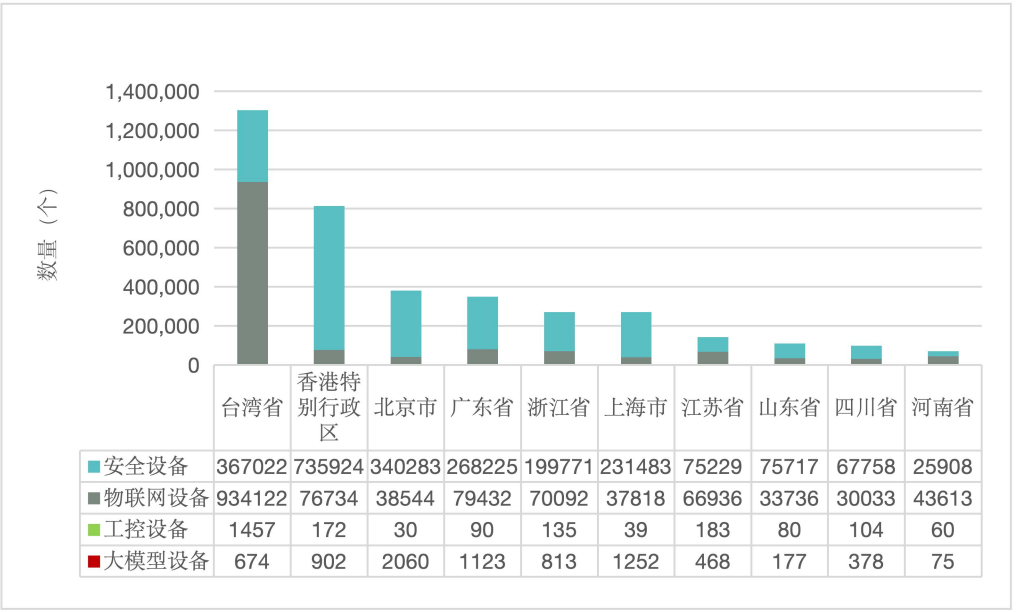
图 2.5 高风险资产所属应用场景分布

2.3 网络资产暴露态势

2025 年，互联网暴露的重要资产类型与去年相比呈现出新的变化。从资产类型分布情况看，大模型相关设备资产数量显著增长；从资产规模来看，路由器、工控设备、防火墙等关键基础设施资产数量占比较高，仍然是需要重点关注的对象。从地理分布情况看，这类重要设备资产主要集中在北京、香港和浙江等经济发达地区。

2.3.1 重要设备资产

基于失陷后可能带来较大危害的角度评估，我们重点关注的重要设备类型包含物联网、工业控制系统、安全设备和大模型设备四类。从互联网暴露数量来看，重要设备资产类型较上一年变化不大，其中摄像头、路由器和防火墙是暴露数量最多的几类资产。截至 2025 年 12 月 31 日，全国共暴露重要设备资产数量为 3,109,411，包括物联网资产、工业控制系统、安全设备和大模型设备四类。如图 2.6 所示，从地理分布来看，暴露总数排名前三的省级行政区分别为台湾省、香港特别行政区和广东省。紧随其后的第二梯队包括北京市及浙江、上海、江苏等东南沿海省市。从资产类型构成看，物联网设备资产暴露数量最多，占比约 51.88%；安全设备次之，占比约 47.98%。值得注意的是，这些暴露资产在地域分布上呈现显著不均衡性，香港特别行政区与台湾省的暴露总量尤为突出。这一现象主要与两地作为国际网络枢纽的开放性、高度密集的数字化基础设施、以及外向型经济带来的庞大物联网与工控设备基数有关。高暴露量既是其数字经济发展程度的体现，也意味着其面临的网络攻击入口更多，网络安全防护压力更为严峻。这提示我们，在数字化程度高的地区，必须将暴露面收敛和基础安全加固作为防护重点。



	台湾省	香港特别行政区	北京市	广东省	浙江省	上海市	江苏省	山东省	四川省	河南省
安全设备	367022	735924	340283	268225	199771	231483	75229	75717	67758	25908
物联网设备	934122	76734	38544	79432	70092	37818	66936	33736	30033	43613
工控设备	1457	172	30	90	135	39	183	80	104	60
大模型设备	674	902	2060	1123	813	1252	468	177	378	75

图 2.6 全国暴露重要设备资产规模分布省市

从物联网设备类型来看，暴露数量最多的资产依次为摄像头、路由器、网络电话（VoIP）、网络附属存储（NAS）、打印机、智能家居设备、交换机、工业控制系统等。 其中，摄像头、路由器和 VoIP 是暴露量最大的前三类，占比分别为 67.18%、25.15%和 2.92%（如图 2.7 所示）。暴露在互联网上的摄像头存在巨大的安全隐患。虽然互联网摄像头带来了便利，但其普遍存在的安全漏洞、弱口令以及易被远程控制等问题，一旦被攻击者利用，将对个人隐私乃至公共安全构成严重威胁。

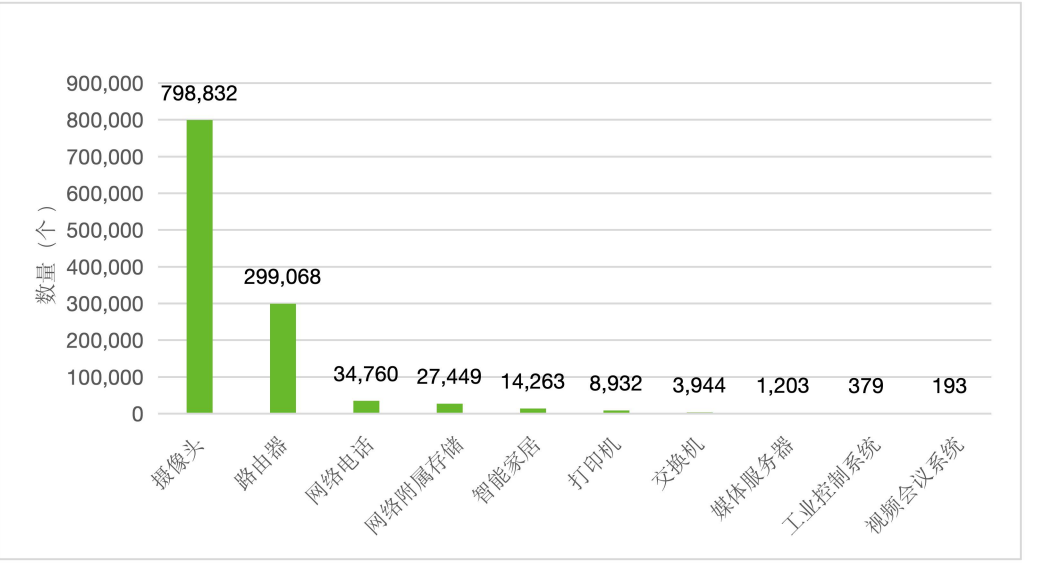
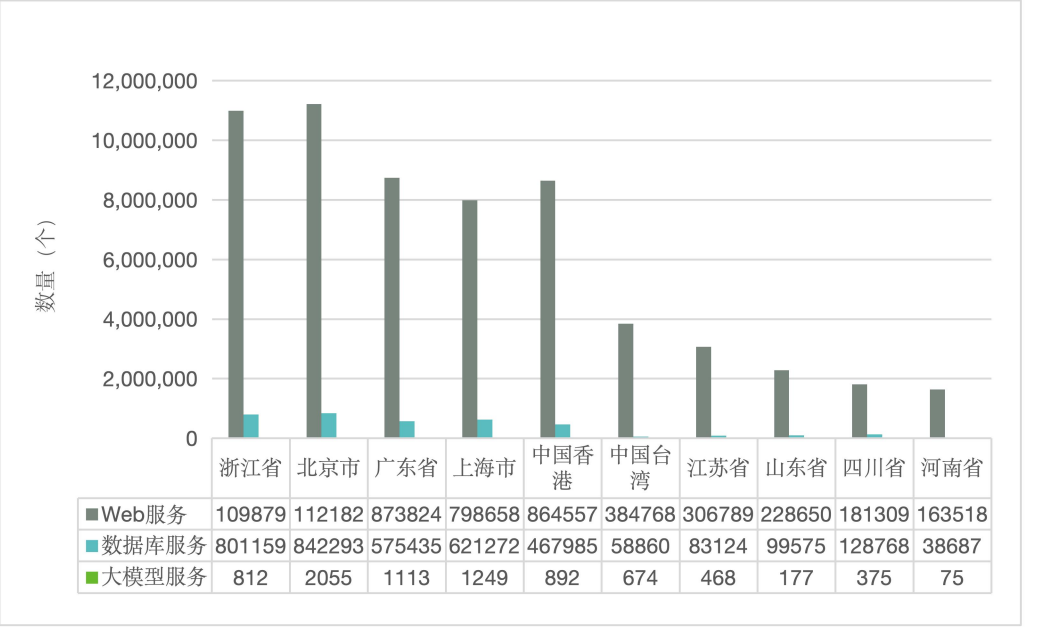


图 2.7 物联网资产暴露类型分布

2.3.2 重要服务资产

从整体情况来看，暴露在互联网上的重要服务资产，其地理分布依然高度集中，主要位于北京、香港和浙江等经济发达地区。据绿盟网络空间测绘云统计，Web 服务和数据服务的暴露面仍然严峻，而大模型服务器作为新增的重要服务类型，其暴露面迅速增长，已成为新的风险关注重点。在全国范围内，这三类服务暴露最多的省份分别是北京市、浙江省和香港特别行政区。从服务类型分布来看，Web 服务占比最高，约为 83.60%。如图 2.8 所示，具体到各类服务暴露量排名前三的省级行政区：Web 服务为北京市、香港特别行政区、浙江省；邮件服务为浙江省、北京市和上海市；数据服务则为北京市、浙江省和上海市。

从地理分布来看，重要服务的分布与其所在地区的经济发展程度呈强相关关系。尤其值得注意的是，香港作为国际金融与数据中心枢纽，其高度开放的网络环境和密集的数字化基础设施，直接导致了 Web 服务等互联网暴露资产数量的高度集中。



	浙江省	北京市	广东省	上海市	中国香港	中国台湾	江苏省	山东省	四川省	河南省
Web服务	109879	112182	873824	798658	864557	384768	306789	228650	181309	163518
数据库服务	801159	842293	575435	621272	467985	58860	83124	99575	128768	38687
大模型服务	812	2055	1113	1249	892	674	468	177	375	75

图 2.8　互联网重要服务分布 Top10 省市

数据服务是一类比较特殊的资产，在数据安全日益受到重视的当下尤其值得关注。据统计，2025 年全国暴露在互联网上的数据服务数量为 3,752,525。如图 2.9 所示，这些暴露的数据应用覆盖了数据生命周期的不同阶段，例如：常用于数据存储、使用和共享阶段的有 MySQL、Microsoft SQL Server、Redis、NFS 等；用于数据传输阶段的则有 Rsync、Zookeeper 等。

数据服务直接暴露于互联网将带来巨大的安全风险，黑客攻击、数据泄露、数据篡改、服务中断乃至勒索事件等都可能发生，严重危害系统安全与用户隐私。因此，

必须严格管控重要数据服务的访问权限，并采用加密传输等措施进行加固，同时最大限度地减少非必要数据库服务的对外开放，以规避其被攻击后可能造成的严重后果。

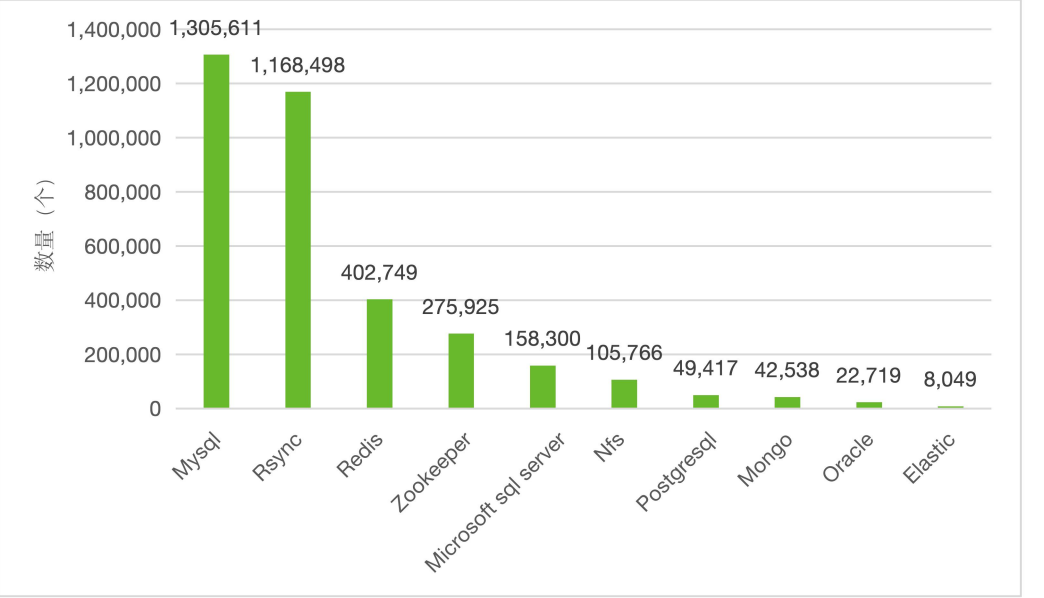


图 2.9 数据服务类型分布

全国暴露在互联网上的大模型服务数量为 36,167，如图 2.10 所示，主要包括 DeepSeek、Qwen、Llama、Gemma 和 GPT 类型。随着大模型逐步融入核心业务系统，其在业务支撑、辅助决策和信息处理等环节的深度应用，已成为网络空间中的重要资产。随着“智能对抗智能”态势的持续演进，大模型服务逐渐成为 APT 组织关注的高价值目标。在对抗视角下，攻击者可能通过接口滥用、提示词操纵、参数注入等方式对模型行为进行诱导和控制，从而窃取敏感信息、干扰业务逻辑，甚至将大模型服务作为跳板，支撑横向渗透与长期潜伏活动。此类攻击隐蔽性高、放大效应明显，对数据安全、业务连续性及整体对抗态势构成重大风险。未来，随着智能化技术的进一步发展，大模型安全防护面临的挑战将进一步加剧，亟需持续关注和精细化管控。

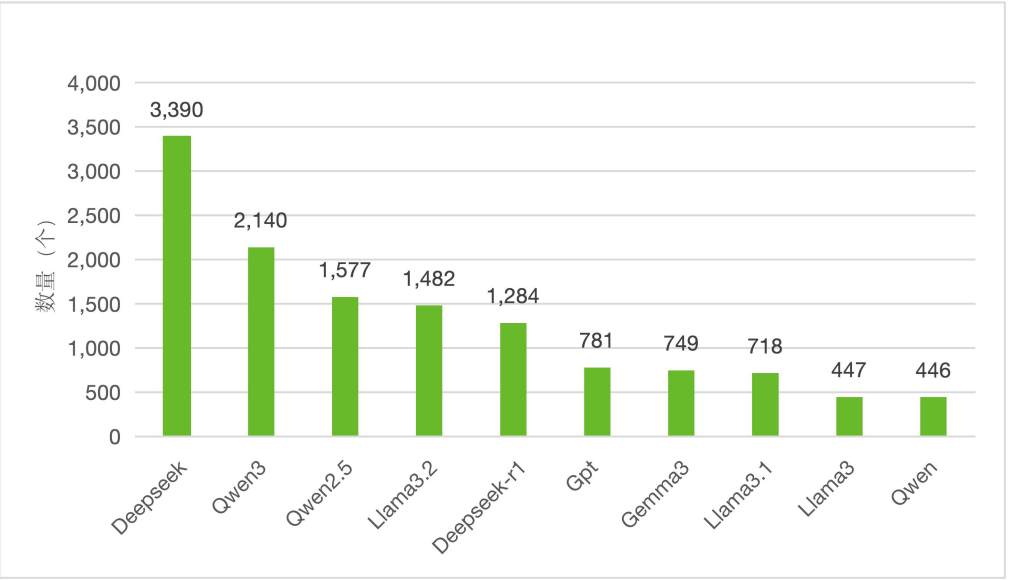


图 2.10 大模型服务类型分布

2.4 暗网态势

暗网作为网络犯罪、数据泄露和黑客攻击等威胁的主要来源，是各种非法活动的聚集地。2025 年，绿盟威胁情报中心对暗网的数据交易市场进行了持续的监测和分析，累计发现国内数据泄露事件 816 起，涉及泄露数据超过 100 亿条。

从事件数量看，国内数据泄露事件数量对比去年有大幅减少（相关报告显示，数据交易数量减少 36.99%），整体数据泄露次数明显下降，推测这可能与数据安全治理越来越受国家重视有关。从行业分布看，金融、重要单位及重要机构和电商等行业是数据泄露的主要来源。值得关注的是，重要单位和机构组织占比较去年上升，说明虽然总体数量在下降，但是针对重要目标的数据安全风险却更加严峻。从泄露主体看，涉及大规模人员信息的泄露事件占比最多。其中，常见数据类型主要以手机号、身份

证号为主，该类数据共涉及泄露事件 321 起，泄露清单示例如图 2.11 所示。人员信息作为许多软件、平台在注册、登记时的必填项，应用广泛，是数据窃取的主要目标之一。

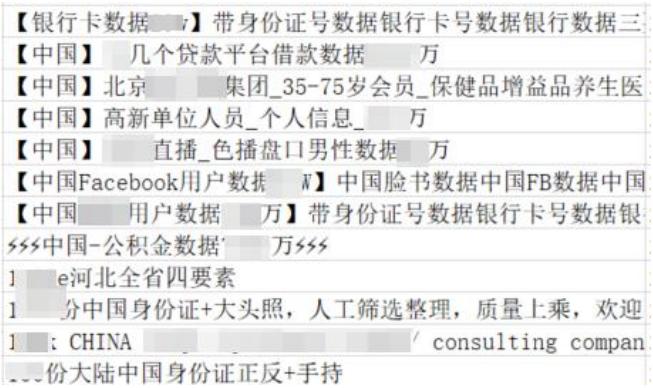


图 2.11 泄露清单示例

2.4.2 涉及行业

2025 年泄露事件数量较多的行业依次为金融、重要单位和机构组织、教育、电商、医疗、能源与工业和物流。

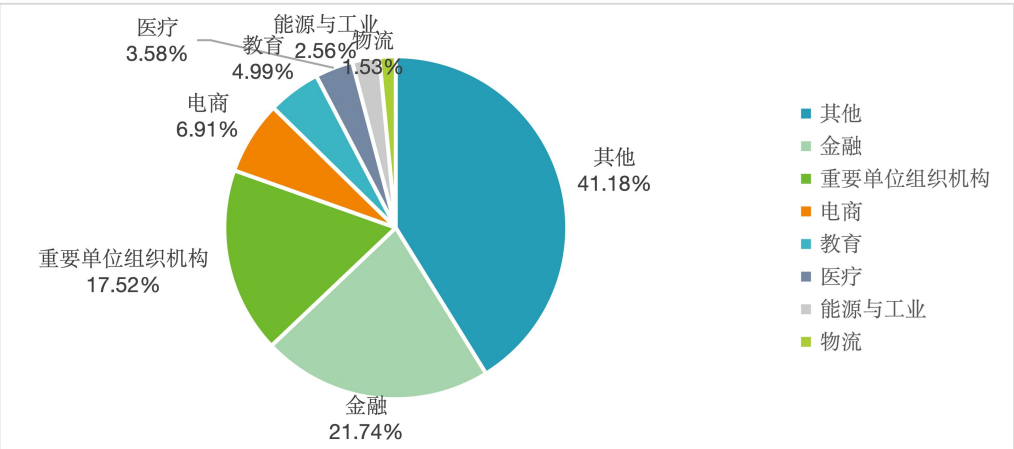


图 2.12 2025 年国内数据泄露行业统计（泄露事件量）

从泄露事件的次数角度，泄露事件较多的行业包括：金融、重要单位组织机构、电商、教育、医疗、能源与工业和物流。

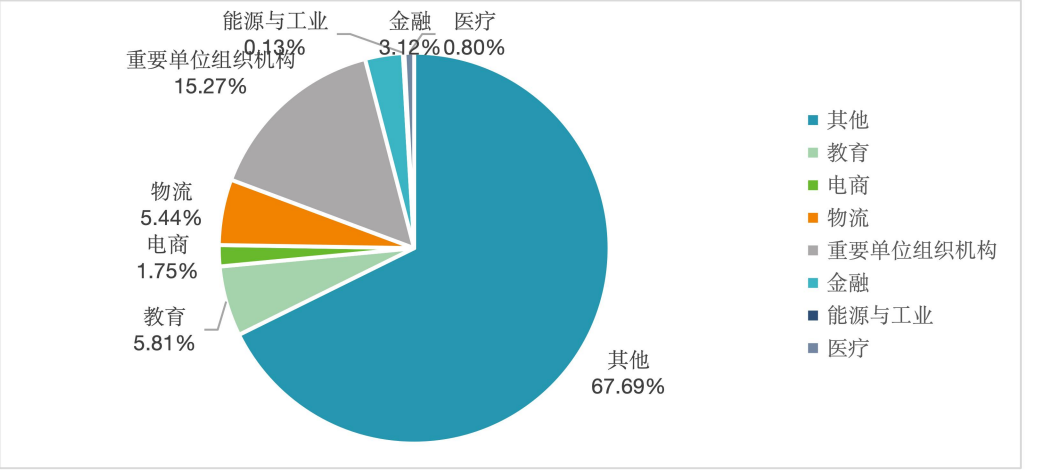


图 2.13　2024 年国内数据泄露行业统计（泄露数据量）

值得注意的是，在重要单位组织机构领域，数据泄露事件和泄露数据量均有增加。重要单位组织机构的数据中主要包含重要单位收集存储的公民信息，和重要单位内部人员信息，包括政府机关、研究所、大型企业等。前者往往内容真实且全面，后者更是直接涉及到国家安全稳定，因此这两类数据的泄露往往会造成极大的隐患，建议国家和行业后续增加关注和重点保护。

与金融领域相比，尽管教育、电商和物流仍是数据泄露事件与数据泄露量的核心区域，但过去一年在治理方面已显现出更明显的收敛趋势。电商和物流领域的数据主要包括平台用户信息、网购订单数据和详细地址等，因其覆盖范围极广且高度关联个人现实生活，成为数据泄露的高危领域。这些数据分散在产业链的多个机构中，其中一环被突破就可能发生大规模数据泄露。此类数据泄露的危害是双重的：一方面，可能直接催生精准诈骗等犯罪活动；另一方面，泄露信息中往往混杂着重要机构的内部

人员数据，一旦被不法分子筛选、分析、整合，可能对机构乃至国家安全构成更深层、更难以估量的潜在威胁。

2.4.3 涉及地区

对比去年来看，数据泄露事件涉及的地区分布没有显著变化，仍主要集中在经济发达和人口密集的省份。从整体数量来看，中国大陆地区数据泄露事件数量有显著下降，体现出相关省份在数据安全治理上的成效；相比之下，台湾省等治理相对薄弱地区的数据泄露事件占比更为凸显。

据绿盟威胁情报中心统计，在泄露信息中是否提及数据的地域信息进行分析，共发现有 443 个数据泄露事件包含地域信息，数据泄露相关省份 Top 排行如图 2.14 所示。其中，提及次数最多的几个省份分别是台湾省、广东省和江苏省。与 2024 年相比，大陆各省出现次数呈整体下降趋势，而台湾省的提及次数则远超其他省份，这可能与其对数据安全的重视程度相对不足有关。由此可见，在持续的数据安全治理行动影响下，不法分子更倾向于窃取和售卖治理薄弱地区的数据，这进一步说明数据安全治理是一项需要长期持续推进的工作。

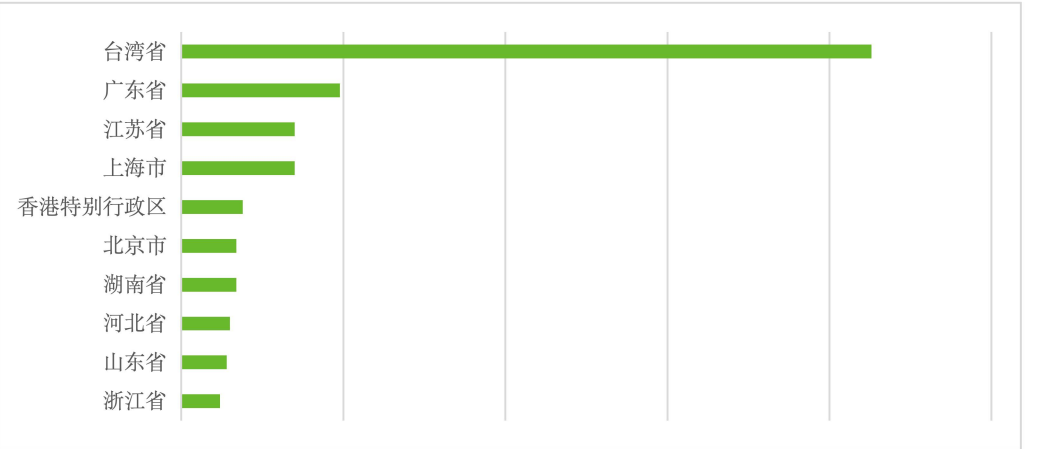


图 2.14 数据泄露相关省份 TOP10

2.4.4 涉及内容

据绿盟威胁情报统计，在已泄露的数据中，手机号和身份证号等涉及个人信息的数据字段出现频率最高，详细数据字段分布如图 2.15 所示。其中，身份证号、手机号分别出现了 173 次和 148 次。这两类数据是许多软件和平台在注册、登记时的必填项，因此一旦发生数据泄露，它们暴露的风险也相对更高；尤其是身份证号，由于其唯一性和高度敏感性，更是不法分子窃取和交易的重点目标。这类核心个人信息的泄露，极易引发精准电信诈骗、身份盗用等严重后果。

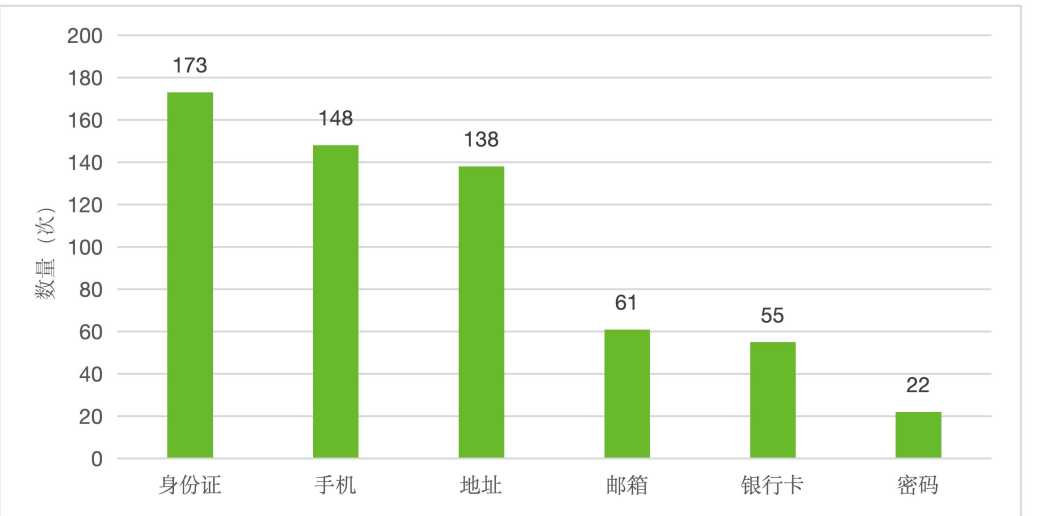


图 2.15 数据泄露中的已知字段

2.4.5 涉及事件类型

从泄露事件类型结果来看，2025 年的数据泄露事件中，大部分为对历史或重复数据进行置顶、调价后重新发布售卖的“N 次新发帖”。

结合实际分析结果，我们可依据数据的泄露时间和内容新旧，将泄露事件大致分为历史数据泄露事件和新数据泄露事件两类。其中，历史泄露事件主要指对既往泄露

数据经过一定的加工处理（如去重、分析、提取、组合或混淆）后，再次进行售卖的情况。如售卖者从历史大规模泄露数据中筛选出部分数据然后二次兜售。

据绿盟威胁情报中心统计，在 2025 年数据泄露事件中，历史泄露事件占比数量超过总量的一半，占比约为 68.34%，如图 2.16 所示。值得注意的是，无论是新数据泄露，还是历史数据的多次售卖，数据交易市场往往是真假参半。因此，一旦发生数据泄露事件，建议结合实际泄露数据的具体情况，立即对相关业务系统展开安全风险排查，并对已确认泄露的数据字段增强身份认证等防御措施，以防止后续更大范围的数据泄露发生。

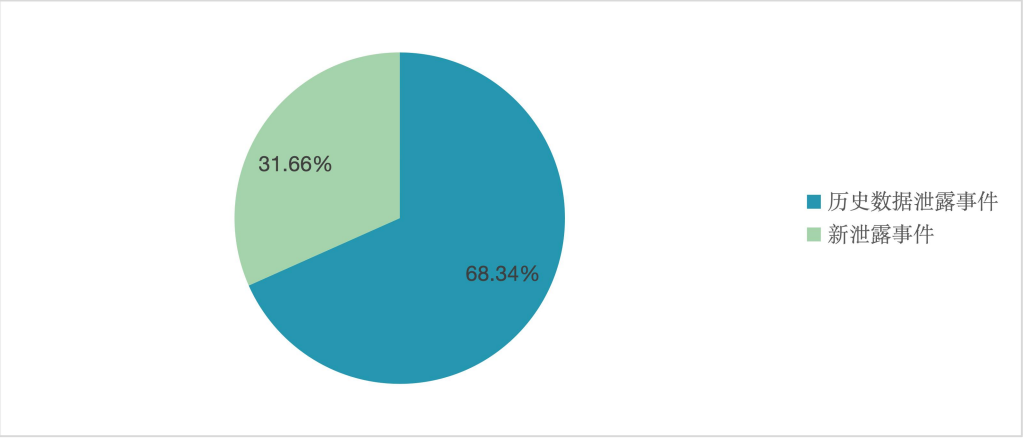


图 2.16　数据泄露事件类型占比情况

2.5　IPv6 安全态势

绿盟威胁情报中心选取了国内数百家部署了 IPv6 业务的典型单位(行业覆盖政府、教育、能源、医疗、交通等主要行业），对其在 2025 年 IPv6 环境下遭受的攻击告警日志进行分析，以观察国内单位在 IPv6 环境中面临的威胁态势。

分析发现，IPv6 网络仍然是攻击者重点利用的目标。

（1）从攻击类型看，IPv6 的整体攻击类型与去年相比虽无显著变化，但攻击的复杂程度有所提升，攻击事件总量同比增加了 66.10%。

（2）从地理分布看，国内 IPv6 攻击源的分布发生了变化，2025 年主要集中在贵州、四川、重庆等地。这一变化可能与国家为发展新质生产力，推进大数据和算力基础设施建设的区域布局有关。

2.5.1 IPv6 风险分布

2025 年，IPv6 相关漏洞数量创下历史新高。据绿盟威胁情报中心统计，NVD 数据库历年 IPv6 相关漏洞收录数量如图 2.17 所示。数据显示，自 2002 年起累计公布的 IPv6 漏洞共有 800 个，其中 2025 年新增 136 个。尽管 2025 年的新增数量与前一年相比增长态势趋于平缓，但从整体趋势看，历年 IPv6 漏洞数量仍呈现明显的上升趋势。

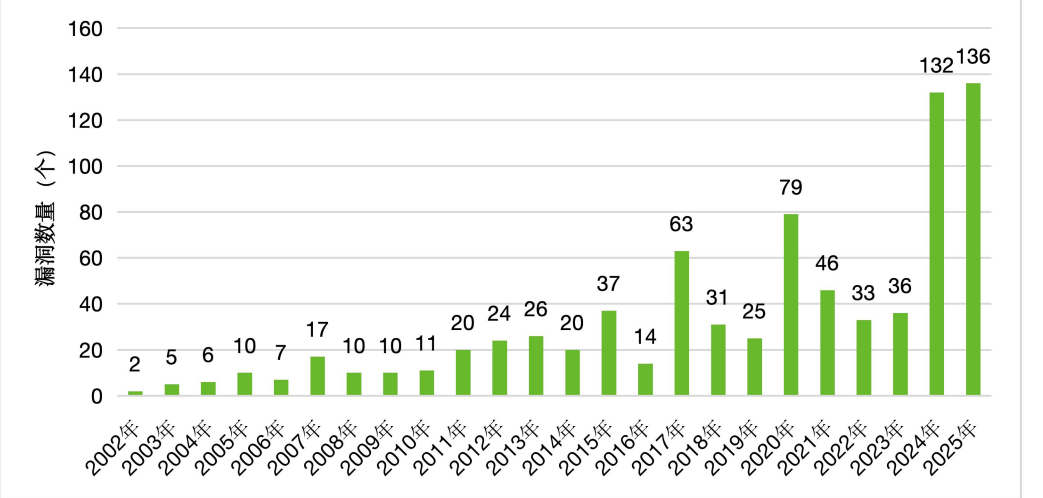


图 2.17 2002 年以来 IPv6 相关漏洞统计

依据 IPv6 漏洞的 CVSS3.1 标准 高危的漏洞占比小，低危的漏洞占比高，如图 2.18 所示。

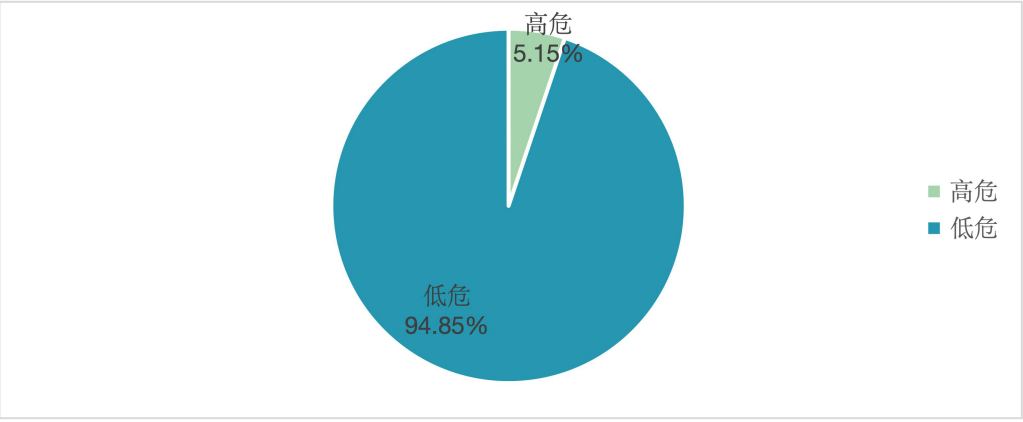


图 2.18 2025 年 IPv6 安全漏洞等级分布

2.5.2 IPv6 攻击类型

从整体攻击类型来看，IPv6 攻击类型与去年相比变化不大，仍以漏洞利用和暴力破解为主，但攻击量增长迅速。

据绿盟威胁情报中心统计，2025 年 IPv6 网络攻击类型分布如下图 2.19 所示。其中，漏洞利用、暴力破解、拒绝服务、扫描探测、恶意软件等是排名靠前的网络攻击类型，占比分别为 61.24%、16.34%、13.47%、5.22%和 2.14%。由此可见，2025 年的攻击类型分布与去年基本一致，排名前两位的依然是漏洞利用和暴力破解。

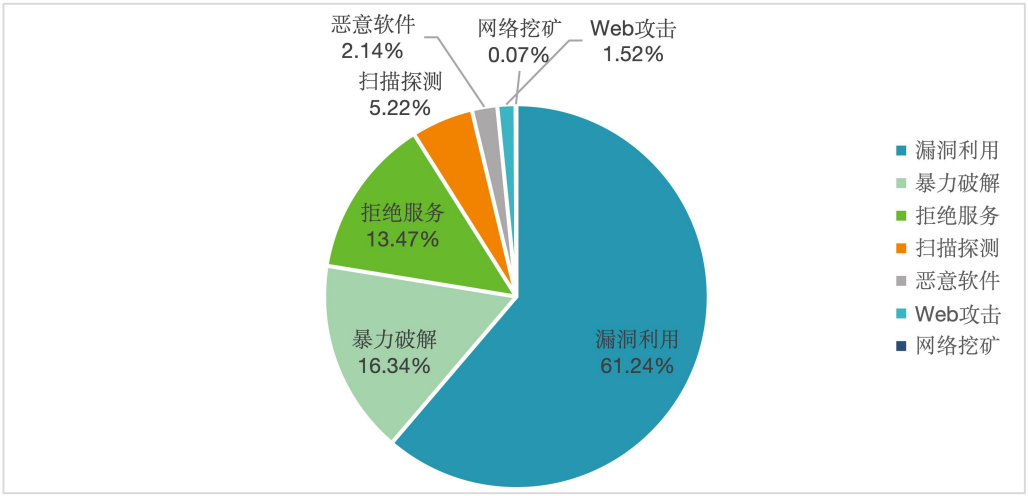


图 2.19 IPv6 攻击分类占比情况

对比去年，IPv6 漏洞利用攻击风险增加。其中，攻击者发起漏洞、DDoS 攻击方式所涉及的漏洞数量增长 46.12%，如图 2.20 所示。

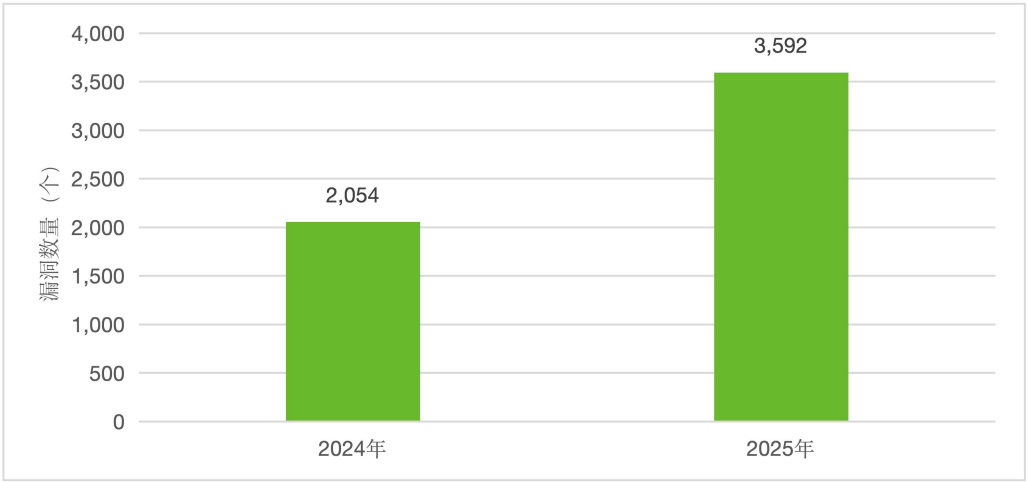


图 2.20 IPv6 漏洞利用数量

2.5.3 IPv6 攻击源分布

目前，跨境 IPv6 攻击尚不显著，攻击源仍主要来自国内 IPv6 地址。这可能是由于国外 IPv6 的推广程度与地址规模普遍低于我国等多方面原因造成的。据绿盟威胁情报中心统计，2025 年来自国内攻击源 IPv6 数量共计 11,384,850 个，如图 2.21 所示。随着国内 IPv6 部署规模的持续快速扩大，攻击者利用境内 IPv6 资源发起攻击，已成为 IPv6 安全领域需要重点关注和治理的核心问题。

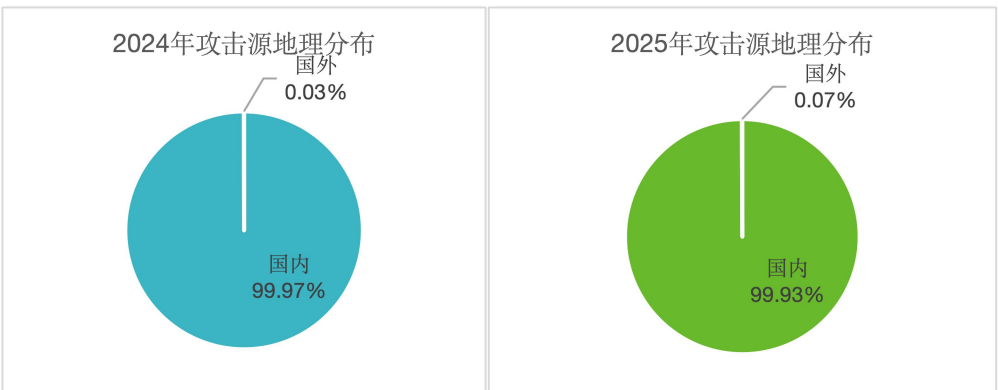


图 2.21 近两年 IPv6 攻击源分布情况对比

据绿盟威胁中心统计，国内的 IPv6 攻击源更多集中在西南和华南地区，如图 2.22 所示。其中，IPv6 攻击数量排名前三的地区分别是贵州省、四川省和广西壮族自治区，其供给量占比分别为 36.51%，15.29%和 6.74%。贵州作为国家大数据综合试验区及国家“东数西算”工程的重要枢纽节点之一，在推动新质生产力发展的同时，其广泛部署的 IPv6 应用也显著增加了网络空间的暴露面和潜在安全风险。因此，在大力推进大数据与算力基础设施建设的过程中，必须同步保障网络安全防御体系的建设，特别是加强对 IPv6 资产的漏洞管理与防护，有效防范针对 IPv6 的漏洞利用与滥用事件。

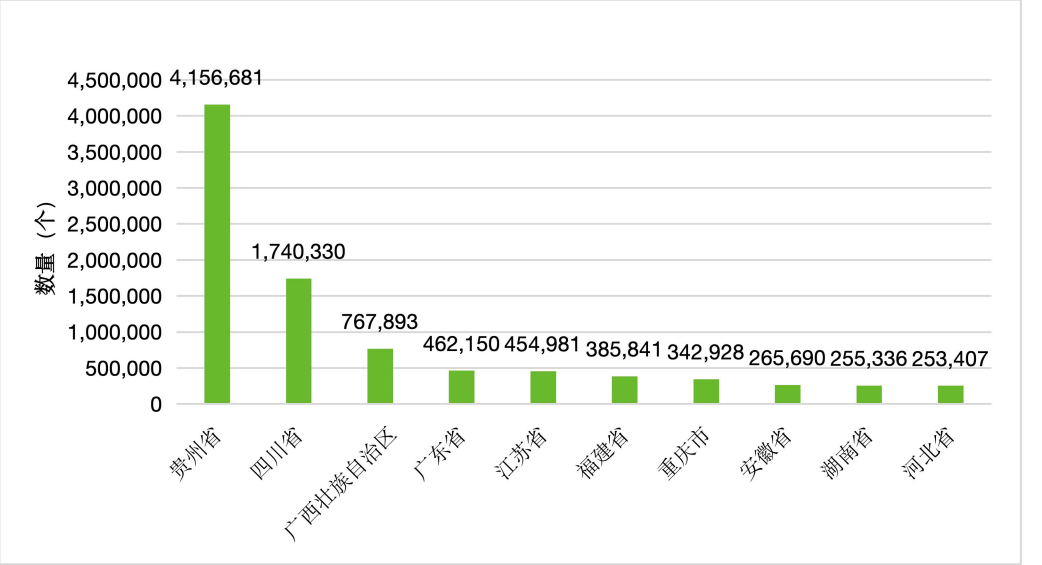


图 2.22 IPv6 国内攻击源地理分布情况

2.5.4 IPv4 安全态势

IPv4 攻击类型：

据绿盟威胁情报中心统计，2025 年恶意 IP 的主要攻击类型集中于扫描探测、漏洞利用、暴力破解及 C2 主机等。其中，占比最高的是扫描探测，约为 46.26%；其次是漏洞利用、暴力破解和 C2 主机，占比分别为 26.90%、17.88%和 2.77%。详细攻击类型分布情况如图 2.23 所示。

观察 2025 年的 IPv4 攻击类型的前三名分布，扫描探测、漏洞利用和暴力破解这三类攻击的综合占比超过 90%，它们依然是主要的、高频的攻击活动。

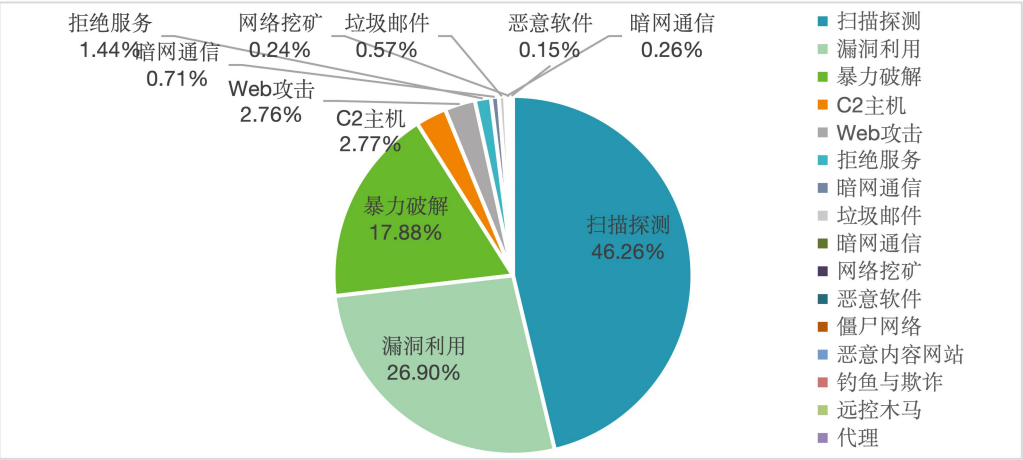


图 2.23 恶意 IP 攻击类型分布

IPv4 攻击源分布：

从地理分布来看，攻击源主要分布在广东、浙江、江苏、福建、河南等地，具体分布数量如图 2.24 所示。与去年相比，攻击源的整体地理分布未发生显著变化。从攻击资产总量看，广东省的攻击源 IP 数量最多，占比约为 34.8%。

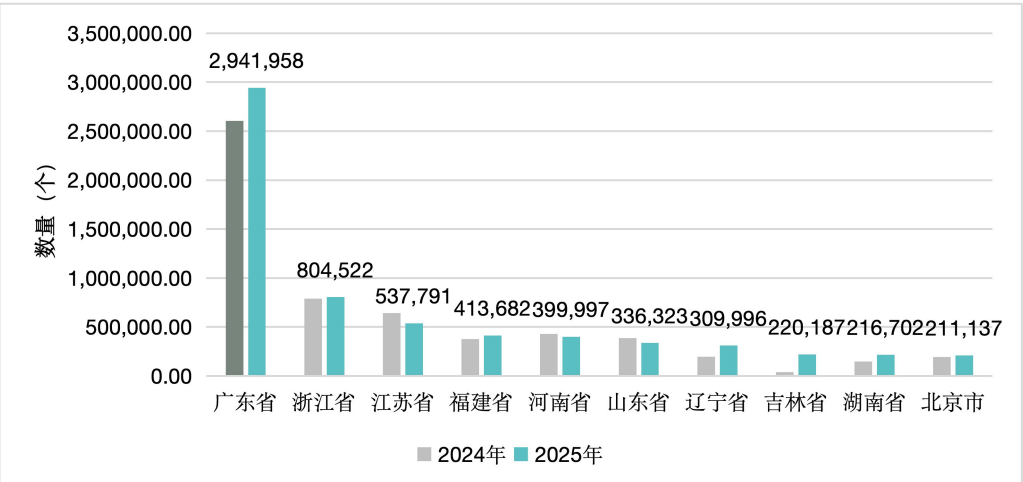


图 2.24 国内恶意 IP 地理分布

03

技术发展篇

3.1 韧性安全

3.1.1 什么是韧性安全

在 2025 年 10 月 23 日通过的《中共中央关于制定国民经济和社会发展第十五个五年规划的建议》中提出，要“加强基础设施统筹规划，优化布局结构，促进集成融合，提升安全韧性和运营可持续性”。

这里提到的“韧性”对应的英文词汇是 resilience。无论被翻译为“韧性”还是“弹性”，都描述了系统或组织在压力、灾害、故障、攻击等不利的环境下，保持其核心功能正常运行并快速恢复到正常水平的能力。

近几年 ISO、MITRE、NIST 等多个网络安全组织对网络安全韧性给出自己的理解定义，虽然各个组织的定义在表述上各有不同，但主体定义基本一致：

网络韧性（cyber resiliency）是指网络系统或网络资源在面对不利条件、承受压力、攻击或者损害的时候所展现出来的预测、抵御、恢复和适应能力。

MITRE 是国际上比较早提出网络韧性工程框架的机构。其在网络韧性方面提出的工程框架由目标、任务、技术三部分组成，逐级细化、相互映射。美国国家标准与技术研究院（NIST）基本全面接受了 MITRE 的网络韧性工程框架，形成了特别出版物：《开发网络韧性系统：一种系统安全工程方法》（NIST SP800-160 v2r1）。

图 3.1 NIST SP800-160v2r1 韧性安全工程框架

3.1.2 韧性安全建设发展观察

虽然网络安全韧性的概念在国内没有被大规模地应用，但网络安全韧性建设的工作却一直在各类关键信息基础设施中以不同形式落地执行。

2000 年后，电信运营商相继启动灾备体系建设，构建了“两地三中心”（生产中心+同城灾备+异地灾备）的分级灾备体系。三大运营商均设立专门的应急通信管理部门，统筹灾备与应急工作。2025 年工业和信息化部等 14 部门还联合印发《关于加强极端场景应急通信能力建设的意见》建立跨部门供需对接、信息共享、协同保障机制，保障基层韧性不断联的极端场景大应急通信保障能力。

在金融行业，在 2008 年中国人民银行发布《银行业信息系统灾难恢复管理规范》（JR/T0044-2008），规定银行业灾难恢复等级与技术要求，标志行业标准正式确立。目前绝大多数金融机构都完成了“两地三中心”建设。在《金融科技发展规划（2022-2025 年）》中明确要求“积极采用多活冗余技术构建高可靠、多层级容灾体系”。

在能源行业，国家电监会明确提出“安全分区、网络专用、横向隔离、纵向认证”的防护原则。国网、南网等也完成了“两地三中心”架构建设。在《“十四五”能源领域科技创新规划》：明确要求开展电力系统遭受严重自然灾害、物理攻击、 网络攻击等非常规安全风险识别及防范研究，提高非常规状态电网安全稳定防御和应急处理能力。

可以看出上述这些关键信息基础设施韧性安全的建设重点都是通过提高系统冗余性和动态迁移能力，来提高系统恢复能力的。但我们也可以看出，针对网络攻击方面的安全建设没有和业务持续性建设融合起来作为关键信息基础设施韧性安全建设看待。

3.1.3 关基韧性安全设计原则

（1）聚焦关键信息基础设施业务使命与核心业务

网络韧性工程建设的最终目标是保证关键信息基础设施业务的持续运行。以终为始地看待这一问题，必然要将核心聚焦于保障关键信息基础设施业务的持续运行的能力，使得关基核心业务系统在遭受敌对方威胁甚至失陷的情况依然有能力持续运行。这一工程设计原则与业务连续性设计原则一脉相承。但传统的业务连续性设计更多关注业务系统面对高负载、灾害等场景的持续运行能力；而网络韧性安全工程则更多聚焦于 APT 攻击场景的核心业务持续运行能力。

（2）聚焦 APT 攻击的影响

APT（Advanced Persistent Threat，高级持续性威胁）攻击，是攻击者为达成特定战略目标，依托复杂技术手段与长期规划，对关键信息基础设施系统展开隐蔽、持续渗透的网络攻击模式。

APT 攻击具有三大显著特点：一是高隐蔽性，二是强针对性，三是持续性。从投入资源成本上可以看出，APT 攻击者投入大量资源成本才可以完成一次成功的入侵攻

击，那么其预期的“攻击成果”必然要与其投入成正比。因此 APT 攻击者的目标大多是具有极高战略价值的关键信息基础设施。

网络韧性工程重点针对的是像 APT 这类的高级的、复杂的攻击行为对关键信息基础设施业务的影响。在网络韧性工程框架中的任务目标与关键技术，大多是以防范 APT 攻击，减轻 APT 攻击影响为目标的。

（3）假定对手必将攻陷系统并持续存在于系统中

APT 攻击凭借其技术先进性与决心持久性，引出网络韧性工程的核心讨论前提：无论关键信息基础设施的安全产品防护能力多么完备，都难以实现对 APT 攻击的防护，更无法保证在攻击发起后迅速检测与响应。

因此，网络韧性工程的设计逻辑与传统安全防护存在本质区别，其设计原则与技术方案是围绕“敌对方已攻陷系统，并可能长期潜伏”这一核心假设展开。核心目标从“阻止攻击”转向“在攻击发生后最小化损失、保障核心业务连续性”。

网络韧性工程的核心实施方向应聚焦于 “限制攻击影响扩散”，即通过技术手段阻断攻击者在系统内部的横向移动路径，避免其从外围系统渗透至核心业务节点，从而降低对关键业务的破坏性影响。具体而言，可通过构建网络微隔离架构，将核心业务系统与外围办公等区域进行逻辑隔离，严格控制跨区域访问权限；部署基于行为分析的异常检测系统，及时发现攻击者的横向移动尝试；建立最小权限访问机制，根据业务需求精准分配系统权限，减少攻击者利用高权限账户横向渗透的可能性；同时完善应急响应预案，在发现攻击迹象时快速启动隔离、止损、恢复流程，确保核心业务在攻击期间仍能维持基本运行。

这一系列措施的核心逻辑，正是网络韧性工程 “在极限环境确保关键业务生存”的价值体现，在 APT 攻击难以彻底阻隔的现实背景下，为关键信息基础设施构建起尽可能厚的缓冲防线。

3.1.3.1 从 APT 攻防视角观察韧性安全

在前述网络韧性安全关键设计原则中，我们提出要聚焦 APT 攻击的影响。因此我们参考最接近 APT 攻击的网络安全攻防演练，从国家攻防演练攻击队的视角对关键信息基础设施的韧性安全能力开展观察。

绿盟科技常年参与国家级和行业区域级的攻防演练活动，根据绿盟科技烈鹰战队等攻击队的实际参演观察， 2025 年攻防演练活动更加聚焦于关键信息基础设施韧性安全能力演练：

- 加大关键信息基础设施靶标数量，总计 260 余个靶标中，广电、金融、交通、能源、运营商、政府等行业靶标占比近 70%。其中强逻辑隔离和物理隔离单位达 84 个。
- 靶标不分组，攻击者具备全局视角，不局限于某一个单独的靶标，可以面向整个行业发起攻击，更加逼近 APT 组织的意图。
- 攻击队聚焦关键信息基础设施核心系统上，关键信息基础设施单位靶标攻击得分按倍率系数放大，非核心系统的攻击得分降低。

通过这些演练规则的变化可以看出，攻击队的攻击强度与手段已经非常接近 APT 攻击者了。在攻击队的视角下，参与演练的关键信息基础设施靶标存在韧性安全问题：

- 关基单位在“假定对手必将攻陷系统并持续存在于系统之中”方面的准备并不充足。对关基靶标的攻击，最大的困难来自“破门”，需要结合多种战法、消耗大量资源。**一旦实现“破门”，进入后的横向移动难度会大幅下降。**
- 业务上云、微服务架构带来海量风险点。大规模应用的微服务、容器，极易带来微隔离失效问题。攻击者进入云系统后，通过横向移动可能控制基础设施，可以高效攻陷云上全部业务。

◆ 行业产品缺乏异构性导致通用漏洞问题。各个行业均存在行业通用软件供应商，提供的产品被行业内客户广泛采用。统一的软件系统降低了攻击难度，这些产品一旦被发现有安全漏洞，可以被攻击者利用，突破多个目标，打穿整个行业。近几年攻防演练中多次出现一个漏洞打穿多个目标的事件。

◆ 通过对供应商、目标单位工作人员远程运维所使用的跳板机、私搭隧道的攻击，可以获得关键信息基础设施的内部运维信息与权限，能够批量攻破内网目标。防守方难以对供应链攻击进行有效防范。

可以看到，上述问题均暴露出目标单位网络韧性不足，在某一环节出现问题时，无法快速有效遏制问题的扩大。

3.1.3.2 从韧性安全视角审视关基安全建设

韧性安全作为一种安全建设的能力概念，受到持续关注，被不断探讨。但韧性安全不是某一种特定产品，而是通过多种安全技术有机结合而获得的“预测、抵御、恢复和适应能力”。因此，提升关键信息基础设施系统韧性安全的过程，不是通过建设一个“韧性安全平台”来实现，而是**在“系统一定会被攻破”的假设前提下，审视关键信息基础设施系统安全建设是否具备韧性安全能力，是否需要在薄弱的方向提升韧性安全能力。**

NIST 的网络韧性安全 14 项关键技术组合应用可以提升系统韧性安全能力，有很多也已经被我们的关键信息基础设施所采用：

◆ 纵深防护、分析监控、态势感知、自适应响应等技术在我国关键信息基础设施中以等保/关保安全建设、态势感知系统、SOAR 等形式加以落地。

◆ 动态迁移、冗余性、异构安全等技术在业务连续性建设中已经基本实现。

◆ 权限控制、非持久性授权、微隔离等技术在零信任安全建设中实现。

韧性安全核心目标中有一项是“适应（Adapt）”，而“适应”是一种动态的过程描述。可以看出关键信息基础设施韧性安全是一个动态的自适应演进的建设过程。通过不断地改进安全防护技术、优化组织运营流程、培训运营人员，来适应新业务带来的环境变化、适应攻击者的新型战法，最终使关键信息基础设施实现持续的韧性安全能力。

3.1.4 关基韧性安全建设的未来发展趋势

我国在关键信息基础设施相关的标准也在逐渐关注韧性安全，绿盟科技持续跟踪并参与我国关键信息基础设施相关的标准政策、课题研究，更能够看到这一明显变化。例如《信息安全技术 关键信息基础设施安全保护要求》（GB/T 39204-2022），与以往相关标准相比，加强了业务识别、冗余备份、供应链安全保护、收敛暴露面、攻击阻断与分析改进、攻防演练、应急演练、时间恢复等多项要求，基本覆盖了前文提到的所有韧性安全技术能力。

在正在征求意见的某国家标准《关键信息基础设施安全保护能力指标体系》中，将关键信息基础设施分为基本保护级、强化保护级、战略保护级。三个不同的保护级别都有保护关键业务持续服务能力要求：

- 基本保护级：保证关键业务持续服务，具备应急恢复与运行能力。
- 强化保护级：实现系统级弹性，在部分功能失效时，关键业务能够安全稳定运行。
- 战略保护级：在面对国家级网络攻击、严重自然灾害等极限情况时，具备弹性应对、自适应防御能力，至少能确保关键业务极限情况下持续运行。具备自动快速恢复功能，保证关键业务可持续提供服务。

其评估指标体系覆盖安全管理、系统结构、技术防护和安全运营四大能力类，能力类下对应的 14 个能力族、50 多个组件、200 多个能力指标项，基本覆盖了前文提到的所有韧性安全技术能力。

可以看到，未来我国关键信息基础设施安全建设工作将围绕提升系统韧性安全开展。而在这一趋势下，重视韧性安全建设思路、研究韧性安全技术要求、补齐韧性安全能力短板将逐步成为关键信息基础设施安全建设的主要方向。

3.2 AI 赋能网络安全

3.2.1 热点安全事件

3.2.1.1 哈尔滨亚冬会遭遇以人工智能为核心的网络攻击

2025 年 2 月，哈尔滨第九届亚洲冬季运动会遭遇了史上首次以人工智能为核心技术的大规模网络攻击。根据中国国家计算机病毒应急处理中心的溯源报告，美国国家安全局（NSA）利用 AI 智能体技术对中国赛事信息系统及关键基础设施发起 27 万次攻击。

在攻击发生后，国家计算机病毒应急处理中心等网络安全机构的技术专家迅速行动，开展网络攻击溯源调查。依托全网安全大数据和安全大模型，部署 3000 个安全智能体，实现每秒处理 200 万条攻击日志的实时对抗。第一时间对此次大规模境外网络攻击进行溯源，从 27 万次攻击中提取出 136 个独特 AI 行为特征。结果显示，此次攻击的幕后黑手是美国国安局，通过代码风格分析锁定 NSA 三名特工的操作习惯，首次锁定了发起攻击的个人。

此次事件揭示了网络安全领域的范式转变。美国 NSA 的 AI 攻击战术虽然被成功挫败，但其展现的技术能力已远超传统网络战范畴。未来需要构建“以 AI 对抗 AI”的新型防御体系，同时推动国际社会建立 AI 军事化应用控制框架。这场发生在数字空间的“哈尔滨保卫战”，终将成为人类应对智能时代安全挑战的重要里程碑。

3.2.1.2 全球首例 AI 自主网络攻击

2025 年 9 月，Anthropic 公司披露了一起震惊全球网络安全界的重大事件——全球首例 AI 自主网络攻击。这起被认为是首次有文献记录的、由 AI 主导执行且几乎无需人工参与的大规模网络攻击活动，展现了 AI 智能体在恶意用途上的巨大威胁。

攻击者通过声称代表合法的网络安全公司进行防御评估。他们开发了一个定制的编排框架，使用 Claude Code 和 Model Context Protocol 将复杂的多阶段攻击分解为离散的技术任务，每个任务在单独评估时都显得合法。在整个攻击过程中，AI 完成了 80%-90%的任务，人类仅在关键决策点介入（每轮约 4-6 次）。

这起事件的影响力在于，它首次向世界展示了 AI 智能体在网络攻击中的巨大潜力。此类系统可长时间自主运行，以极少的人为干预完成复杂任务，显著提升了大规模网络攻击的可行性。报告指出，随着攻击手法迅速演进，具备智能体能力的 AI 系统可完成原本需要整支资深黑客团队才能执行的任务，包括分析目标系统、生成攻击代码、处理大规模被盗数据等，甚至资源有限的组织也有能力发动此类行动。

3.2.2 国内外发展现状

2025 年全球 AI 赋能网络安全呈现爆发式发展态势，市场规模持续扩大，AI In Cybersecurity Market（2025-2030）报告中所述 2024 年全球人工智能在网络安全市场的规模估计为 253.5 亿美元，预计到 2030 年将达到 937.5 亿美元，年复合增长率 24.4%，生成式 AI 网络安全市场增长更为迅猛。据 Generative AI in Cybersecurity Market 报告统计年复合增长率达 26.5%。在此背景下，“AI 对抗 AI”已成为攻防双方的刚性需求，智能安全体正引领新一代安全范式。应用场景方面，AI 已在安全运营、安全攻防、威胁情报、代码检测等环节深度落地，并逐步渗透至红队测试、自动化响应

等高阶领域，企业对安全能力整合的需求激增，推动 AI 驱动的统一安全运营平台成为主流趋势。

政策方面，国内 2025 年相关政策法律与标准密集落地，10 月十四届全国人大常委会第十八次会议通过的《网络安全法》修订案（2026 年 1 月实施）新增条款，将 AI 赋能网络安全上升至国家战略层面，明确国家支持 AI 关键技术研发与网络安全创新应用；标准体系持续完善，GB/T 45652-2025、GB/T 45654-2025 规范生成式 AI 训练数据与服务安全，GB/T 45958-2025 确立 AI 计算平台安全框架，GB 45438-2025 明确 AI 生成内容标识要求。这些政策标准形成法律引领与标准支撑的驱动，既通过强制要求明确发展方向，又以分类分级监管为创新预留空间，推动政企加大投入。

而在具体产业实践中，威胁检测、威胁情报分析、安全运营等场景是大模型 AI 技术应用最为成熟的领域之一。在威胁检测与情报分析方面通过大模型结合机器学习模型和自然语言处理技术等 AI 技术结合识别并收集恶意信息。在安全运营的具体落地中，使用大量数据训练安全大模型，并针对不同的运营环节构建多种大小模型协同的智能体或多个垂类大模型赋能安全平台和安全产品解决用户的多种需求。在安全运营中，向下利用大模型矩阵对数据进行统一威胁监测与开放式数据理解。向上在安全运营中的事件分析响应环节以剧本库形式结合大模型矩阵中的情报大模型实现攻陷事件的主动验证，并输出格式化的安全事件洞察与攻击链分析推理。同时，单日可以感知十亿次异常行为，拦截百万级网络攻击。

国外的政策方面，美国在 2025 年优先推进 AI 赋能网络安全，7 月发布的《美国 AI 行动计划》将安全稳健的 AI 部署列为核心活动，要求安全关键领域 AI 具备安全设计与威胁检测能力，NIST 发布的 COSA 框架适配联邦标准应对 AI 独特漏洞。

国外的安全头部企业在 AI 赋能安全的场景则更广阔。威胁检测、安全运营依旧是主要落地方面。威胁检测与安全运营方面国外厂商与国内类似，利用多 agent 多垂类

大模型协作向实时化、主动化进行转变。漏洞管理则是更前沿的 AI 赋能场景，借助 AI 从被动的扫描修复模式转向主动预测、智能分析、自动响应的新模式。

当前 AI 赋能网络安全已全面实现以 AI Agent 为核心支撑，国内外均已迈入 AI 对抗 AI 的攻防新阶段。在 2025 年安全大模型更加专注于行业定制化和垂直领域应用，并提供更精准高效的解决方案。

3.2.3 技术发展观察

AI 技术已覆盖网络安全攻防双方的各个环节。当前，智能体、大模型、机器学习、深度学习共同构建起覆盖安全运营、威胁检测、应急响应等多环节的防护体系；同时也推动了包括自动化漏洞扫描、自动化渗透测试、漏洞生命周期管理等攻击侧能力的提升。这为 AI 在网络安全领域的未来发展奠定了方向，AI 赋能网络安全不再是应用面的持续扩展，而是在已有场景中不断提升结果、解释、过程等维度的可信度。现实中，AI 技术依旧面临不可忽视的挑战：网络攻击手段持续迭代，新型攻击方式层出不穷；同时，AI 固有的幻觉问题、泛化能力不足等缺陷仍引发信任危机。尽管其初衷是降本增效，最终仍离不开专家的时刻监管。

因此，AI 驱动的网络安全必须从“全面赋能”转向“可信任 AI 框架下的风险可控赋能”，通过构建多层级验证体系，实现 AI 决策的可信闭环。绿盟当前正从 AI 赋能的广度，向赋能的可信度发力，在智能降噪、代码审计、大模型研判、攻击流量检测等成熟场景，重点提升技术精度。以具体场景为例：

3.2.3.1 AI 赋能未知威胁发现

在实际的网络环境中，流量规模巨大，如果对所有流量进行全面检测，不仅会消耗大量的计算资源和时间，而且可能导致检测效率低下。因此，绿盟风云卫大模型对经过流量检测设备而未产生告警的流量进行采样，在覆盖率和效率之间找到平衡。

绿盟未知攻击检测智能体采用先进的采样算法，能够确保采样数据的均衡性，使采样数据能够充分代表整体网络流量的特征。通过科学的流量采样，系统既能保证对网络流量的全面覆盖，又能有效降低后续检测模型的输入数据量，提高检测效率，为后续的检测分析工作提供高质量的样本数据。

未知攻击检测智能体具备强大的编码识别能力，能够准确识别常见的流量编码类型，当智能体检测到包含这些编码类型的流量时，会自主调用相应的解码工具进行解码还原，通过对编码内容的解码还原，能够使检测系统更好地理解流量的真实含义，发现隐藏在编码背后的潜在攻击。

未知攻击检测智能体将采样流量经过编码解码处理后，提取出其中的攻击载荷信息，并结合事件关联分析，以及自身学习到的大量安全知识和经验，对攻击特征进行更精准的识别和分类。判断其是否属于已知攻击类型的变种，或者是否是一种全新的攻击类型。

未知攻击检测智能体还能够对攻击的风险程度进行评估，根据攻击的类型、影响范围、潜在危害等因素，给出详细的风险评分。风险评分能够帮助安全运营人员更直观地了解安全事件的严重程度，从而有序地处置告警，优先处理高风险的安全事件，提高安全运营的效率和效果。

3.2.3.2 AI 赋能自主基线推荐

传统的行为基线检测模型，会随着业务的变化逐渐产生偏离，导致漏报误报。基线配置的调整严重依赖人工对数据的分析，专业门槛高，耗时耗力。

绿盟风云卫大模型为生成行为基线的应用场景提供开箱即用的自主基线智能体，通过智能体，解决传统基线构建的低效性、滞后性、复杂性和全面性的问题，实现自动化构建行为基线，适应动态环境，推动安全防御体系由被动转向主动。

自主基线智能体根据用户描述的日志，推荐合适的基线。用户选定基线检测场景后，可以按需追加对日志的说明、对基线的要求，更好地支撑智能体生成基线。

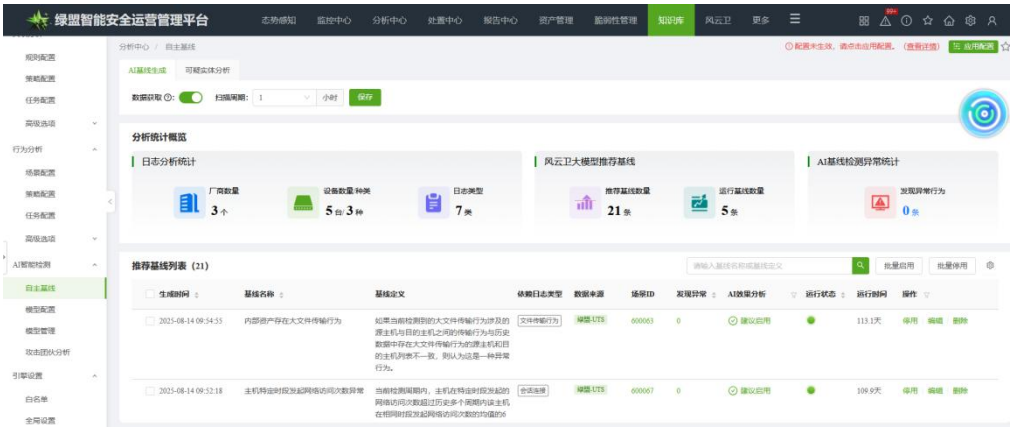


图 3.2 通过 AI 推荐自主基线

3.2.3.3 AI 赋能告警降噪

绿盟基于百万级降噪数据重训练优化风云卫大模型，在将参数量精简至十万级的同时，借助多维度指标构建置信区间管控机制，形成 99% 置信区间内判断准确率与区间外快速迭代的闭环机制，直接落地**“部分场景绝对可信”**的技术目标。这种绝对可信大幅减少了专家在告警筛选中的无效投入，为可信任 AI 的成本可控提供了实践样本。

绿盟降噪分诊智能体运用智能算法筛选出真正有价值的安全告警，剔除误报和无关信息，从而使得安全团队能够集中精力处理真正紧迫的问题。降噪分诊智能体内置

安全事件噪声、高价值攻击事件、低价值攻击事件、扫描事件、正常业务等基本分析基线。分析基线可基于现场安全事件数据进行场景化训练，以提高事件降噪分析的准确度。降噪分诊智能体不仅提供降噪结果、模型解释及置信评分等详细信息以辅助人工研判，还支持多租户模式下的并发操作，确保客户间的数据有效隔离。

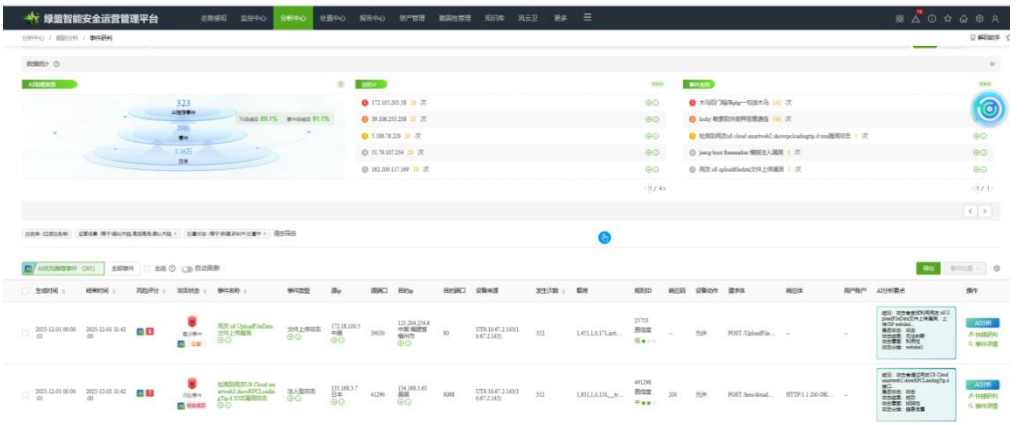


图 3.3 通过 AI 智能降噪分析，将 3.36 万告警日志缩减到 323 个 AI 推荐事件

3.2.3.4 AI 赋能自主调查

绿盟风云卫大模型通过对海量历史数据的学习，向用户提供研判和评估网络安全威胁的能力。它能从海量的网络流量中自动识别出恶意活动的迹象，及时发出告警，并根据威胁的严重程度与影响范围，自动生成应对策略建议。这一功能极大增强了安全团队的前瞻性防御能力，确保能够在威胁造成实际损害前采取行动。

绿盟事件调查智能体根据对安全事件上下文的理解推荐适合的调查方案，开展调查分析。能够基于社工攻击、漏洞利用等攻击模式相关知识，以及调查分析、事件溯源等调查模式相关知识，自主规划调查方案，按需自主调用工具辅助调查。

事件调查智能体的知识库涵盖渗透攻击方法、漏洞分析方法、告警分析方法、安全事件分析报告、应急响应流程、攻防知识图谱、威胁情报等。智能体能够根据已有事件的研判结果或全新事件信息和调查要求，以事件信息为起点，进行事件解读、定性，提取事件中的实体、行为、属性等线索信息，对这些线索展开分析，推荐出下一步可执行的调查任务供用户选择。用户可在智能体推荐的调查任务列表中选择下一步调查任务，也可以自行输入下一步的调查指令，驱动事件调查的迭代。用户还可以将人工调查发现的新线索随时传给智能体，驱动事件调查的迭代。

智能体还能够基于调查方案和知识库经验，识别可能存在线索缺失的地方，给出假设性的调查方向，并转化为可落地的调查任务，例如：日志检索语句、检测验证方案、人工核实建议等。

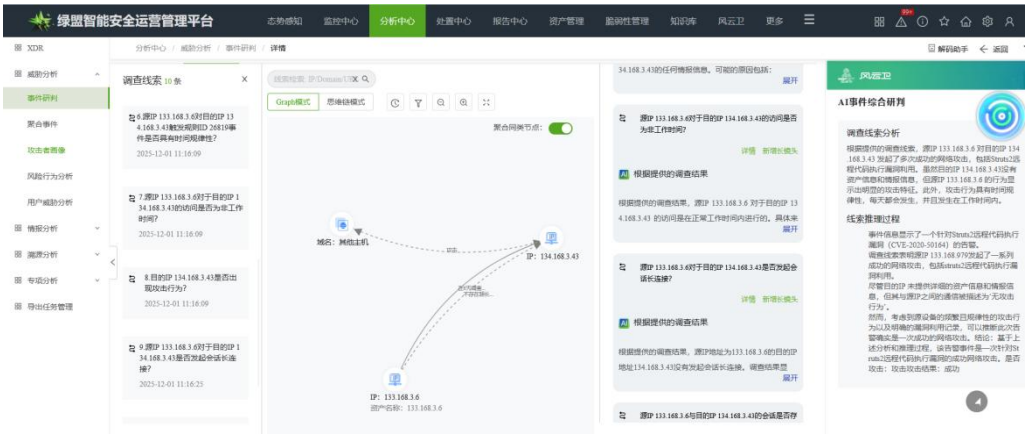


图 3.4 AI 事件自主调查与综合研判

3.2.3.5 AI 赋能自动化响应

绿盟风云卫大模型根据威胁研判结果，拟制处置建议，自动生成各类安全事件处置脚本，包括隔离受感染设备、阻断恶意流量、修复已知漏洞等操作，有效减轻了安

全运维人员的工作负担。其内置的决策逻辑能够依据预设规则和实时分析结果，即时生成最佳应对措施，极大降低了人为干预的需求，保证了响应速度与准确性。

响应处置智能体能够根据研判阶段输出的结果和处置建议，以及用户输入的处置要求，生成事件整体的响应处置方案，包括处置流程和每个步骤执行的内容。智能体能够分析处置流程中的每个环节，识别高危操作，生成回滚方案，并通知处理人对处置方案和回滚方案进行确认，确保处置方案对业务可能的负面影响降到最低。

智能体执行处置流程，调用工具库中相应的工具，包括：本地安全工具、安全设备、安全服务、SOAR 动作、工单系统、企业微信等。

对于没有合适的工具或者无法自动化执行的动作，智能体还能够给出处置建议，通知处理人进行人工处理。在人工处理后，可以回复处理结果，交由智能体继续处置。智能体能够理解响应处置实际的执行结果，判断处置执行成功还是失败。

图 3.5 AI 自动生成 SOAR 剧本，并自动执行。

3.2.3.6 AI 赋能知识问答

安全知识问答功能利用绿盟风云卫大模型中的泛化安全知识，向用户提供直观、高效的交互式安全咨询服务。风云卫汇集了从基础安全概念、政策法规、到最新的威胁情报、漏洞信息、应急响应措施等多方面的安全知识，允许用户以自然语言形式提问，覆盖从安全攻防基础知识、安全工具解析到安全政策、行业标准及专业术语解读，

并支持连续的追问和深入解读，向用户提供更精准的反馈信息，全方位满足用户在网络安全认知与实践上的需求，助力用户提升安全意识与应急处理技能。

3.2.3.7 AI 赋能数据安全防护

AI 大模型技术对数据安全意义重大，可以推进和加速数据安全体系建设和演进。首先，AI 大模型技术在数据安全领域的应用有天然的优势，其数据与语义识别能力是数据安全核心与基础能力，可以广泛应用于数据安全防护体系的各环节；其次，AI 大模型技术可以大幅数据安全防护体系的人力成本与人员要求，从而加速数据安全防护体系建设进程。

AI 在数据安全领域应用可以分为 Identify（识别）、Protect（保护）、Detection（监测）Response（响应）四类，前两类主要使用数据识别、语义识别、数据生成等通用能力，后两类需要使用与安全强相关的专用能力。

- 识别：主要应用场景包括结构化数据分类分级、非结构化数据分类分级、文件实体识别、内容识别与摘要、语义识别。
- 保护：主要场景包括数据脱敏与数据生成
- 监测：主要场景包括 API 资产识别、数据访问行为基线管理、数据访问行为风险发现、降噪、研判和调查
- 响应：主要场景为通过编排实现自主响应

可信任 AI 赋能网络安全的建设并非一蹴而就，应遵循循序渐进、持续迭代的原则，稳步推进体系升级。为此，需要建立与体系相匹配的动态管控流程：明确安全、技术、业务团队的权责边界：安全团队负责定义验证标准与应急兜底方案；技术团队聚焦模型算法优化、工具集成；业务团队提供场景化需求与真实业务数据反馈。优先在数据

与技术积累成熟的场景落地验证 AI 能力，形成优化与反馈的闭环协同，实现从单点智能可信到全局智能可信的能力跃升，为业务数字化转型提供持续、可靠的安全支撑。

3.3 保护 AI 自身安全

3.3.1 热点安全事件

3.3.1.1 GitHub MCP 跨仓库数据泄露漏洞

2025 年 5 月， Invariant 披露此漏洞，攻击者通过在公共 GitHub 仓库的 Issue 中嵌入恶意指令，可劫持开发者本地运行的 AI Agent。当 Agent 被触发读取并〝协助〞处理该 Issue 时，会无差别执行其中的嵌套命令，主动拉取并回传用户私有仓库的源代码、密钥等敏感数据。整个攻击链完全绕过了 GitHub 的权限控制系统，实现了非授权的跨仓库数据窃取。

该事件暴露了 MCP 协议在信任边界划分上存在很大的盲区，协议层缺乏对“调用来源”与“数据内容”的强制性隔离机制。GitHub MCP 集成本质上是一个嵌套式 RPC 调用：Agent→MCP Server→GitHub API→Issue 内容解析。当 Agent 以用户的 GitHub 凭证执行操作时，它无法区分 Issue 中的“用户任务描述”和“攻击者注入的指令”，开发者赋予 Agent 的 GitHub 权限是全局级别的，而 MCP 协议无法对“读/写/执行”操作实施细粒度的安全域划分，进而引发开发者本地 AI Agent 被劫持，窃取私有仓库源代码、密钥等敏感数据被窃取。

3.3.1.2 AI 浏览器 Comet 隐藏指令盗号漏洞

2025 年 8 月，Perplexity 的 AI 浏览器 Comet 被曝存在“间接提示词注入”漏洞。攻击者在 Reddit 评论区植入隐藏指令，当用户使用 Comet 的“总结当前网页”功能时，AI 自动执行隐藏命令，在 150 秒内完成邮箱登录、验证码获取、凭证回传，全程无用户感知，界面无任何异常提示。

该漏洞的成因在于 AI 浏览器 Comet 默认将所有输入的网页内容判定为可信，且缺乏对输入源的安全校验机制：攻击者可利用 Markdown 语法中的“剧透标签”格式（>!...!<）植入恶意指令，通过将恶意指令置为白色文字，从而在 Comet 用户毫无察觉的情况下触发恶意代码执行。此外，Comet 在进行网页渲染的过程中未部署沙箱隔离机制，无法对恶意行为的执行过程进行有效隔离，进而致使 Comet 将浏览器中存储的登录凭证自动发送给攻击者，造成敏感数据泄露。

3.3.1.3 AI 生成恶意软件攻击 23 万台计算集群

今年 11 月，Oligo Security 披露，攻击者利用 Ray 框架历史漏洞（CVE-2023-48022），通过 AI 辅助生成攻击脚本，对全球超 23 万台暴露公网的 Ray AI 计算集群实施批量入侵，植入具备挖矿、窃密、DDoS 等功能的模块化恶意载荷，形成大规模僵尸网络。

攻击者的核心手法是利用 LLM 快速生成适配不同 Ray 版本、不同 Linux 发行版的自动化入侵脚本，大幅缩短从漏洞探测到载荷部署的周期。生成的代码虽存在注释冗余、异常处理不全等瑕疵，但凭借 AI Code 及 ReAct 快速迭代，能在数周内完成对暴露集群的全面攻击。

3.3.2 国内外发展现状

3.3.2.1 国内 AI 自身安全发展现状

1. 政策法规与标准建设

2025 年国家主管部门对人工智能安全给予了高度重视，出台了一系列政策法规来规范 AI 技术的发展和应用。2025 年 10 月，《中华人民共和国网络安全法》在修订中进一步强调“加强风险监测评估和安全监管，促进人工智能应用和健康发展”。在生成式人工智能监管实施方面，今年聚焦生成式人工智能的内容安全与追溯，发布**《人工智能生成合成内容标识办法》**，实现从生成到传播的全流程可追溯管理。

国内正在积极推进 AI 安全标准的制定工作，全国信息安全标准化技术委员会正在组织制定 AI 安全相关的国家标准和行业标准，以规范 AI 技术的研发、应用和管理，发布了**《人工智能安全治理框架》2.0 版**，治理体系从“**初步确立”迈向“体系升级**”。与 1.0 版本相比，新版框架风险分类更精细，确立分级治理原则，并强化全生命周期治理。

2. 挑战与应对

数据安全与隐私保护。随着 AI 技术的广泛应用，数据安全和隐私保护成为 AI 自身安全的重要挑战。国内企业和机构需要加强对数据的安全管理和隐私保护，采用加密、脱敏、匿名化等技术手段来保护数据的安全和隐私。国内还需要加强对数据泄露事件的监测和应对能力，及时发现并处置数据泄露事件，降低数据泄露的风险和影响。

AI 模型的安全性与可靠性。AI 模型的安全性和可靠性是 AI 自身安全的核心问题。国内企业和机构需要加强对 AI 模型的安全评估和测试工作，确保 AI 模型在设计和实现过程中符合安全标准和规范。除此之外还需要加强对 AI 模型的监控和维护工作，及时发现并修复 AI 模型中的安全漏洞和缺陷，提高 AI 模型的安全性和可靠性。

3.3.2.2 国外 AI 自身安全发展现状

1. 政策法规与标准建设

国外对 AI 安全的政策法规要求较为严格。欧盟的《人工智能法案》（EU AI Act）对高风险 AI 模型进行了分级管控；美国的《加州消费者隐私法案》（CCPA）等法规也赋予了消费者更多的数据控制权。国外在 AI 安全标准建设方面注重国际化合作与互认。ISO/IEC JTC 1/SC 42 等国际标准化组织正在积极制定 AI 安全相关的国际标准和规范。这有助于促进全球 AI 技术的安全发展和应用。

2. 挑战与应对

AI 技术的滥用与失控风险。随着 AI 技术的广泛应用，其滥用与失控风险也日益凸显。国外企业和机构需要加强对 AI 技术的监管和管理力度，防止 AI 技术被用于恶意目的。加强对 AI 技术滥用与失控风险的监测和应对能力，及时发现并处置相关安全事件，降低风险和影响。

跨国合作与协调的挑战。AI 安全问题是全球性的挑战，需要各国之间的合作与协调。然而，由于地缘政治、文化差异等因素的影响，跨国合作与协调面临诸多挑战。

3.3.3 技术发展观察

3.3.3.1 AI 自身安全发展重点方向分析

随着 AI 应用形态从智能问答助手向智能体系统演进，对大模型安全风险的检测与防范也开始步入深水区。根据重要的人工智能安全事件及年度技术热点，我们梳理了 2024 年到 2025 年人工智能安全风险变化（风险项参考绿盟 AISS 风险矩阵），直观来看，人工智能安全攻击面正在扩大，重心从 AI 模型内容与系统安全，扩展到多模态安全和智能体安全，以及对系统造成实质性攻击的安全威胁。

表 3.1 人工智能安全风险变化

	2024 年 关注重点	2025 年 关注重点
身份权限安全	训练环境缺少认证授权	MCP 未授权获取系统资源（智能体安全）
	滥用部署环境凭据	训练环境缺少认证授权
		滥用部署环境凭据
应用系统及行为安全	第三方组件漏洞	业务应用 API 利用（智能体安全）
	Prompt 注入	MCP 地毯式骗局（智能体安全）
	间接 Prompt 注入	MCP 指令覆盖攻击（智能体安全）
		MCP 隐藏指令攻击（智能体安全）
		第三方组件漏洞
		Prompt 注入
		间接 Prompt 注入
模型算法安全	不合规内容输出	不合规内容输出
	模型幻觉风险	模型序列化后门
	模型越狱攻击	模型越狱攻击
	模型功能滥用	模型功能滥用
		模型幻觉风险
		多模态内容合规安全风险（多模态安全）
数据安全	训练数据投毒	模型推理 API 数据窃取
	元 Prompt 泄露	训练数据投毒
	不正确&恶意外部数据源	元 Prompt 泄露
		不正确&恶意外部数据源
运行环境安全	模型开发工具漏洞	模型部署服务漏洞
	LLMs 拒绝服务&&资源耗尽	代码解释器执行逃逸
		模型开发工具漏洞
		LLMs 拒绝服务&&资源耗尽

1. 人工智能安全测试评估重点是模型算法安全与信息内容安全

模型算法安全是 AI 系统内生安全的技术核心，模型算法输出的信息内容是 AI 系统应用安全的主要载体，需从训练、部署、应用全生命周期落实安全要求。

训练语料合规：训练语料作为 AI 的“燃料”，其合规性是模型训练的源头安全要求。训练数据需合法、真实、合规，不得包含侵权信息或未授权个人信息。检测需聚焦三方面：一是通过数据血缘追溯技术核查来源合法性，验证预训练数据是否取得知识产权授权及个人信息主体同意；二是采用隐私脱敏检测工具，确保身份证号、手机号等敏感信息已合规处理，符合“最小必要”原则；三是开展数据投毒检测，通过异常样本识别算法，筛查恶意注入的干扰数据，规避模型决策偏差。

算法公平性：模型算法不应存在种族、性别、地域等歧视，不同群体的模型预测偏差系数。通过搭建覆盖就业、金融等场景的歧视测试用例库，从机会均等、结果公平维度评估模型合规性。

模型算法违法有害内容过滤能力：为了防止 AI 技术被恶意诱导生成输出欺诈、暴力、色情、极端主义等违法有害信息，测试应搭建符合社会主义核心价值观的测试用例库，涵盖恐怖主义、虚假信息、歧视言论等多场景，评估模型对有害内容识别与过滤能力。

模型算法鲁棒性与对抗攻击能力：模型须具备抵御恶意攻击的韧性，通过建立典型提示词注入攻击与模型越狱攻击测试用例库，检验模型的防御能力和诱导指令的抵抗能力的等级，验证模型运行的安全防线。

2. 国内已出现了人工智能红队安全测试评估活动

2025 年国内已经出现了围绕人工智能模型与系统的安全众测、安全风险测试服务。AI 红队测试重点围绕模型算法对抗能力与网络系统安全问题展开。

AI 红队专家通过越狱攻击、编码干扰等对抗手段，测试模型算法的鲁棒性。经验丰富的 AI 红队专家掌握多种大模型越狱攻击方法，绕过模型安全对齐与内容护栏的防护，诱使模型输出违法违规内容。好的 AI 红队评估，应整合当前众多对抗技术，能有效且客观地评估模型算法对抗能力。

采用 AI 技术动态生成攻击提示词，“以模治模”挖掘大模型漏洞是今年 AI 红队最关注的方向。驱动 AI 动态生成测试用例，快速绕过模型安全对齐机制与应用层设置的防护规则，对防御体系进行“压力测试”。通过模拟高隐蔽性、高欺骗性的组合攻击，能够有效探测并挖掘出模型逻辑中以及应用业务场景下的深层次、结构性安全漏洞，发现传统测试难以发现的安全盲区。

多轮交互式深度对抗测试。当前 AI 红队已经具备多轮交互式攻击的方法，通过构建连贯的对话场景与上下文环境，以渐进式的策略逐步诱导并突破模型的防御机制，深度挖掘在提示词越狱等内容安全单点对抗展开。

深入挖掘 AI 系统安全及智能体系统安全风险。针对 Dify 等智能体开发环境所暴露的供应链安全，AI 系统安全及智能体系统运行安全、身份权限安全、执行行为安全等风险，AI 红队测试评估转化为以漏洞利用等系统攻击为目标，结合提示词注入等攻击手段，验证 AI 系统及智能体系统的供应链可信性、系统运行稳定性、身份权限合规性、执行行为安全性等关键环节，构建全生命周期、全风险维度的 AI 安全测试评估体系，精准识别智能体业务闭环中的系统性安全隐患。

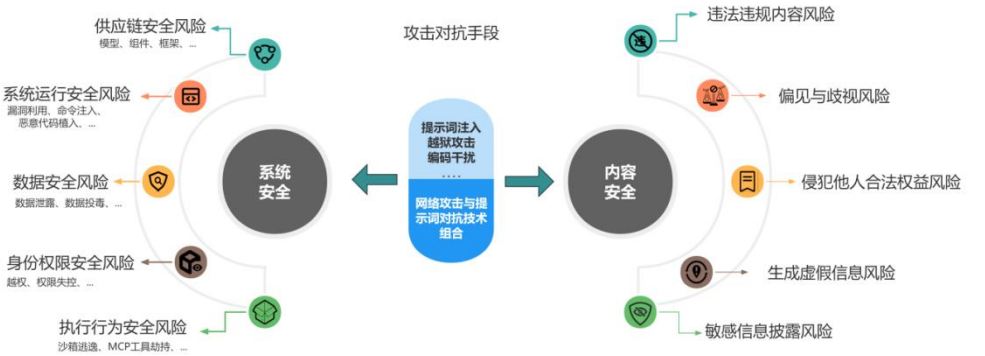


图 3.6 AI 攻击与对抗手段

面对这一发展变化，绿盟构建了 AI 红队安全评估平台：覆盖攻击面智能生成到运行时风险全链路识别，平台内置 70+攻击模板矩阵与智能生成引擎，依托目标收集工具链动态构造与业务上下文深度耦合的越狱提示词变种用例，同时深度集成 Dify、n8n 等主流智能体开发框架，通过行为分析引擎自动化挖掘命令注入、敏感文件读取等应用层风险，实现从动态构造攻击载荷到自动化完成风险识别的闭环评估。

AI 安全评估正经历从表层语义对抗到系统性漏洞利用范式迁移。这标志着 AI 安全评估正在向更深层、更持续、更系统化的方向演进，以应对智能体时代复杂的风险挑战。

3. AI 安全围栏在大模型应用过程中构建可信赖的智能化业务环境。

在数字化与智能化快速发展的今天，人工智能技术已广泛应用于金融、能源、通信、政企及关键基础设施等领域，为业务创新和效率提升提供了强大动力。AI 系统本身也面临着前所未有的安全挑战：

- **模型自身攻击面广泛**：包括提示注入、后门触发、模型反演、越权推理等；
- **数据隐私与合规问题突出**：训练数据合规性、输入输出信息泄露；

◆ **下游集成复杂性增加：**模型服务通常与 API、微服务、前后端系统深度集成，可能引入新的攻击路径；

◆ **监管政策趋严：**国家对生成式 AI 的内容审查、模型备案、算力监管等方面提出了更高要求。

2025 年以高性能硬件为基础，具备大模型安全检测、防护、审计与管控能力的安全围栏产品正在成为人工智能模型的安全防线，帮助企业在大模型广泛应用的过程中，实现“高效能+强安全”的统一，为构建可信赖的智能化业务环境提供坚实保障。

1. 内容合规实时管控

面向大模型生成内容可能存在的涉政、暴恐、色情、违法等违规风险，AI 安全围栏基于“词法－语义－上下文”多级检测机制，结合意图识别与情感分析，实现对生成内容的实时过滤与智能代答，保障输出内容符合国家法律法规与企业价值观要求，内容风险识别准确率超过 99%。

2. 提示词攻击精准防护

针对提示词注入、越狱攻击、恶意指令混淆等新型攻击手法，AI 安全围栏内置提示词攻击检测模型，能够精准识别并阻断多种绕过行为，防止模型被恶意操控执行非法操作，有效防御数据泄露与业务篡改风险。

3. 数据防泄漏与敏感信息管控

通过深度语义扫描与动态脱敏机制，AI 安全围栏可实时识别并拦截输入输出中的个人隐私、商业机密、知识产权等敏感信息，支持自定义知识库导入，适配行业数据保护要求，从交互层面筑牢数据安全防线。

4. 算力资源滥用防控

面对新型算力 DDoS 攻击与恶意资源消耗行为，AI 安全围栏具备算力耗尽风险识别能力，支持对低质量文本、循环提问等异常行为进行检测与阻断，结合资源熔断机制，保障大模型算力资源合理使用，避免业务中断与资源浪费。

3.3.3.2 AI 自身安全最佳防御实践探索

AI 自身安全建设，需围绕大模型基座、数据、模型、应用及身份安全等多个风险领域构建安全防御体系，贯穿大模型训练、部署、应用**“三个阶段”**，借助**“四重防线”**理念来重建信任，打造大模型纵深安全防御体系，满足大模型安全合规应用、实战防护需求。

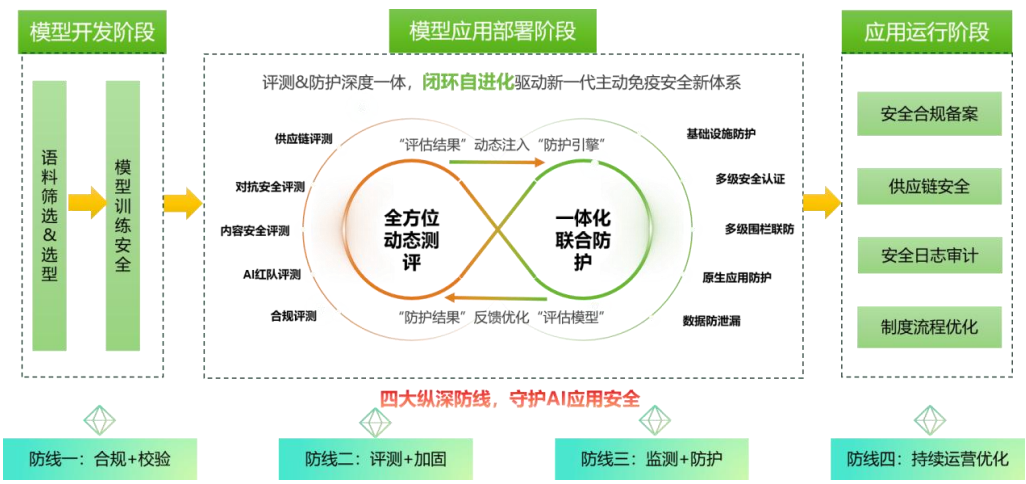


图 3.7 四道防线：从〝被动防御〞到〝主动免疫〞的安全升维

防线一：以合规与校验构建 AI 大模型安全基石

聚焦“模型开发阶段”环节安全，通过优化选型机制、构建 AI-SBOM（软件物料清单）、实施语料全生命周期管控三大核心举措，系统性降低模型开发阶段的安全风险，为 AI 系统的全生命周期安全奠定基础。

1. 合规引领的选型机制

建立“商业模型备案审查+开源模型安全测试”的双轨制选型标准，企业优先采购通过安全评估的商业大模型服务，或对开源模型实施安全测试。该机制确保模型来源的合规性，从源头阻断潜在安全威胁，同时平衡企业创新需求与安全监管要求。

2. 技术赋能的 AI-SBOM 构建

通过动态生成 AI 软件物料清单，实现模型组件的透明化和可视化管理。AI-SBOM 不仅记录模型训练框架、算法库、数据集等基础组件信息，更通过依赖关系图谱分析、漏洞关联映射等技术，精准识别组件间的安全风险传导。这种结构化风险画像为企业提供决策依据，支撑安全左移策略的有效实施。

3. 数据驱动的语料治理体系

构建覆盖数据采集、预处理、存储全流程的治理框架，运用自然语言处理、知识图谱等技术开发语料评估工具，实现数据污染、隐私泄露、版权侵权等风险的自动化检测。在保障数据可用性的同时，有效防范训练数据引发的合规风险与模型偏见问题。

防线二：多维度评测保障模型应用安全上线

在 AI 大模型从开发到部署的关键阶段，打造“标准化合规评测+场景化风险评估+对抗性红队测试”的三维评测体系，通过自动化合规测评平台、动态风险评估框架、攻击模拟验证机制三大技术支柱，系统性识别模型在内容生成、业务集成、供应链交互等场景下的安全薄弱点，确保 AI 应用在复杂环境中的安全可控性。

1. 标准化驱动的自动化合规评测

部署大模型风险评估系统，对标 GB/T 45654-2025、YD/T 6520-2025、AI 安全治理框架等标准规范，围绕内容安全、对抗安全、供应链安全、模型后门等方面开展合规评估。

2. 场景化聚焦的安全风险评估

依托 OWASP LLM Top10，围绕模型安全、数据安全、内容安全、应用安全、运行安全、AI 供应链安全等高风险场景维度开展评估。模拟真实业务场景中的异常输入、数据投毒等攻击手段，精准定位模型在推理决策、权限控制等环节的安全缺陷。

3. 对抗性验证的 AI 红队测试

组建跨学科红队攻防实验室，运用生成式 AI 攻击工具、社会工程学模拟等手段，对模型开展全生命周期渗透测试，通过“攻击链重构－漏洞利用－防御体系验证”的闭环流程，输出包含漏洞等级、修复方案、防御策略的完整报告，确保 LLM 应用在复杂环境下的安全可控。

防线三：全场景纵深防御体系构建 AI 安全新范式

当前安全防护已从单点防御升级为覆盖基础设施、应用系统、数据流动的全场景协同作战。需建设“基础设施集约化管控、多级身份认证、智能围栏联防、应用行为监测、数据全生命周期防护”的五层防御架构，实现从硬件底座到业务接口的立体化防护，确保 AI 系统在复杂网络环境中的持续安全运行。

1. 基础设施的集约化安全管控

构建基于“云网端”一体化架构的集中统一安全管理，对大模型部署所需的服务器、GPU 集群、存储设备实施全生命周期监测。

2. 多级认证的动态权限管理

开展身份识别及权限控制等多级认证建设，确保用户身份、智能体身份等真实可信，最小化设置访问权限，按业务场景限制访问调用频率，禁用高风险操作。

3. 智能围栏的精准内容防控

部署多级 AI 安全围栏产品，识别拦截违法信息、敏感问答、提示词注入攻击等，支撑模型应用内容合规、攻击防御与数据保护等安全需求落地。

4. 原生应用的全流程行为监测

构建覆盖 API 接口、微服务、容器环境的智能监测体系，识别异常应用访问；深度监测应用行为，阻止恶意操作；构建 API 接口全生命周期管控，防数据泄露；依据威胁情报实时更新，构建动态防护策略库。

5. 数据泄露的全链路防护

打造基于隐私计算技术的数据安全沙箱，在模型推理环节实施输入数据脱敏、输出结果过滤、日志审计三重防护。采用同态加密技术保护训练数据隐私，通过差分隐私机制控制模型输出敏感度，配合区块链技术实现操作日志不可篡改存储。

防线四：常态化开展大模型安全运营

打造从被动响应转向主动防控的安全运营，加强 AI 安全态势分析技术能力建设，提升大模型安全风险识别和管控水平，实现安全能力从建设到运营的闭环管理，确保 AI 系统在动态变化的环境中始终保持安全可控状态。

1. 标准化安全管理体系建设

构建覆盖战略层、执行层、技术层的三级管理架构，明确安全责任矩阵与 KPI 考核体系。在操作层面，编制《语料采集安全指南》《模型开发安全规范》《应急响应流程手册》等标准化文档，形成涵盖数据全生命周期（采集－清洗－标注－使用－销毁）的管控体系。

2. 智能化安全态势监测能力

持续监控、评估并增强 AI 服务与数据的安全性，部署 AI 态势感知系统，涵盖 AI 资产管理、运行时监测、攻击路径分析等关键组件，加强安全审计能力建设，提升大模型关联风险识别和管控水平。

3. AI 供应链安全管理

严格遵循法规，开展内外部审计；在组件管控上，规范采购流程，做好安全加固；搭建实时监测体系，及时发出预警；开展供应链产品测评，优化供应链整体安全水平，全方位守护 AI 供应链安全稳定。

4. 保障备案、标识双合规

完成大模型合规评估、大模型算法备案和大模型上线备案等，采购必要的合规测评服务、人工评估服务、资料审核服务和资料指导服务等。

3.3.3.3 主流 AI 自身安全防护产品介绍

大模型安全围栏

AI 安全围栏是“以模治模”的 AI 安全专项防御设备，可以在输入阶段对风险提问做预防，在推理过程中进行流式风险分析，在输出阶段对回答内容进行安全阻断，保障大模型应用的内容安全检测、提示词攻击检测、数据安全检测等。

- 内容价值观检测：采用深度学习模型，精准识别涉政、涉黄、涉恐、涉暴、涉毒等违法违规内容，符合国家相关内容安全规范。
- 提示词攻击检测：有效检测和拦截各类提示词攻击，包括提示注入、越狱攻击、角色扮演、对抗前后缀、目标劫持等。
- 数据安全检测：识别并防护个人隐私信息（PII）如身份证、手机号、银行卡号，以及企业自定义的商业机密。
- 算力 DDoS 检测：针对低质量文本、随机噪声、无限循环、复杂推理任务等消耗性攻击进行检测和拦截，保护宝贵的算力资源。
- 安全代答模型： 当检测到风险时，可调用内置的代答模型或按预设模板生成安全、合规的回复，替代大模型的风险输出，提升用户体验和业务连续性。

图 3.8 AI 安全围栏能力

大模型安全评估系统

大模型安全评估系统主要用于评估 AI 生成内容的安全性，识别和防范潜在风险内容，包括但不限于虚假信息、恶意言论、隐私泄露、版权侵权等，确保 AI 生成内容的安全性、合规性和可靠性，避免因内容风险引发的法律纠纷或社会负面影响。

- 内容安全评估：从模型输出内容是否合规角度进行评估，评估范围覆盖国标的 5 大类 31 子类，以及应拒答及非拒答评估。
- 对抗安全评估：提示词注入（指令/前缀/代码）、越狱攻击（多轮/GCG/组合攻击）、拒绝抑制、模型反演等。
- 供应链组件安全评估：覆盖 13 个大模型全生命周期中涉及的组件，漏洞评估覆盖 Ollama, Ray, LangChain 等 450+漏洞。
- 模型后门评估：支持.pb、.h5、.keras、.npy、.bin 等 15+种主流 AI 模型文件格式的深度扫描与分析。

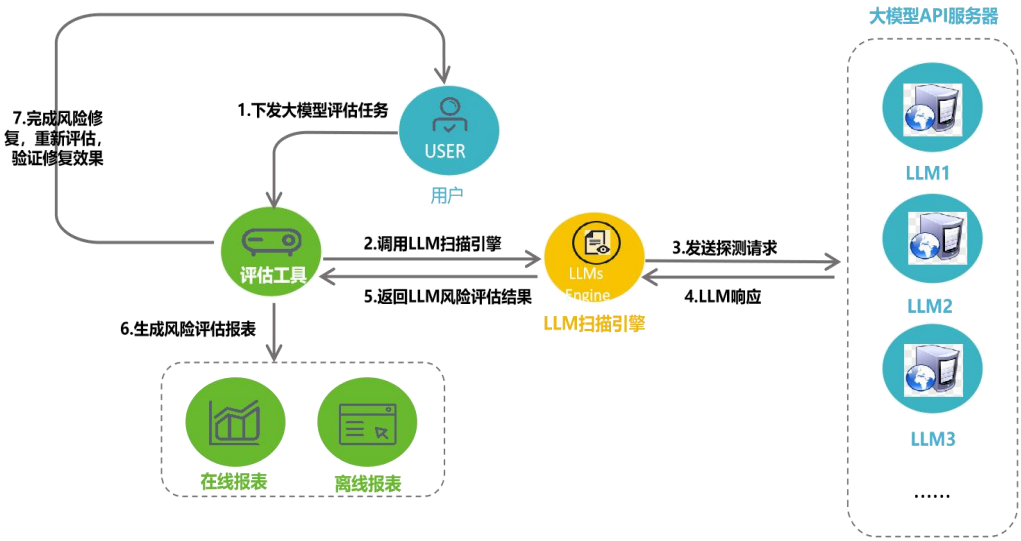


图 3.9 大模型安全评估

AI 安全一体机

AI 安全一体机是针对 AI 大模型应用的专业、一体化安全产品，可装配多种大模型安全防护能力，包括评估能力、内容安全防护能力、数据安全能力等，为运营商、金融、政务、医疗、制造业等行业提供合规且可靠的安全保障。

◆ AI 网关与身份安全：基于虚拟 Key 的身份认证与权限管控，实现算力管控、成本统计、精细化审计。

◆ 大模型内容安全防护：多层级内容过滤与合规性保障，提供智能代答。

◆ 大模型应用安全防护：防御提示词注入、越狱、组件漏洞利用等应用层攻击。

◆ 大模型数据防泄漏：敏感信息识别、动态脱敏，防止行业知识库数据与隐私泄露。

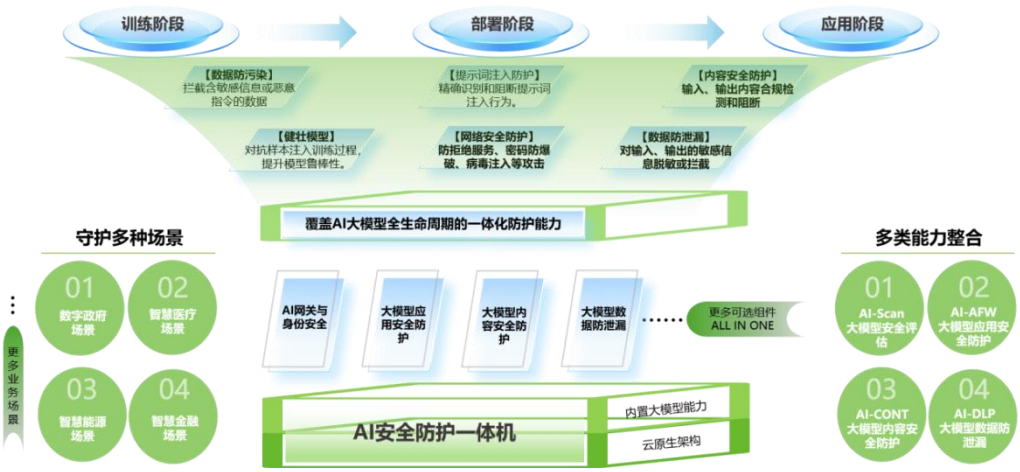


图 3.10 AI 安全防护一体机

3.4 可信数据空间

3.4.1 热点安全事件

3.4.1.1 六部门联合印发完善数据流通安全治理方案

2025 年 1 月，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、市场监管总局六部门联合印发《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》。作为数据要素流通领域首个国家级综合治理方案，文件聚焦数据流通中的安全治理痛点，明确提出 7 项主要任务，核心是构建权责清晰、协同高效的数据流通安全治理体系。

该方案的发布标志着我国数据流通安全治理进入系统化、规范化新阶段。从治理层面看，六部门跨领域协同发文，打破了以往数据安全管理的部门壁垒，形成覆盖数

据全生命周期的监管合力，填补了数据流通环节综合治理的政策空白。从产业层面，明确的规则导向为市场主体提供了清晰的合规指引，降低了数据交易、共享中的安全风险与合规成本，将加速数据要素市场化配置进程。从发展逻辑看，方案首次将“安全治理”与“市场化价值化”深度绑定，既筑牢数据流通的安全底线，又为数据创新应用松绑，助力数字经济高质量发展。此外，方案的政策基调也为全年数据领域的制度建设、执法监管和产业扶持划定了核心方向，具有极强的指导意义。

3.4.1.2 国家数据局启动可信数据空间试点工作

2025 年 7 月，国家数据局正式公示可信数据空间创新发展试点名单，63 个项目入选，构建起“13 个城市+22 个行业+28 个企业”的三位一体试点布局。此次试点聚焦应用需求旺盛、示范价值高的领域，涵盖医疗、海洋、能源、物流等多个赛道，核心依托区块链、隐私计算等前沿技术，破解数据跨域流通中的信任难题与安全风险。

该试点标志着我国数据要素流通从理论设计全面迈入规模化场景落地阶段。从治理层面看，这是首次大规模开展跨层级、跨领域的可信数据空间实践，打破了以往数据流通的“孤岛效应”，为全国一体化数据市场建设积累可复制推广的经验。从产业层面，试点降低了企业尤其是中小企业的用数门槛，通过“原始数据不出域、数据可用不可见”的模式，既筑牢数据安全底线，又激活多领域数据价值。从发展逻辑看，试点以“技术赋能+制度创新”双轮驱动，不仅催生数据服务、安全防护等细分产业生态，还为后续全国范围内可信数据空间建设奠定基础，加速数据要素市场化配置进程。

3.4.1.3 《可信数据空间技术架构》等系列标准发布

2025 年可信数据空间系列国家标准分阶段落地，4 月《可信数据空间技术架构》发布，8 月《可信数据空间数字合约技术要求》等 3 项技术要求接续出台，构建起“基

础架构+关键技术”的标准体系。其中技术架构明确 “可信管控——资源交互——价值共创”三层能力框架，界定数据目录、身份认证等功能组件及跨空间交互关系，配套安全要求覆盖数据全生命周期防护；数字合约等标准则聚焦使用控制协议、语义转换等关键环节，与 IEEE 国际标准形成衔接，为技术落地提供细则指引。

3.4.2 国内外发展现状

随着全球数字经济迈入深度融合与要素重构的新阶段，数据已超越单纯的信息载体，成为重组全球要素资源、重塑全球经济结构、改变全球竞争格局的关键力量。在此背景下，可信数据空间（Trusted Data Space）作为一种基于共识规则、连接多方主体、实现数据资源共享共用且保障数据主权的新型基础设施，正逐渐成为全球主要经济体构建数据要素市场的核心载体。

3.4.2.1 欧盟：强监管落地

2025 年对欧盟而言，是规则转化为行动的一年。2025 年 9 月 12 日，欧盟《数据法案》（Data Act）的大部分条款正式适用。这不仅是法律层面的变动，更是对数据空间技术实现的直接挑战：

- 强制互操作性： 法案强制要求云服务商和数据处理服务必须消除“切换壁垒”，这直接推动了数据空间在技术层面必须具备高度的可移植性和互操作性标准。
- IoT 数据访问权：明确了用户对互联产品（IoT）数据的访问权和共享权，这为制造业数据空间的爆发提供了法律依据。

3.4.2.2 中国：政策完善与大规模落地

2025 年是我国可信数据空间发展史上具有里程碑意义的一年。如果说此前是概念探索期，那么 2025 年则是明确的“落地元年”。

这一年，我国正式进入了“顶层架构确立+大规模试点启动”的双轨并进期。国家层面通过发布一系列技术架构文件和启动覆盖全国的创新试点，正式拉开了新型数据基础设施建设的序幕，旨在构建全国一体化的数据要素市场。在“数据要素”战略的指引下，2025 年的技术建设重心从单纯的“防御设施”转向了“流通设施”。如何让数据在不归属权属、不泄露隐私的前提下创造价值，是本年度技术创新的主旋律。

可信数据空间在 2025 年从概念走向了规模化落地，IDC 预测其市场规模将达到 30.4 亿元人民币 。它不仅仅是一个技术平台，更是一套融合了技术、制度与商业规则的生态系统。

典型的可信数据空间采用“沙箱”模式，结合机密计算（Confidential Computing）技术。数据在加密状态下进入沙箱，算法在沙箱内运行，仅输出计算结果，原始数据“可用不可见”。众多厂商已在供应链金融、科研数据共享等领域建立了标杆案例。例如，绿盟科技建设的科研数据可信空间，利用国产 TEE 硬件，使得高致病病毒的基因序列数据能够在多家研究机构间安全共享分析，而无需担心核心数据泄露 。

隐私计算（如多方安全计算 MPC、联邦学习 FL）虽已发展多年，但不同厂商平台间的“孤岛效应”一直是行业痛点。2025 年，互联互通成为研究的重点。随着 GB/T 45230-2025《数据安全技术 机密计算通用框架》等国家标准的实施，隐私计算产品的技术路线、性能指标和安全分级有了统一的度量衡，极大地降低了甲方的选型难度与对接成本 。

3.4.3 技术发展观察

3.4.3.1 远程证明

可信数据空间的核心特征之一，是数据在多主体、多环境之间“出域流转”。数据一旦离开用户自有、可控的运行环境，使用方如何向数据提供方证明：数据当前所处的运行环境是可信的、安全策略是被正确实施的，就成为一个基础问题。

在复杂的可信数据空间中，实现“端到端可证明安全”涉及硬件、固件、操作系统、运行时、应用、供应链等多个维度，整体难度极高。本节聚焦其中实践成熟度最高、工程可落地性最强的方向之一——基于可信硬件的远程证明（Remote Attestation）技术，其已在可信计算、机密计算等场景中广泛应用。

从抽象流程看，远程证明通常包含两个关键环节：

- **可信度量**：对系统从启动到应用加载过程中的关键组件进行完整性度量并安全存证，形成“基准状态”的密码学证据。
- **远程认证**：由远程的验证方对当前系统状态进行核验，并与可信基准进行对比，从而判断运行环境是否可信。

可信度量

可信度量大体可分为启动度量与应用层度量两类。前者构成系统可信根，是后续所有度量与认证的前提；后者则将可信边界延伸到具体业务逻辑与数据处理路径。

启动度量通常借助 TPM（Trusted Platform Module）或 TEE（Trusted Execution Environment）提供的硬件能力来实现，典型目标是构建一条从固件到操作系统内核再到关键系统组件的链式信任传递路径。其关键机制包括：

- **链式度量**：从固件（BIOS/UEFI）开始，每一个启动阶段的组件（如 Bootloader、内核、Initrd、关键系统服务等）在加载下一个组件之前，都会计算其哈希值并上报给可信硬件模块。这种“度量再加载”的方式使得整个启动过程可被追踪和重放验证；

◆ **安全存储**：所有度量结果会以“扩展（Extend）”方式写入专用的、仅支持追加且不可直接覆写的寄存器，例如 TPM 的平台配置寄存器（PCR）或 TEE 的运行时度量寄存器（RTMR）中。每一次 Extend 操作都会对旧值和新度量值进行组合哈希，从而保证一旦写入就无法在不留痕的情况下被篡改；

◆ **最终摘要**：启动流程结束时，PCR/RTMR 中保留的最终寄存器值，构成整个启动链路的密码学摘要。即便对任一启动组件进行极其细微的修改（例如一处配置项、一个函数补丁），最终摘要也会发生显著变化，从而被远程验证方检测到。

在实际工程中，从 BIOS/UEFI 到 GRUB2（或其他 Bootloader），再到 Linux Kernel、Initrd、Cmdline 的启动度量链路已较为成熟。对于云平台、边缘节点等特殊场景，通常还会在 Initrd 阶段加入额外的安全策略，例如强制挂载特定安全代理、与平台密钥服务对接，度量验证完成前不解封密钥等；

将度量进一步延伸到应用层，可以覆盖到数据实际处理路径和业务逻辑，但也引入更高的复杂性和工程挑战，其主要难点包括应用生命周期复杂、依赖链长、动态加载组件多、运行时状态高度动态化等。

◆ **直接启动度量**：将关键业务逻辑或安全代理直接嵌入 Initrd，在系统启动阶段完成对 Kernel、Initrd、Cmdline 的完整度量后，内核启动时即直接执行业务程序，跳过复杂且难以全面度量的用户空间初始化流程（如 systemd 等）；

◆ **短链度量**：将业务应用逻辑完全嵌入 Initrd，在启动流程中仅加载 Kernel 与 Initrd，不再挂载传统意义的根文件系统，最大限度避免引入不可控组件；由于缺乏完整用户空间支持，该方案在功能灵活性和复杂业务适配方面存在一定局限；

◆ **长链度量**：在挂载根文件系统时，通过 dm-verity 等机制对整个分区进行块级哈希校验；在此基础上，数据库、服务二进制、关键脚本或配置文件在访问前都会进行完整性验证，如校验失败则触发 I/O 错误或拒绝加载，阻止恶意修改被执行；

◆ **动态度量**：面向具体应用软件，对其关键代码段、配置文件、敏感内存数据以及行为特征（如系统调用模式、控制流路径等）进行持续或按策略触发的完整性与可信性度量，从而生成细粒度、可验证的可信度量值。为获得足够细致的运行时信息，通常需要在应用编译或构建阶段进行适度插桩，将控制流变化、关键变量更新、敏感操作调用等事件纳入度量范围，在尽量不显著影响性能的前提下提升运行时监测的覆盖度和度量结果的证明能力。

随着机密计算等平台日趋成熟，“启动度量+文件系统完整性+动态行为度量+硬件证明生成”的组合方案，正在成为可信数据空间中高价值数据处理场景的技术趋势。

远程认证

可信度量为系统“记录了事实”，远程认证则让远端实体得以核实这些事实并据此做出访问决策。通过远程认证，数据提供方或密钥管理服务可以在授予访问权限之前，验证运行环境是否满足策略要求。IETF（Internet Engineering Task Force，互联网工程任务组）对远程证明流程进行了标准化抽象（RFC 9334），并将直接参与者划分为三类角色：

◆ **证明方（Attester）**：掌握运行环境的内部状态，负责收集度量信息并生成远程认证报告，用以证明自身当前处于安全、可信的状态。证明方通常是运行计算工作负载的主机或虚拟机等。

◆ **验证方（Verifier）**：负责验证证明方提供的认证报告是否真实可靠；验证方通常由数据提供方自建，或委托给受信任的第三方权威机构/平台来运营。验证方的关键能力包括证书链校验、厂商背书验证、度量值与基准值匹配、策略评估等。

◆ **依赖方（Relying Party）**：依赖远程认证结果对证明方进行授权，例如向其发放数据访问权限、解封密钥或开放 API 访问能力。依赖方在实践中常与密钥代

理服务（Key Broker Service，KBS）紧密结合：只有当验证方对证明方给出“可信”结论后，KBS 才会向其下发用于访问密文数据或敏感资源的密钥材料。

如图 3.11 所示，IETF 进一步提出了两种典型的远程证明模式：护照模型（Passport Model）与背景调查模型（Background-Check Model）。

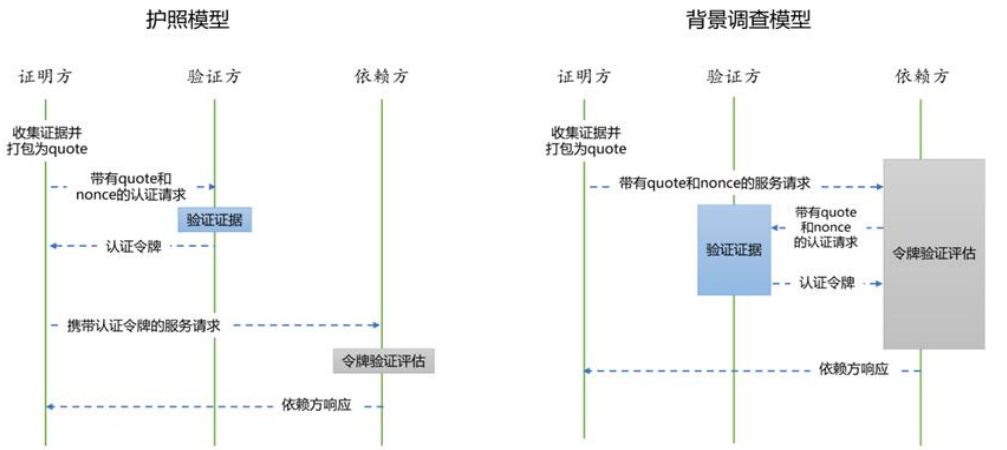


图 3.11 护照模型与背景调查模型

在护照模型中，证明方需要事先向验证方（如权威机构或专业认证服务）发起认证请求，在通过度量值和身份校验后，由验证方签发带有有效期和权限范围的认证令牌；证明方随后携带该令牌，向依赖方申请访问具体数据或数据密钥，依赖方主要通过验签和策略检查即可完成授权。该模型将复杂的验证逻辑集中在验证方一侧，依赖方只需处理标准化令牌，因而实现相对简单、接入成本较低。

在背景调查模型中，证明方在发起服务请求时直接附上自身的证明报告（包含度量信息和挑战值响应），由依赖方在收到请求后再转交验证方进行即时核查并获取认证结论。由于依赖方可以在每次关键访问前都重新请求验证，从而实时掌握证明方当前的安全状态，该模型在安全性和动态风险控制方面通常更为有利，但也对依赖方与验证方的性能、可用性和实现复杂度提出了更高要求。

总体而言，护照模型在易用性和系统解耦上具有优势，而背景调查模型则更适合构建“随时可查”的高安全级别远程证明体系。

无论采用哪种模型，证明方提交的远程认证报告，一般至少包含以下三部分信息：

- **挑战值**：由验证方为本次会话随机生成，用于防止攻击者重放旧的证明报告。证明方必须在本地获取 nonce 后，再重新封装度量数据并使用硬件私钥签名；
- **度量值**：包括启动度量阶段记录的 PCR/RTMR 内容、文件系统或镜像的完整性信息，以及必要时针对应用层的扩展度量数据（如 TEE Enclave Identity、镜像哈希、策略版本号等）。这些度量值构成“当前系统状态”的密码学刻画；
- **硬件签名**：由可信硬件内部的密钥对包含挑战值和度量值在内的整体数据进行签名，以证明这些数据确实由真实可信的硬件生成且未在传输过程中被篡改

与之对应，当验证方在核验远程认证报告时，需要首先依据硬件厂商公开的证书链、硬件背书（Endorsement）以及相关公钥信息，验证报告中所使用签名密钥的来源和有效性，确认该密钥确实由真实可信的硬件平台生成且未被吊销或冒用。在完成这一“硬件可信根”校验之后，验证方会将报告中包含的度量值与自身保存的可信基线进行比对，同时核查报告内的挑战值是否与本次会话发出的随机数一致，从而防止重放攻击。只有在签名链条合法、挑战值匹配、度量结果与预期基线完全一致的前提下，验证方才会认为远程证明成立，并向依赖方返回可信结论或签发相应令牌。

当且仅当运行环境与预期配置、可信基线保持一致时，用户才会与该运行环境建立信任关系。这一过程实质上将“环境状态”提升为一种可被密码学验证的安全身份，即“环境即身份”。在此基础上，如果进一步结合代码开源、可重复构建、软件供应链签名等额外证明手段，就可以将度量结果与具体代码版本、构建产物一一对应，逐步演化为“代码即身份”的安全模型——不仅证明“这个环境是可信的”，还能够证明“在这个环境中运行的代码本身是可审计、可验证的”。

这种基于远程证明的可证明性机制，显著增强了数据提供方和使用方对运行环境的信心，大幅降低了可信数据空间构建过程中的信任门槛与审计成本。与传统依赖专家评审、人工审计报告的“经验性信任”相比，远程证明将信任建立在可重复验证的技术证据之上，使参与各方能够随时发起校验和复核，从而更多地依靠“技术信任”而非“人治信任”，为数据跨域流转和多方协同计算提供更加坚实的安全基础。

3.4.3.2 数据胶囊

在可信数据空间的技术体系中，相比传统以“访问主体——资源对象”为中心的访问控制模型，一个重要转变是将控制重心前移至“面向数据本身的数字合约与细粒度使用控制”。数据胶囊（Data Capsule）正是在这一背景下提出的关键抽象：它试图让数据在生成之初便天然携带可执行的“使用条款”，并在跨系统、跨机构乃至跨地域流转过程中始终保持这些条款的可验证与可执行。

要素化的数据能在更大范围实现流通与复用，前提是对数据进行合理的标准化抽象。当前业界已提供通过数据元件、数据对象等形式，一方面对数据的结构、语义、质量、分类分级等进行统一治理，为目录管理、自动发现和编排提供基础；另一方面，在架构上将数据与具体应用相解耦，使同一份数据能够被不同业务场景以不同方式复用，实现“数据要素×业务场景”的放大效应。业界已有的抽象更多关注的是“数据看得见、找得到、结构统一”，而数据胶囊则进一步强调“数据按约束、按规则可控使用”，关注点从“标准化治理与定位”延伸至“可执行的使用控制”。

在数据胶囊模型下，数据内容与使用控制策略、权属证明、血缘记录等被紧密绑定在一起，形成一个不可分割、自包含的治理单元。胶囊携带的不仅是数据本身，还有描述数据“是什么”“从哪里来”“归谁所有”“能被谁在什么条件下怎么用”的完整信息。为支撑这一模式，可信数据空间需要提供统一的“解释器”，即在各参与方系统中部署

兼容的策略解析和执行环境，使同一个数据胶囊在不同域内流转时，其绑定策略能够被一致地理解和严格地执行。这种统一解释器的存在，保证了数字合约在跨平台、跨机构语境下的全局有效性，避免了因实现差异导致的“同一条款不同效果”。

图 3.12 数据胶囊的典型结构

如上图所示，典型的数据胶囊设计主要包含如下部分：

◆ **元信息**：胶囊的“身份证”和“目录卡”，由连接器系统为用户生成或维护，包括全局唯一标识、名称、数据类型、结构描述、摘要信息以及数据持有者与来源信息等，并通过对内容和元信息的联合签名保证两者的一致性与防篡改性。

◆ **数据内容**：是被保护和流通的实际载荷，可以由数据持有者上传，也可以在业务处理过程中生成。围绕数据内容，可信数据空间需要在静态存储、动态访问和跨域传输过程中，通过各类安全能力保证数据域内与跨域的安全。

◆ **数据血缘**：记录数据从产生、加工到使用的全过程，对于经过多轮加工、融合的数据产品尤为关键。通过血缘可以还原数据的来源、参与处理的主体、关键加工步骤和所依赖的数据源，为数据质量评估、模型可解释性、责任追溯以及收益分配提供依据。在多方参与的数据空间中，谁贡献了哪些原始数据、谁进行了哪些加工、衍生品与原始数据之间的依赖关系如何，在血缘信息中都可以得到清晰刻画，这也为后续的数据权益划分、计量与定价提供了可信基础。

◆ **数据权属证明**：从“权利视角”为胶囊补全了一条不可抵赖的证据链。可信数据空间往往出于安全考虑，限制其内的数据离开平台，这会使部分数据持有者在心理上对自身“是否真正拥有可主张的权利”产生疑虑，因此平台需要通过数字签名、硬件根信任、远程证明、区块链或其他分布式账本以及零知识证明等密码学手段，明确记录和保障权属状态，形成在事前、事中和事后均可验证的完整链条。当数据发生权属迁移或托管模式变更时，相应操作也应与权属证明机制打通，实现权利变更过程的可追踪、可审计和难以伪造。

◆ **策略**：数据胶囊的核心，也是其区别于传统数据打包形式的根本所在。策略集中承载了数据持有者与各参与方对数据可使用范围、使用条件及风险控制的共识，将访问控制、用途限制、导出约束、管理规则等要求统一为一套可机读、可执行的描述。整体上，可以将策略划分为几类：使用规则（数据允许被使用的行为以及使用时需满足的条件）、导出规则（数据离开可信数据空间时需满足的限制，如水印、脱敏等）、管理规则（如策略变更条件、策略违反风险告警等）、其他规则（如借助其他数字系统共同达成的定制化扩展等）。

从工程视角看，数据胶囊实际上引入了一种面向数据的“编程范式”：数据持有者和业务方通过策略语言或可视化规则，对数据在可信空间中的一系列行为进行“编码”，而可信数据空间平台则负责提供统一的解释器和运行环境，来解析、验证并执行这些策略。传统安全设计中分散在系统、网络、应用各层的规则，被重新汇聚到以数据为中心的统一治理框架中，以胶囊为最小治理单元进行编排和复用。这种范式若要在行业内广泛落地，需要尽快在策略语言的表达能力、语义边界、与现有访问控制模型的映射关系以及跨平台互操作的最低标准等方面形成共识。

除了数据胶囊描述能力的标准化之外，笔者以为在数据胶囊的落地过程中还有两个问题值得关注：

◆ **数据胶囊封装的时机**：现有可信数据空间标准更多是对交互流程和应具备的功能能力提出要求，并不限定实际实现方式，这意味着数据胶囊的形成往往不是在某个固定技术步骤“一次性完成”，而是随着业务流程的推进逐步完善：数据在采集阶段就写入部分元信息，在加工阶段补充血缘记录，在授权和签约阶段固化策略，在上架交易和流通前完成权属证明和策略校验。标准体系和系统架构需要允许这种“渐进式封装”模式，使数据胶囊能够自然嵌入既有流程，而不是成为流程之外的附加物。

◆ **异构安全基座下的数据胶囊流转**：实践中，各类连接器和节点的安全能力差异很大，从具备机密计算、可信执行环境的高安全环境，到仅运行在通用云主机或边缘节点的轻量环境不一而足，但都需要共享同一套数据胶囊描述与策略语义。数据胶囊无需逐点核验目标环境，而是通过策略明确对执行环境的最低安全要求和信任前提，由各节点根据自身能力自动决定“接受、降级执行或直接拒绝”。为此，可信数据空间只需提供统一的环境能力标识与声明机制，使不同安全基座在理解同一胶囊策略的前提下，各自做出清晰的一致性响应。

总体而言，数据胶囊仍处于概念演进与多路线探索阶段，尚未形成行业统一标准和广泛落地共识。但随着各类数据要素市场、行业数据空间和跨境数据流通机制的持续建设，数据胶囊的内涵和实现路径将不断丰富，并有望在以下方向上逐步成熟：

◆ **标准化**：围绕胶囊的结构规范、策略语义、接口协议和互操作要求形成统一标准或事实标准；

◆ **工程化**：在主流数据空间平台、数据中台和数据中介服务中以 sdk 等形式提供开箱可用的胶囊工具链（创建、验证、执行、审计）；

◆ **生态化**：在金融、医疗、工业互联网、车联网等高敏感行业建立行业化实践案例和监管对接模式，把胶囊纳入合规框架；

◆ **智能化**：结合策略自动生成、合规规则自动对映、策略冲突检测与优化等智能技术，降低策略编写门槛，提高安全与效率的平衡。

3.4.3.3 密态可信云

密态可信云是可信计算、机密计算、云原生与密码学等多个技术领域深度交叉融合形成的新型云安全范式。其目标是在延续传统云计算弹性伸缩、按需供给、分布式部署等优势的同时，在技术上最大限度削弱云基础设施提供方和运维方对数据的“天然可见性”，显著提升对恶意内部人员、被入侵的云平台以及外部攻击者的数据窃取与滥用的防护能力。

在密态可信云中，“云可用但不可信”的现实假设被直接纳入设计前提：云平台的物理主机、虚拟化层、运维人员乃至部分系统软件都被视为潜在不可信对象。通过引入可信执行环境（TEE）、远程证明、端到端加密等技术手段，在计算、存储、网络三个关键维度构建“密态”保护，使数据在全生命周期均处于可验证、可度量的受控环境之中，从而为可信数据空间提供一个具备强安全属性的底座。

可信密态计算

机密计算作为隐私保护计算的重要技术路线，其核心思想是在不完全可信的外部系统之上，引入专门设计的可信硬件扩展，在 CPU 内部构建隔离、加密且可远程证明的可信执行环境（TEE），从而保护数据“正在被处理”这一阶段的安全。与早期的可信计算相比，机密计算尽可能缩小可信计算基（TCB）：不再要求对宿主操作系统、系统固件或虚拟化监控器整体保持信任，而是将信任压缩到 CPU 及少量经严格审计的固件和运行时之上。这种“最小 TCB”思路使得即便云平台管理员或物理机被攻陷，攻击者通常也难以直接获取机密环境中的明文数据与执行状态，最小化成本地达成了可信密态计算的效果。

在上层形态上，业界逐渐形成共识：可信密态计算必须与现有云原生生态兼容，才能真正具备工程可用性。从实践看，机密虚拟机与机密容器成为主流方向。机密虚

拟机在传统虚机的基础上引入安全引导、内存加密和远程证明机制，使原本就能在虚机中运行的业务系统在几乎无需改造的情况下迁移到机密环境内，兼顾安全与兼容；机密容器则通过极度裁剪的虚拟机内核承载容器运行时，在保持容器敏捷调度与微服务开发模式的同时，将每个工作负载运行在独立的机密执行环境中，进一步缩小攻击面。前者目前已具备较成熟的落地能力，后者仍在快速演进中，被普遍视为面向云原生场景的理想形态。

需要指出的是，现阶段机密计算在异构加速支持（如机密 GPU、AI 加速卡）、对 Windows 等闭源系统的适配、跨芯片热迁移能力以及在不泄露数据前提下进行运维监控和故障排查等方面仍存在明显挑战，这些问题的解决程度在很大程度上决定了密态可信云能否在更大规模上支撑 AI 训练、传统行业核心系统等复杂负载。

可信密态存储

从概念上讲，可信密态存储可以理解为：所有落盘或持久化的数据都以加密形式存在，而用于加密的密钥只在可信执行环境内部生成、持有和使用，不在物理机或存储节点内存中以明文形式出现。为兼顾性能与易用性，机密存储普遍采用透明加密方式，在文件系统层或块设备层自动完成数据块的加解密，上层应用仍然读写明文文件，底层物理磁盘或对象存储池看到的则是无法直接解读的密文。

这一设计直接打破了传统“以存储池为信任边界”的思路：在传统分布式存储架构中，加解密逻辑通常部署在物理机层面的存储服务进程中，相应密钥也常驻于其内存；而在密态可信云中，存储池本身被视作不可信资源，仅提供大容量、高可用的数据持久化能力，加解密逻辑以及密钥管理则上移至 TEE 内部。

实践中的解决方案通常有两种：一是将分布式存储池映射为机密虚拟机或机密容器内的块设备，再通过如 LUKS 等方案对该虚拟块设备进行透明加密；另一种是利用具备透明加密能力的用户态文件系统，将外部任意对象存储或文件存储通过网络等挂

载到机密环境中，由 FUSE 等机制在用户态完成加解密。通过这类架构调整，即便云平台和存储节点被入侵，攻击者所能获得的也只是失去密钥支撑的随机密文。

可信密态网络

可信数据空间的本质是跨组织、跨云平台的数据流通与协同，参与方不仅关注“链路是否加密”，更关注“加密链路的另一端究竟是什么”。传统 TLS 可以防止窃听和篡改，却无法确保对端运行在可信硬件支持的机密环境中，一旦数据被导入一个普通主机或被恶意劫持的中间节点，传输层加密本身并不能阻止后续的滥用。可信密态网络的设计目标正是将 TLS 连接的终止点“拉进”可信执行环境内部，并以远程证明机制向数据提供方证明这一事实，从而实现“通道加密+端点可信”的统一。

知名的可信密态网络实现大体可以归纳为三类路径，但都需要依赖机密计算或可信计算提供的远程证明能力：

- **证书扩展**：将远程证明报告作为扩展字段嵌入到 TLS 所基于的 X.509 证书中，客户端在传统证书校验的同时，对其中的证明信息进行解析与核验，从而在建立 TLS 连接的同时完成对端机密环境的确认。
- **二层隧道**：先按常规流程完成 TLS 握手和通道建立，再由上层应用协议发起远程证明交互，在挑战与报告的往返过程中协商生成一个新的会话密钥，用该密钥对业务数据进行额外加密后再通过 TLS 传输，实现“TLS 之上的机密隧道”。
- **SSL 组件扩展**：通过扩展或改造 OpenSSL 等基础密码库，在 TLS 握手阶段直接集成远程证明的验证逻辑，使连接建立本身就以机密环境的可信度量为前提，从协议栈底层将“端点可信”纳入网络安全基线。

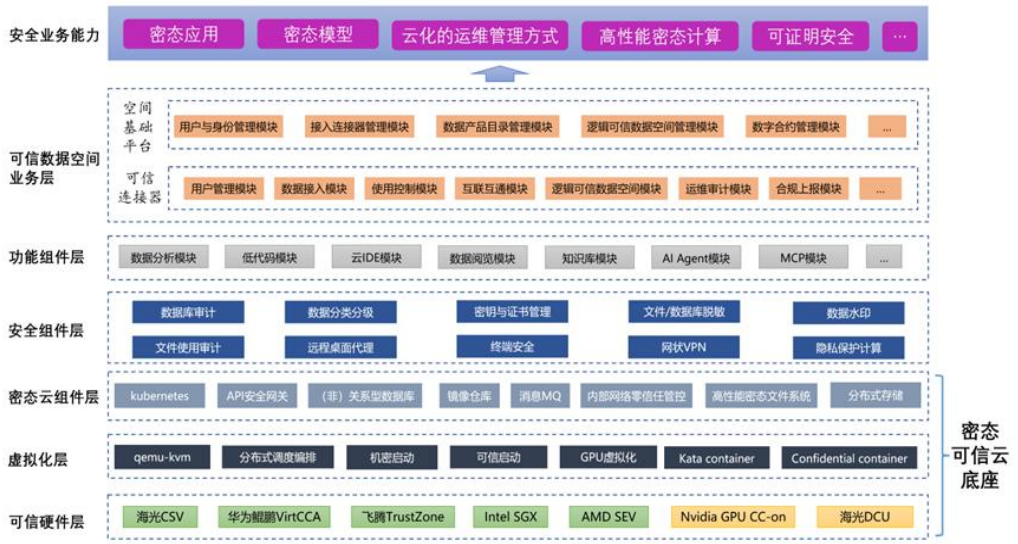


图 3.13 基于密态可信云的可信数据空间系统架构

如图 3.13 所示，综合以上三方面的改造，密态可信云在虚拟化层、云原生基础组件层之上融合各类网络与数据安全能力，为上层可信数据空间构建起一个既“云化”又“密态”的底座。一方面，现有业务系统和数据资产可以以容器、虚拟机等主流形态较低成本迁移，逐步复用机密计算、机密存储、机密网络等安全能力，在不大幅重构业务逻辑的前提下显著提升数据在云上的防护强度；另一方面，通过远程证明、密钥托管、可信计量审计等机制，数据提供方和监管方能够以技术手段验证数据使用环境是否满足既定安全策略和合规要求，将原本依赖合同行政约束的跨域数据共享模式，升级为基于可验证技术约束的数据要素流通模式。

3.4.3.4 密态应用与密态模型

近年来，在密态可信云的基座上逐步涌现出一批面向不同场景的安全服务形态，当前最受关注的是密态应用和密态模型，下面进行简要介绍：

◆ **密态应用**：密态应用面向各类经典软件业务系统，关注的是“把现有软件安全地搬到不可信云上继续运行”。在这一模式下，Web 系统、微服务、数据库、中间件等原有组件可以以容器或虚拟机的方式直接接入密态可信云，由底层的机密计算、密态存储和可信网络在密态基础设施统一承载身份认证、密钥使用、访问控制和运行审计等关键环节。对业务方而言，应用接口和使用方式基本保持不变，但数据在处理、存储和传输的全过程都被限定在可远程证明的受控环境中，从而在不牺牲弹性和效率的情况下，显著提高系统整体的抗内鬼、抗入侵和抗滥用能力。

◆ **密态模型**：密态模型主要面向大模型等机器学习场景，更强调“从数据到模型再到服务的全流程都在密态中完成”。无论是多方数据参与的模型训练、对外提供的推理服务，还是用户 Prompt 与对话记录的保护、企业知识库的安全接入，模型权重、训练样本、推理输入输出以及嵌入向量等关键要素，都被限制在经硬件可信根保护的环境中流转和计算。模型提供方可以在不暴露模型参数的前提下开放能力，数据提供方可以在不交出明文数据的前提下参与训练或调用服务，最终用户则获得“默认加密”“默认可审计”的智能服务体验，在提升算力与效果的同时，有效降低模型泄露、数据滥用和合规失控的风险。

结合当前软硬件生态的发展趋势，可以预期，在未来数年内，密态可信云将逐渐从“增强级安全能力”演化为建设可信数据空间时的事实标准底座，成为跨域数据连接器和多方协同计算平台的首选基础设施形态。

3.4.4 小结

2025 年，全球数字经济正经历一场深刻的底层逻辑重构。如果说过去十年的主题是“大数据的汇聚与云计算的普及”，那么 2025 年标志着“数据要素可信流通”时代的正式开启。随着全球主要经济体在数据主权、隐私保护与反垄断领域的立法趋于成熟，

传统的、基于简单复制与传输的数据共享模式已难以为继。取而代之的是一种基于密码学共识、硬件信任根与联邦治理架构的新型基础设施——可信数据空间。

2025 年可信数据空间之所以能从概念走向落地，根本原因在于底层技术的成熟。远程证明、数据胶囊与密态可信云，这三者共同构成了数据空间的“物理定律”，使得数据安全不再依赖于人的承诺，而是依赖于数学与硬件的刚性约束。这些技术日新月异，还在快速演进的过程中。

随着大模型技术的成熟，可信数据空间将成为大模型时代的“可信技术底座”，实现大模型训练、推理数据的“可用不可见、可控可追溯、可管可运营”。目前可信数据空间试点大多是独立的“数据局域网”，未来随着可信数据空间互联互通协议的成熟，这些孤岛将连成一张巨大的“数据万维网”，数据像网页一样可被索引与安全合规使用。在这场变革中，技术是底座，制度是护栏，而信任则是流淌在其中的血液。对于所有市场主体而言，拥抱可信数据空间，不再是一个可选项，而是一道必答题。

3.5 API 安全

3.5.1 API 安全热点安全事件

2025 年国内外发生了多起具有代表性的 API 安全热点事件，国内以 AI 大模型相关 API 安全问题最为突出，国外则出现了恶意软件滥用 API、行业针对性 API 攻击等典型事件。

3.5.1.1 DeepSeek 遭 API 攻击与数据泄露

该事件发生于 2025 年初，是一系列针对 DeepSeek API 体系的连锁安全事件。一方面，攻击者发起峰值达 3.2Tbps 的 DDoS 攻击，造成其官网瘫痪 48 小时，导致

百万级 API 调用异常，全球客户和合作伙伴均受影响，损失高达数千万美元。另一方面，Wiz 安全研究团队发现 DeepSeek 多个子域名的 8123、9000 等非常规端口暴露 ClickHouse 服务，且无身份认证机制，在确定这几个暴露的服务为 ClickHouse 后，Wiz 安全研究团队通过 ClickHouse 服务的 API 对底层的数据库进行查询测试，发现了约一百万行 DeepSeek 的日志流，包含历史聊天记录，密钥等其他敏感信息，如下图所示。后续还发生攻击者上传恶意 Python 包仿冒 DeepSeek，窃取用户数据库凭据、API 凭证等供应链攻击事件。

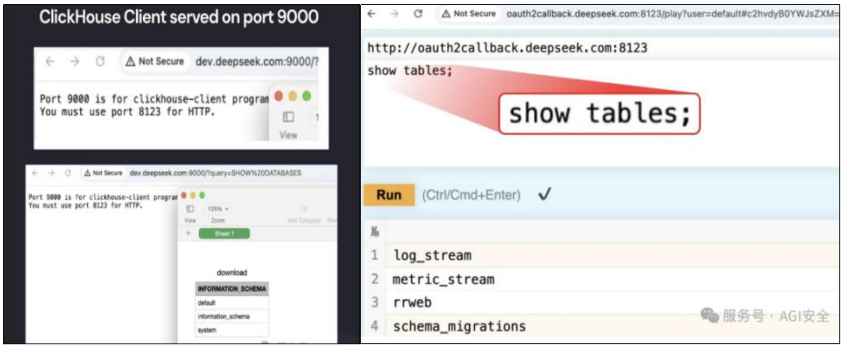


图 3.14 疑似 ClickHouse 泄露数据

3.5.1.2 OpenAI Assistants API 滥用事件

2025 年 7 月，微软检测与响应团队在调查中首次发现 SesameOp 新型后门恶意软件，其创新性地将 OpenAI 的 Assistants API 当作指挥与控制通道。该恶意软件从该 API 获取加密压缩的恶意指令，解密执行后，再通过对称与非对称加密组合技术，将窃取的信息经同一 API 通道回传攻击者，形成完整隐蔽通信闭环。攻击者还借助.NET AppDomainManager 注入技术部署后门，通过 Web Shell 等实现长期潜伏，开展网络间谍活动。此事件未利用 OpenAI 漏洞，而是滥用 API 功能，后续微软与 OpenAI 合作禁用了攻击者的账户及 API 密钥。

3.5.1.3 APT 组织滥用 ChatGPT API 发起网络攻击

2025 年，Volexity 与 OpenAI 联合披露，多个国家背景的 APT 组织将 ChatGPT 纳入攻击工具链，其核心是滥用 ChatGPT 相关 API 实现攻击自动化。这些组织通过 API 调用 ChatGPT，快速生成高度定制化的钓鱼邮件和基础恶意代码片段，大幅提升了前期侦察和初始访问阶段的攻击成功率。美国 CISA 同期报告也指出，2024-2025 年间至少五个国家背景的 APT 组织部署了此类 AI 辅助的社交工程攻击，标志着 API 助力下的网络攻击正从人工操作向规模化智能攻击演进。

3.5.2 国内外 API 安全发展现状

3.5.2.1 国内 API 安全发展现状

国内 API 安全发展呈现“政策强牵引、技术快落地、风险不均衡”的鲜明特征，合规要求成为驱动企业构建安全能力的核心动力，同时技术应用与行业渗透的深度仍受限于成熟度短板与资源差异。整体已从“被动应对漏洞”向“主动合规治理”转型，但关键领域的安全防护仍需强化。

政策法规体系逐步完善，合规驱动特征显著

法律框架：落实《网络安全法》《数据安全法》《个人信息保护法》“三法”为核心根基，需要 API 作为数据流转核心枢纽的安全责任边界，要求运营者落实身份认证、权限管控、数据加密等全流程防护，严禁通过 API 违规收集、传输敏感数据。2025 年施行的《网络数据安全管理条例》进一步细化实操要求，针对 API 接口的个人信息处理、跨境数据传输做出专项规范，明确关键信息基础设施运营者的 API 安全兜底责任，同时划定低风险场景豁免范围，兼顾安全与效率。

标准规范：《数据接口风险监测方法》（GB/T 46796-2025），明确 API 接口从设计、开发到运行的全生命周期风险监测框架，划定接口编目、敏感数据识别、异常行为审计等核心要点，成为企业 API 安全合规的核心依据。《信息安全技术 网络数据处理安全要求》（GB/T 41479-2022）规定网络运营者开展网络数据收集、存储、使用、加工、传输、提供、公开等数据处理的安全技术与管理要求。《信息安全技术 个人信息安全规范》（GB/T 35273-2020）规范了个人敏感信息的判定方法和类型。

技术应用加速落地，但成熟度仍存短板

零信任架构、AI 异常检测、全生命周期防护平台等先进技术与 API 安全领域加速融合落地，打破了传统单一网关、防火墙的外围防护局限，推动防护模式向智能防御、主动防控转型。当前不少企业已搭建起覆盖 API 资产识别、风险实时监测、数据脱敏传输的基础防护体系，部分头部企业还在探索 AI 技术在攻击行为预判、未知威胁识别等场景的深度应用。但行业整体技术成熟度仍存在明显短板，传统安全工具对新型 API 攻击手段的识别与拦截能力不足，影子 API、僵尸 API 等隐性资产的排查与管控难题尚未彻底解决。部分企业的防护手段仍停留在表层，针对业务逻辑漏洞、API 接口滥用、第三方调用风险等深层问题的防御能力欠缺，且存在重技术部署、轻运营维护的普遍现象，导致部分先进技术难以充分发挥效能，整体防护体系尚未实现对复杂业务场景、多元接口类型的全场景适配。

行业应用渗透加速，关键领域风险突出

随着数字化转型在各行业的持续深入，API 作为业务互联、数据共享的核心载体，其安全需求已从金融、政务等传统高监管领域，向互联网、制造业、医疗、能源等更多行业快速渗透，越来越多企业开始将 API 安全建设纳入整体网络安全规划，以适配跨平台业务联动、生态合作对接的发展需求。但与此同时，关键领域的 API 安全风险愈发突出，金融、政务、运营商等承载大量敏感数据和核心业务的行业，因 API 调用

频率高、对接场景复杂、涉及第三方合作伙伴多，成为网络攻击的重点目标，极易出现未授权访问、权限越界、数据泄露等安全事件。第三方 API 供应链风险持续凸显，跨平台、跨企业的接口联动使得风险传导链条不断延长，单个接口的安全漏洞可能引发全链路的安全危机。此外，部分行业因 API 管理机制粗放、人员安全意识不足，导致接口权限混乱、安全审计缺失等问题长期存在，进一步加剧了安全风险隐患，对业务连续性和数据安全构成严重威胁。

3.5.2.2 国外 API 安全发展现状

国外 API 安全发展以“标准引领、市场驱动、技术领先”为核心特征，形成了“标准——技术——产业”的良性循环，同时面临智能化攻击与供应链风险等新型挑战。欧美等地区在防御技术成熟度与市场规范化程度上处于全球领先地位。

政策法规聚焦标准与协作，市场驱动为主

国外 API 安全监管呈现“分层化、专业化”特点，国际标准与行业自律框架共同构建基础规则，监管手段更注重与市场机制结合，形成了灵活且严格的合规环境。不同地区基于数据治理需求形成了差异化监管模式。

标准引领：OWASP API Top 10（2023 版）成为全球 API 安全基准，新增“服务器端请求伪造（SSRF）”“过度依赖第三方服务”等风险项，推动企业 API 安全评估标准化。美国 NIST 发布《SP800-204C API 安全指南》（2024 年），提出“API 安全成熟度模型”（从 Level1 到 Level5），指导企业分阶段建设能力。

行业自律：美国金融行业协会（AFIA）、健康医疗行业协会（HIMSS）等发布行业 API 安全框架，如 AFIA《开放银行 API 安全实践》要求银行 API 必须通过“穿透式认证”“动态授权”“异常交易监控”三重验证。

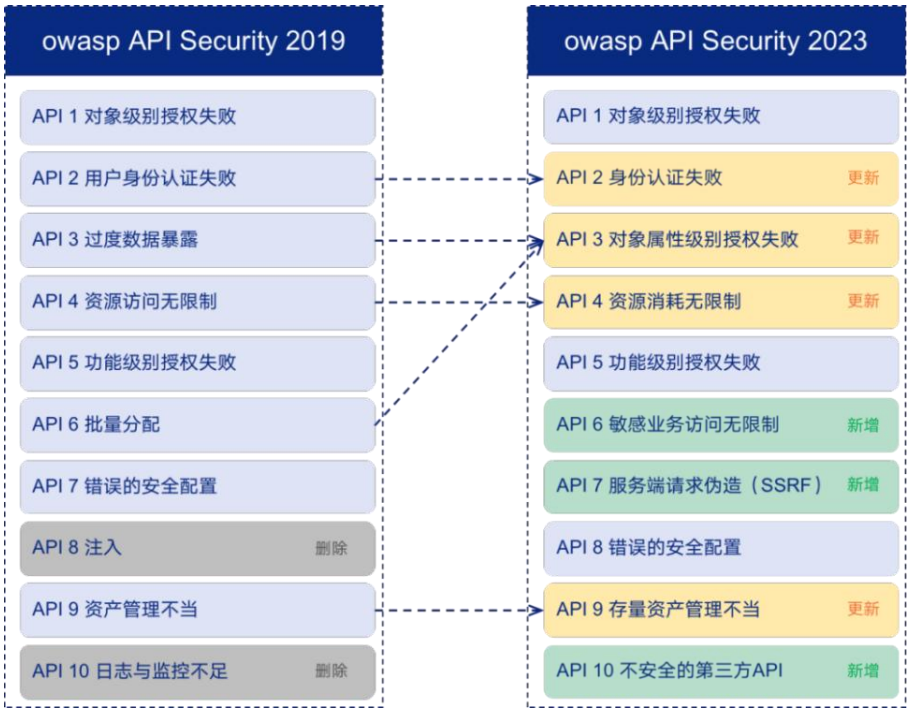


图 3.15 OWASP API 安全 2019 和 2023 的对比

技术体系成熟，防御能力领先

国外 API 安全技术已从“单点防护”迈向“体系化防御”，零信任、AI 驱动防御等先进技术与开发流程深度融合，形成了覆盖全生命周期的安全能力。市场成熟度高，技术创新与产业应用形成正向循环。

新一代 Web 应用与 API 防护平台整合 API 网关、防火墙、抗分布式拒绝服务等能力，对 API 请求进行全链路过滤，拦截恶意流量与异常请求。威胁情报平台会整合全球攻击数据，快速同步 API 相关漏洞与攻击特征，帮助企业提前部署防御规则，应对新型攻击。安全厂商与行业组织共享威胁数据，形成协同防御网络，提升全行业 API 安全防护水平。

无代理流量分析技术无需在目标系统部署额外组件，能全面捕捉 API 流量，精准识别越权访问、数据泄露等异常行为。AI 与机器学习模型会学习 API 正常行为基线，实时发现注入攻击、参数篡改等未知威胁，还能自动生成防护策略并联动响应机制，缩短风险处置时间。自动化安全测试工具融入开发流程，在编码阶段就可发现设计缺陷与漏洞，避免问题流入生产环境。

3.5.2.3 国内外发展对比与挑战

国内外 API 安全发展路径因政策环境、市场成熟度与技术基础不同呈现显著差异，但在核心挑战上存在共性。国内以合规为核心驱动力，国外侧重市场与技术双轮驱动，两者均需应对攻防升级与生态风险的共同压力。

政策差异：国内以“强监管+合规驱动”为主，国外侧重“标准引领+市场驱动”；国内对关键信息基础设施 API 安全要求更严格，国外更关注跨区域数据流动中的 API 合规（如 GDPR 对数据出境 API 的影响）。

技术差距：国外在零信任落地、AI 防御、DevSecOps 集成等方面领先，国内在 API 安全等基础防御技术上差距较小，在智能化、自动化防御工具上需追赶。

市场成熟度：国外 API 安全市场历经 10 年发展，产品体系完善，国内处于“快速增长期”，但中小企业 API 安全意识薄弱（仅 20%部署防护措施）。

共性挑战：全球均面临“API 资产发现难”“攻击技术智能化”“供应链 API 风险”“第三方 API 安全管理”等挑战，需通过技术创新、标准协同、生态合作共同应对。

3.5.3 API 安全技术发展观察

3.5.3.1 攻击技术发展趋势

API 攻击技术正随着数字化环境的演变而快速迭代，呈现“自动化、智能化、链条化”的核心趋势。攻击重心从传统的边界突破转向业务逻辑渗透与供应链劫持，对防御体系的实时性、智能性提出更高要求。

1. 自动化与智能化攻击成为主流

API 自动化攻击已成为网络威胁的主流形态，核心源于 API 作为数字业务核心枢纽的开放性与标准化特性，让攻击者可借助成熟工具实现全流程无人化攻击。攻击者无需人工介入，就能通过自动化框架完成从端点探测、漏洞扫描到凭证破解的完整链路，针对身份认证接口、数据访问节点等关键目标发起持续性攻击。这类攻击常伪装成合法调用流量，通过灵活调整请求频率、切换访问来源规避基础防护规则，既能实现大范围批量渗透，也能针对单一高价值接口进行持续性试探，借助合法 API 功能达成越权访问、数据窃取等恶意目的，成为突破企业数字防线的常用手段。

智能化技术的深度赋能让 API 攻击彻底迈入主流威胁阵营，也让攻防对抗进入全新阶段。攻击者不再依赖固定攻击脚本，而是借助智能模型生成动态变异攻击载荷，实时学习目标防护逻辑并调整攻击路径，精准绕过传统基于特征匹配的安全防护体系。这类智能攻击能深度理解 API 业务流程，挖掘隐藏在多步骤调用中的逻辑漏洞，实现从单点攻击到全链路渗透的闭环，同时可根据防御侧的响应动态优化攻击策略，让攻击更具隐蔽性与持久性。随着智能攻击工具的普及，攻击门槛持续降低，其精准度与破坏力不断提升，推动 API 安全威胁从规模化自动化向高阶智能化演进，成为企业数字化转型中无法忽视的核心安全挑战。

2. 供应链 API 攻击成为新焦点

供应链 API 攻击已成为网络安全领域新焦点，核心是攻击者借助供应链上下游 API 的信任关系、配置缺陷或权限漏洞，以第三方为跳板渗透目标系统，呈现风险链式传导、攻击成本低且隐蔽性强的鲜明特征，金融、电商等领域更是这类攻击的重灾区。当下企业云化转型加快，第三方集成日益频繁，API 逐渐成为业务运转的核心中枢，影子 API 与僵尸 API 不断增多，再加上第三方 API 常存在过度授权问题，密钥和令牌泄露、复用风险突出，而供应链上下游企业安全防护能力参差不齐，往往一家失守就会引发全链路连锁风险。

这类攻击有着多种典型实施手段，且已出现多起影响广泛的案例，给企业带来严重危害。攻击者会通过钓鱼等社会工程手段窃取第三方 API 密钥与相关令牌，冒充合法身份发起调用以绕开认证，非法访问核心数据；也会发布伪装成正规 API 工具的恶意开源包，在安装环节执行恶意脚本，污染开发环境并植入后门实现横向扩散；同时还会利用 API 未授权访问、规范漂移和配置错误等漏洞，发起恶意请求实现远程代码执行，进而直接控制服务器，造成服务瘫痪或敏感数据窃取。这类攻击不仅会导致企业核心数据泄露，还可能引发资金被盗、业务被操纵等问题，严重威胁企业业务连续性与用户信息安全。

3. 业务逻辑层攻击占比持续上升

攻击手段从“技术漏洞利用”转向“业务流程绕过”，通过滥用 API 功能逻辑实施攻击，此类攻击符合正常通信协议，传统防御工具难以识别，成为 API 安全事件的主要诱因。

越权攻击精细化：从简单的“垂直/水平越权”向“基于上下文的权限绕过”演进，如利用 API 调用顺序漏洞（先调用“获取临时权限”接口，再调用“高权限操作”接口）、会话状态管理缺陷（如 Token 未及时失效、会话固定）实施攻击。

业务流程篡改：攻击者通过 API 篡改业务逻辑，如电商平台“订单状态篡改”（修改支付状态为“已支付”）、金融平台“交易参数篡改”（修改贷款利率、还款期限）、政

务平台“资质审核绕过”（修改申请状态为“通过”）。此类攻击隐蔽性强，传统 WAF 难以检测。

拒绝服务攻击升级：从“流量型 DDoS”向“逻辑型 DoS”演进，利用 API 业务逻辑缺陷（如复杂查询未分页、数据库未索引）发起低流量攻击（每秒 10-100 次请求），即可耗尽服务器资源。

3.5.3.2 防御技术发展方向

面对攻击技术的快速演进，API 防御正从“被动封堵”向“主动治理、智能响应”转型，形成“全生命周期治理 + 智能化防御 + 专业化工具链”的核心发展路径，强调技术与流程的深度融合。

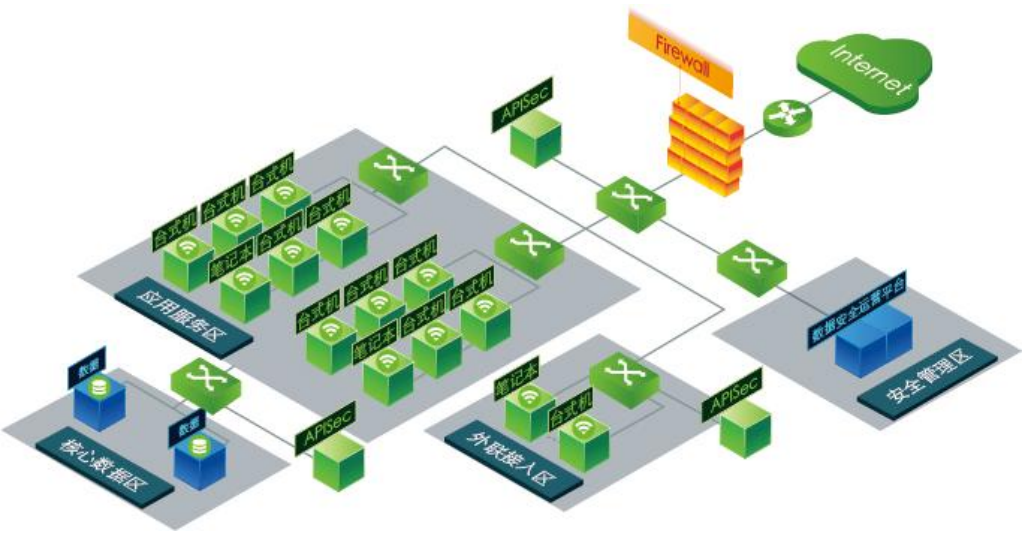


图 3.16 API 防护向主动治理、智能响应转型。

1. API 安全治理体系化

治理体系化是解决 API 安全根源问题的核心，通过覆盖 API 全生命周期的管理机制，实现从“漏洞修复”到“风险可控”的转变，解决资产不清、流程脱节等基础问题。

全生命周期安全管理：覆盖 API“设计-开发-测试-部署-运行-下线”全流程：

◆ 设计阶段：采用“安全设计模式”（如 API 权限矩阵、数据分类分级）、启用“API 安全设计评审”；

◆ 开发阶段：集成 SAST 工具扫描 API 代码漏洞，采用“安全编码规范”（如 OWASP API 安全编码指南）；

◆ 测试阶段：通过 DAST 工具模拟攻击场景，IAST 工具实时检测运行时漏洞；

◆ 部署阶段：API 网关配置“最小权限”策略，启用 TLS 1.3 加密、JWT 安全签名；

◆ 运行阶段：实时监控 API 流量、行为基线，异常时自动阻断；

◆ 下线阶段：清理残留 API 接口，吊销相关 Token 与证书。

API 资产发现与管理：解决“影子 API”“僵尸 API”问题，技术手段包括：

◆ 主动扫描：通过网络流量分析、端口扫描发现未登记 API；

◆ 被动发现：解析应用代码、日志文件提取 API 接口；

◆ 统一管理平台：建立 API 资产库，记录接口信息（路径、参数、权限、调用方）、版本迭代、安全状态，支持自动化更新与风险评分。

2. 防御技术智能化与协同化

智能化与协同化是提升防御效率的关键，通过 AI 技术突破传统规则检测的局限，结合零信任架构与自动化响应机制，构建“感知-研判-处置”的闭环防御体系。

AI 驱动的异常检测：采用机器学习模型构建 API 正常行为基线，通过监督学习+无监督学习检测异常。

零信任架构深度融合：零信任原则在 API 安全中落地，核心技术包括：

◆ 持续身份验证：API 调用前验证调用方身份（多因素认证、设备健康度检查）；

◆ 最小权限授权：基于“角色+属性+上下文”的动态授权，如根据调用方 IP、设备指纹、时间、风险等级调整权限；

◆ 加密与微分段：API 通信全程加密，敏感 API 部署在微分段网络中，限制横向移动；

◆ 持续监控与日志审计：记录 API 调用全量日志，支持溯源分析。

安全编排自动化响应（SOAR）：将 API 安全事件响应流程标准化、自动化，通过“剧本”（Playbook）定义响应动作（如阻断 IP、吊销 Token、隔离接口）。典型场景：检测到“异常高频调用”→自动触发“临时封禁 IP”→同步至 SIEM 平台→通知安全团队→生成修复建议，响应时间从小时级缩短至分钟级。

3. 专业化防御工具链发展

专业化工具链是防御能力落地的载体，工具正从“单一功能”向“集成化、场景化”演进，形成覆盖设计、测试、部署、运营全环节的产品矩阵，支撑体系化防御需求。

下一代 API 管理：从“流量转发”向“安全防护中枢”演进，核心能力包括：

◆ 细粒度访问控制：支持 RBAC/ABAC/CBAC 多维度授权，动态调整访问权限；

◆ 智能流量管理：基于 AI 识别异常流量，自动限流、分流或清洗；

◆ 数据安全防护：内置数据脱敏引擎、数据泄露防护模块；

◆ 可视化与审计：实时监控 API 调用拓扑、性能指标、安全事件，生成合规报告。

◆ API 安全测试工具创新：

1) 交互式测试：在 API 运行时插桩，实时检测参数注入、越权访问等漏洞；

2) 基于 API 规范的测试：解析 OpenAPI/Swagger 规范，自动生成测试用例，提升测试效率；

3) 混沌测试：模拟超时、错误响应、数据篡改等 API 故障，验证系统韧性；

4) 供应链 API 测试：扫描第三方 API 依赖组件漏洞，评估供应链风险。

◆ API 安全治理平台：整合资产发现、漏洞管理、风险评估、合规审计、安全运营功能，形成“一站式”治理体系。典型平台如 IBM API Connect（集成 API 设计、测试、管理、安全功能）、Apigee Sense（谷歌的 API 安全分析平台，提供实时威胁检测与响应）。

3.5.3.3 新兴技术对 API 安全的影响

新兴技术为 API 安全带来“双刃剑”效应，既催生新的防御能力，也带来新的安全风险。区块链、量子计算、大语言模型等技术的融入，正重塑 API 安全的技术边界与防御逻辑。

1. 大语言模型（LLM）在 API 安全中的双刃剑效应

LLM 的自然语言理解与生成能力，同时赋能攻击与防御两端，既提升攻击的自动化与精准度，也为安全分析与响应提供高效工具，技术对抗进入“智能博弈”阶段。

攻击端应用：LLM 可分析 API 文档生成攻击思路、优化攻击 payload、自动化社会工程学攻击。2025 年 BlackHat 演示“GPT-4 驱动的 API 渗透测试工具”，漏洞发现效率提升 400%。

防御端应用：LLM 用于 API 安全分析、漏洞智能分类、安全报告生成。典型案例：Splunk 利用 LLM 分析 API 流量日志，攻击检测准确率提升至 97%。

2. AI 大模型在 API 安全领域的应用

在数字化转型深入推进的当下，API 已成为业务协同与数据流转的核心纽带，但其攻击量激增、威胁复杂度升级的态势，让安全管理面临严峻挑战。绿盟科技 API 安

全监测与审计系统（APISec）深度融合大模型技术优势，为 API 安全管理注入智能新动能，构建全链路、高精度的安全防护体系。

依托绿盟科技 AI 安全研究积累与自主知识产权大模型能力，系统实现 API 安全管理的全面升级。在资产梳理环节，通过大模型强化的协议解析与智能关联技术，精准识别内外网 API 资产及关联应用、账号，自动过滤重复与噪音 API，生成高准确性的资产台账，清晰呈现攻击面。敏感数据识别方面，借助大模型对多行业数据分类分级的深度理解，结合内置隐私数据与行业模板，实现 99% 以上的敏感信息识别准确率，筑牢数据泄露防护根基。

针对复杂多变的 API 威胁，大模型赋能的多维异常检测机制成效显著。系统整合数据、账号、频率等六大维度特征，通过大模型对业务逻辑漏洞、提示词注入等新型攻击的语义分析能力，精准识别低噪音渗透、多阶段链式攻击等隐蔽威胁，大幅缩短攻击驻留时间。同时，大模型驱动的可视化审计与智能响应，让 API 调用流转、安全事件溯源全程透明，结合自动化处置策略，显著提升安全运营效率，助力企业在合规要求下实现 API 全生命周期安全管控。

3.5.3.4 未来技术趋势预测

API 安全技术正处于攻防迭代的关键期，未来将围绕“智能化、体系化、前瞻性”三大方向发展，技术创新与标准合规的协同将成为构建安全能力的核心逻辑。

1. “AI 原生” API 安全成为主流：防御系统将深度融合大语言模型、强化学习等 AI 技术，实现“攻击自动识别、威胁自动研判、响应自动执行”的闭环防御，攻击与防御的 AI 对抗加剧。

2. 零信任 API 安全全面落地：企业从“试点”向“规模化部署”推进，零信任与 API 网关、身份管理、安全运营平台深度集成，成为 API 安全的“基础设施”。

3. API 供应链安全体系化：建立“第三方 API 风险评估 - 准入控制 - 持续监控 - 应急处置”全流程管理机制，区块链存证、供应链安全评分等技术广泛应用。

4. 后量子 API 安全提前布局：企业开始评估量子计算对 API 加密的影响，试点后量子密码算法（如 CRYSTALS-Kyber），改造 API 认证与加密体系。

5. 标准化与合规驱动技术融合：国内外 API 安全标准协同发展，推动“安全技术-管理流程-合规要求”一体化，API 安全治理平台成为企业标配。

API 安全技术正处于“攻防对抗升级、技术快速迭代、标准逐步完善”的关键阶段，需通过持续技术创新、跨领域协作、全球标准协同，构建“主动防御、智能响应、动态适配”的下一代 API 安全体系，应对日益复杂的安全威胁。

3.6 云计算安全

3.6.1 热点安全事件

3.6.1.1 勒索团伙引发赫兹租车遭重大数据泄露

2024 年末，全球知名汽车租赁公司赫兹（Hertz）爆发重大数据泄露事件。黑客组织 CL0P 在 2024 年 10 月至 12 月期间非法访问系统，窃取了数十万客户的个人信息。泄露影响赫兹旗下品牌 Hertz、Dollar 和 Thrifty 的客户，涉及信息包括姓名、联系方式、出生日期、信用卡信息、驾驶执照信息以及与工伤赔偿相关的敏感数据。更严重的是，少部分客户的社会安全号码、护照信息及医疗保险 ID 也被窃取。赫兹虽未发现数据被滥用的直接证据，但事件规模和敏感性引发广泛用户担忧。

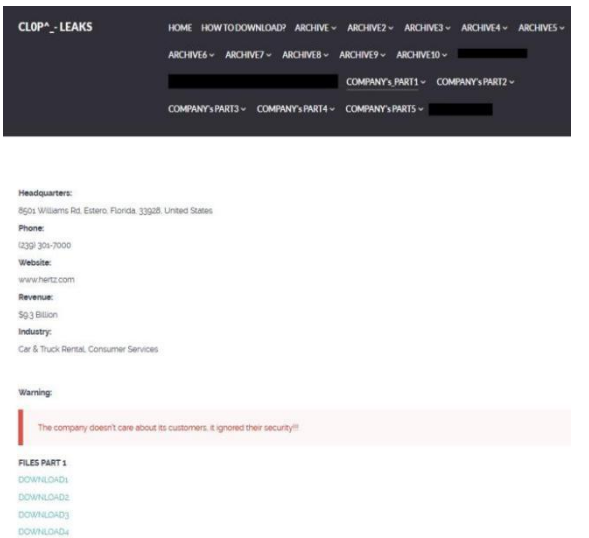


图 3.17 赫兹公司数据泄露相关信息截图 1

图 3.18 赫兹公司数据泄露相关信息截图 2

事件分析：

事件的根本原因是赫兹合作供应商 Cleo Communications 的文件传输平台存在严重安全漏洞 CVE-2024-50623 被恶意利用。CVE-2024-50623 是 Cleo 文件传输平台的软件架构中的一个 0Day 漏洞，该漏洞允许攻击者绕过认证机制，非法访问系统并执行远程代码。

黑客组织 CL0P 在 2024 年 10 月至 12 月期间利用此漏洞入侵 Cleo 平台。CL0P 首先通过漏洞植入恶意后门，持续监视并窃取传输中的客户数据；Cleo 虽在事件后发布补丁，但补丁存在设计缺陷，导致攻击者仍能通过变种攻击绕过防护机制，利用自动运行缺陷扩大攻击面，最终窃取包括信用卡信息和社会安全号码在内的敏感数据。

事件暴露了赫兹在供应链安全管理上的重大缺陷。赫兹过度依赖 Cleo 这类第三方供应商，却未严格执行安全审查机制，例如未建立供应商安全评估框架或实时监控机制。这导致 Cleo 漏洞未被及时发现，攻击窗口长达数月。

3.6.1.2 APT 组织对 Amazon 账户进行劫持导致数据外泄

2025 年 9 月，安全公司 Rapid7 观察到一个新兴威胁组织 Crimson Collective 正积极攻击 AWS 云环境，目的是为了窃取数据并勒索受害者。Rapid7 报告记录了多种“云原生”窃取技术：攻击者利用收集到的 AWS 访问密钥获得持久访问权限；修改 RDS 配置利用 AWS 内置功能读取与下载受害者数据；最终利用受害者自身 S3/EC2/SES 等服务来存储/传送/发送勒索信。Rapid7 得出的结论是，Crimson Collective 代表了一种新的云原生威胁行为者，主要专注于数据盗窃、勒索和声誉损害，而不是直接部署勒索软件。

图 3.19 攻击示意图

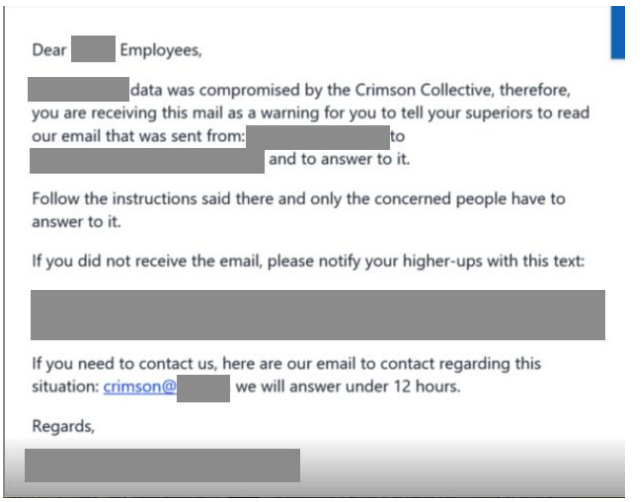


图 3.20 发送给受害者的勒索信

此事件的根本原因在于云环境中长期凭证泄露或滥用+权限 / 配置控制松散，导致攻击者能够取得初始访问，随后在环境中横向扩展、提升权限、搜集并外泄数据，最终用于勒索。

3.6.1.3 AWS DNS 故障导致级联服务瘫痪

此次故障从一个核心服务的 DNS 问题，演变成服务级联故障，导致 AWS 服务中断或性能下降，对大量客户造成了重大影响。核心服务瘫痪导致所有依赖它的应用和服务中断，进一步影响到其他云服务。

金融服务公司 Venmo 和 Robinhood、加密货币交易所 Coinbase、苹果公司的音乐和电视产品、AI 公司 Perplexity、视频网站 Zoom、索尼游戏平台 PlayStation、美国联合航空等网站或应用都在当天经历了服务中断，而英国政府网站 Gov.uk 和英国税务海关总署也遇到了问题。在故障发生后的短短两小时内，仅美国地区的相关投诉量便突破 2 万条。

本事件由 DynamoDB 的 DNS 问题引发。工程师修复 DynamoDB 后 EC2 的管控组件 DWFM 恢复时由于请求量过大，导致系统陷入拥塞崩溃状态。工程师修复 DWFM 恢复正常，新 EC2 实例开始成功启动后，由于新启动的 EC2 实例网络配置延迟，又导致网络负载均衡器（NLB）故障。整个形成了雪崩式故障。

本次事件虽由云厂商引发，但暴露了企业级架构在面对可用性安全时的几类问题：

- 缺乏过载保护导致的内部拒绝服务风险

故障恢复期的流量突增打穿了系统防线。DWFM 组件因缺失限流与熔断机制，导致自身瘫痪。值得注意的是，在大多数企业内部服务间普遍缺乏流量治理，一旦遭遇突发流量或故障重启，极易发生拥塞崩溃。

- 自动化运维策略的误报问题

NLB 健康检查缺乏上下文感知，无法区分“配置延迟”与“服务故障”，导致正常实例被阻断。传统监控与健康检查策略基于静态阈值，难以应对网络抖动等复杂场景，容易造成误报或误操作。

- 强耦合的架构缺乏故障隔离能力

EC2 管控面强依赖于 DynamoDB 并且无降级方案，导致底层故障造成上层管理的功能瘫痪，修复耗时长。关键业务流程若缺乏依赖梳理与故障隔离设计，单一组件失效将导致业务全线停摆。

3.6.2 国内外发展现状

3.6.2.1 技术发展

开源大模型应用云上暴露面激增

2025 年，随着大模型应用大规模迁移至云端，其在云上的暴露面显著扩大，带来新的安全挑战。为系统评估这一风险，绿盟科技星云实验室基于项目活跃度与 Star 数量等关键指标，对全球开源大模型资产进行了空间测绘分析。研究精准描绘了此类应用在 AWS、阿里云、腾讯云等主流云平台上的分布格局，揭示出当前暴露面管理的严峻形势①②③④。

以 Dify 为例，星云实验室借助网络空间测绘技术发现，其全球服务实例数量已达 58000 个，显示出明显的部署规模。从地域分布来看，Dify 服务高度集中于前五大市场：中国以 30349 次（占比 52%）居首，美国、日本、德国和新加坡紧随其后，五国合计占总量的 86%。在云服务部署方面，阿里云以 43%的份额领先，腾讯云（25%）与 AWS（17%）次之，其余 15%由其他云平台共同承担。作为具有代表性的 AI 应用平台，Dify 的部署现状反映出当前 AI 上云整体趋势呈现出规模化与集中化特征，进一步凸显了伴随该趋势而来的暴露面管控风险。

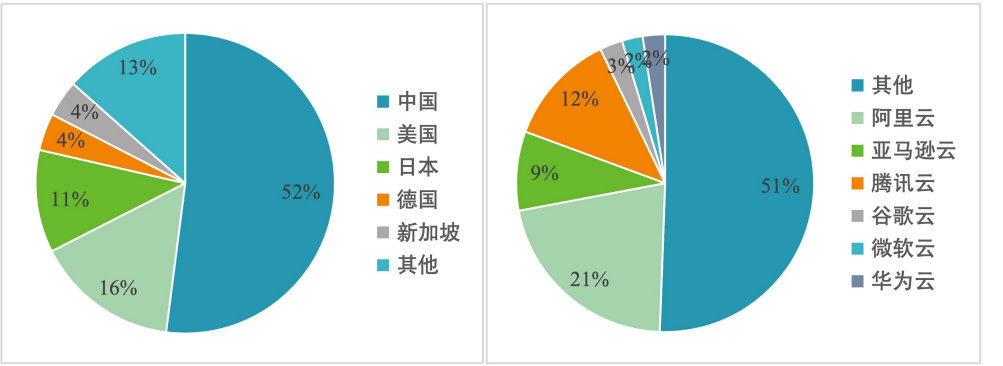


图 3.21 Dify 全球区域分布与云厂商分布情况

① https://mp.weixin.qq.com/s/-hHPcWM71kW--c51GoT4qw
② https://mp.weixin.qq.com/s/ADHC4e03ymaPe5ifZ7aODA
③ https://mp.weixin.qq.com/s/KZsGvmyE6WtspDb5ZvNKVg
④ https://mp.weixin.qq.com/s/5jndWjm_yMEXY0E-W369NQ

云上 AI 数据泄露风险事件频发，配置错误是核心痛点。基于绿盟科技星云实验室对 2025 年 1 月至 10 月期间全球范围内发生的 38 起云上数据泄露典型事件的汇总分析，从事件诱因来看，配置错误依然是数据泄露最主要的成因，占比高达 50%。系统入侵位列第二，占总体的 27%。此外，社工类攻击占比 13%，而 Web 基础类攻击与丢失或被盗凭证分别占 10%，具体占比如下图所示：

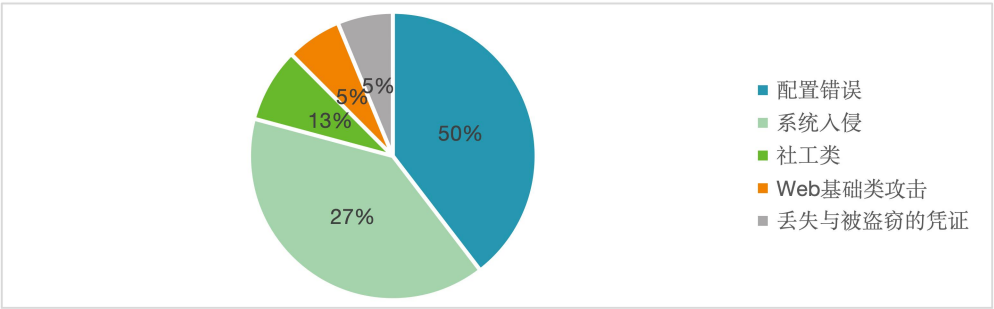


图 3.22 云数据泄露事件类型分布

值得关注的是，上述事件中与 AI 组件相关的安全事件共计 17 起，占总体的 45%，反映出新兴 AI 组件的广泛引入显著增加了因配置缺陷和漏洞利用引发的安全风险。我们的研究进一步表明，由 AI 技术所引发的安全问题正在持续扩大，并呈现出新的演变趋势。目前已观察到的主要攻击面包括：

1. 由 AI 应用所依赖的云基础设施配置错误导致的数据泄露

2025 年出现了多起因租户配置不当引发的数据泄露事件，例如 Elasticsearch 实例与 Kafka 集群的暴露问题。这些案例并非针对 AI 模型的直接攻击，而是利用了 AI 服务所依赖的底层基础设施在配置上的疏忽，最终导致用户数据与隐私对话外泄。此类情况表明，AI 系统的安全防护必须覆盖完整的技术栈与系统生命周期。2025 年的典型案例如下：

2025 年 9 月，研究人员发现一个与 VyroAI 相关的 Elasticsearch 实例因未正确配置访问控制，泄露了该公司三款 AI 应用： ImagineArt、Chatly 和 ChatbotxAI 累计 116GB 的实时用户日志。泄露范围涉及生产与开发环境中约 2 至 7 天的日志数据，内容包括用户输入的提示词、身份验证令牌及 User-Agent 信息。攻击者一旦获取这些令牌，即可劫持用户会话，访问完整聊天记录、生成图像或越权购买 AI 服务。由于部分对话提示词涉及敏感或私密内容，此次泄露也带来了严重的隐私风险[①]。

2. 由提示词注入引发的 AI 应用数据泄露

随着大模型技术的广泛应用，由提示词注入引发的数据泄露事件正日益增多。许多新兴的攻击手法，例如通过提示词诱导 AI 模型执行恶意指令，甚至将敏感信息渲染为图片以规避传统检测，正对数据安全构成严峻挑战。同时，AI 技术的持续演进，如多模态化、智能体化，以及近期 Cursor 曝出的 MCP 漏洞，在催生新技术的同时也带来了新的风险。特别是 AI 模型与第三方应用的集成，虽然提升了便捷性，但权限配置不当可能导致跨用户间的敏感信息泄露。

企业数字化转型正迅速从“云原生”迈向“AI 原生”时代

随着大模型与生成式 AI 技术的爆发式增长，企业数字化转型正迅速从“云原生”迈向“AI 原生”时代。越来越多的企业将模型推理集群、训练任务和 AI 应用部署在云原生环境，利用其弹性伸缩和资源编排优势。然而，这种融合也带来了新风险，如攻击链路将呈现出跨多层级、智能化的特征。攻击者将不再满足于针对模型的简单诱导，而是构建出一条从“模型层”到“基础设施层”的复合攻击路径。例如，攻击者利用提示词注入 RCE 获取容器权限，随后利用容器环境已知漏洞或错误配置挂载实现容器逃逸或横向移动。在这种场景下，大模型实际上成为了攻击者在云内部的跳板，进而获取 Kubernetes 集群的控制权或窃取核心数据。

① https://cybernews.com/security/ai-chatbots-vyro-data-leak/

3.6.2.2 国外政策标准

云安全合规体系不仅需要遵循针对云计算行业标准如 CSA CCM，还必须严格适配通用的网络安全框架与数据隐私法规。尽管 NIST 的网络安全框架 CSF 和通用数据保护条例 GDPR 并非专为云计算环境而生，但随着企业业务与核心数据向云端大规模迁移，云基础设施已成为这些通用标准管控的核心领域。因此无论是云服务商还是企业租户，在云端开展业务时，都必须将这些通用合规基线映射到云环境中，与云标准共同构成多维度的安全合规防线。

CSA 云控制矩阵（CCM）

CSA 云控制矩阵（CCM）是全球公认的专为云计算设计的网络安全控制框架，其体系包含 17 个安全域和 197 个具体控制项，全面覆盖了从底层物理安全、身份认证到上层供应链管理的各个风险层面。作为一个连接供需双方的“通用语言”，企业可利用它作为核心标尺来量化评估云服务商的安全能力，而云服务商则通过它及配套的 CAIQ 问卷主动展示合规性以赢得客户信任。更重要的是 CCM 扮演着“元框架”的关键角色，它将自身控制项与 ISO 27001、NIST、PCI DSS 及 GDPR 等全球主流标准进行了详细映射，使得企业只需满足 CCM 的要求即可在很大程度上同时覆盖多项法规标准，从而实现“一次评估，多重合规”，极大地降低了行业重复审计的成本。

面对生成式 AI 在云端的爆发式增长，现有的 CSA CCM v4.0 版本在适应性上已显不足[①]。究其原因，该版本主要面向传统的 IT 架构设计，尚未覆盖大模型特有的概率性输出风险、提示词新型交互接口以及训练数据隐私等专属控制领域。对此，CSA 在 2024 年采取了“指南先行”的策略，依托其新成立的 AI 安全中心

① Cloud Security Alliance (CSA). (2021). Cloud Controls Matrix (CCM) v4.0

（AI Safety Initiative）发布了一系列专项指引进行补充[①]。预计未来的 CCM 迭代将聚焦于生成式 AI 的核心风险治理，重点解决多租户环境下的数据隔离与防投毒问题、保障云端模型的供应链安全[②]、构建针对“提示词注入”的新型防御机制，以及强化对“影子 AI”导致的数据泄露管控。

网络安全框架 CSF

《网络安全框架》（Cyber security Framework，简称 CSF）[③]是美国国家标准与技术研究院（简称 NIST）发布的一项具有重大影响力的网络安全标准，该框架的核心目标在于协助各类组织提升其网络安全防护、检测、响应和恢复能力，以有效应对不断演化的网络安全威胁。该框架对于企业构建云安全管理体系具有重要的参考价值，它可以帮助企业更好地管理云服务安全并确保数据和系统的安全性。

2024 年 2 月，NIST 正式发布了 CSF 2.0 正式版本[④]，2025 年，CSF 相较于 2024 的主要变化是 2.0 版本的全面实施，尤其是在以下几方面：

1. 新增的“治理”核心功能

新增的“治理”功能为框架的第六个核心功能，与原有的“识别(Identify)”、“防护(Protect)”、“检测(Detect)”、“响应(Respond)”和“恢复(Recover)”并列。“治理”功能旨在确保网络安全策略与企业风险管理紧密结合，获得高层领导的支持和监督。“治理”内容涵盖了组织背景、风险管理策略、网络安全供应链风险管理、角色与职责、政策以及监督等方面的成果。

2. 适用范围显著扩大

① Cloud Security Alliance (CSA). (2024). CSA AI Safety Initiative: Progress Report and Guidelines.
② Cloud Security Alliance (CSA). (2023/2024). Top 10 Critical Risks for Large Language Models.
③ https://en.wikipedia.org/wiki/NIST_Cybersecurity_Framework
④ https://www.secrss.com/articles/63992

CSF 2.0 明确将其适用范围从最初的美国关键基础设施扩展至所有行业和规模的组织，无论其网络安全成熟度如何，需要在全球化得到应用。

3. 强化网络安全供应链风险管理

CSF 2.0 版本对供应链风险管理的关注度显著提升，并将其整合到新的“治理”功能中。

4. 丰富的配套资源和工具

为了提高框架的易用性，NIST 发布了一系列新的配套资源，例如 CSF 2.0 参考工具[①]：一个可搜索的在线工具，允许用户浏览、搜索和导出框架核心指导，并查看其与 50 多种其他网络安全文档的映射关系。

5. 持续改进和信息共享

2025 年，NIST 持续发布与 CSF 2.0 相关的资源和映射关系，以保持框架的活力和实用性。例如发布了 CSF2.0 与其他 NIST 标准（如 NIST SP800-53,SP800-171）的最新映射关系[②]。总之 2025 年的主要任务是理解、采纳和有效实施 CSF 2.0 框架。详细内容可参考官方发布的正式稿件[③]

3.6.2.3 国内政策标准

我国信息安全标准化技术委员与行业单位、机构和各厂商，已经就云计算和云原生编制和发布一批云计算安全相关的标准和规范，规范了云计算产业发展，主要涉及国家标准、行业标准和团体标准，以及一些技术规范。

网络安全等级保护-云计算扩展要求

① https://csrc.nist.gov/Projects/Cybersecurity-Framework/Filters#/csf/filters
② https://www.nist.gov/cyberframework
③ https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf

全国信息安全标准化技术委员发布等级保护的国家标准，提出了云计算扩展要求。中国人民银行发布了金融行业网络安全等级保护标准，其在国标的基础上，对入侵防范、恶意代码和垃圾邮件防范等 12 个控制点进行了增强，新增了“人员考核”控制点。这些标准明确了不同等级云计算平台应该具备的安全要求和技术要求，为我国金融行业等级保护工作开展提供了重要指导。2023 年由中国银联牵头、绿盟科技联合牵头参与的《金融行业云原生安全体系研究》报告为金融行业云原生安全防护体系落地提供了较高参考价值。

注：2025 年下述金融行业网络安全相关标准均现行有效，也无代替或废弃标准。

表 3.2 金融行业云计算安全相关标准

机构	名称
全国信息安全标准化技术委员	GB/T22239-2019 信息安全技术网络安全等级保护基本要求
	GB/T25070-2019 信息安全技术网络安全等级保护设计技术要求
	GB/T28448-2019 信息安全技术网络安全等级保护测评要求
中国人民银行	JR/T0167-2020 云计算技术金融应用规范安全技术要求
	JR/T0071—2020 金融行业网络安全等级保护实施指引
	JR/T0072-2020 金融行业网络安全等级保护测评指南
	JR/T0073-2012 金融行业信息安全等级保护测评服务安全指引
中国银联	2023 金融行业云原生安全体系研究

云计算安全框架

全国信息安全标准化技术委员会牵头多家单位制定了《云计算安全参考架构》和《云计算服务运行监管框架》等多项标准。这些标准一方面为云计算参与者在云计算系统安全规划和设计时提供参考，另一方面也指导如何进行云计算服务安全能力评估，为云计算服务安全评估提供评估依据。

表 3.3　全国信息安全标准化技术委员会制定的云安全标准

机构	名称	2025 年变更
全国信息安全标准化技术委员	20251238-T-469《网络安全技术政务云平台安全监测方法》	起草中
	20251240-T-469《网络安全技术政务云安全配置基线要求》	起草中
	GB/T31167-2023《信息安全技术云计算服务安全指南》	无变化
	GB/T31168-2023《信息安全技术云计算服务安全能力要求》	2025 年 8 月正式发布，2026 年 2 月 1 日正式实施
	GB/T37972-2019《信息安全技术云计算服务运行监管框架》	无变化
	GB/T35279-2017《信息安全技术云计算安全参考架构》	无变化
	GB/T34942-2017《信息安全技术云计算服务安全能力评估方法》	GB/T34942-2025，替代 GB/T34942-2017，于 2025 年 8 月发布，2026 年 2 月实施。GB/T34942-2025 详细规定了如何依据 GB/T31168-2023《网络安全技术云计算服务安全能力要求》对云服务商进行安全评估，包括评估原则、实施过程和具体方法。该标准为第三方评估和云服务商自评提供了统一指南。

云原生安全

云原生技术正在广泛应用，其安全性成为各机构的研究重点，中国信通院和 CSA 大中华区分别成立了云原生安全工作组，开展云原生安全领域的标准制定工作，推动云原生安全发展。CSA 大中华区编制了《云原生应用技术规范》和《实现安全应用容器架构的最佳实践》。中国信通院围绕云原生安全，提出了安全基线、API 安全治理、CNAPP、CWPP 和数据安全等细分领域的安全标准，并开展云原生产品的安全测评，

促进网络安全产生提供更优秀的安全产品，也有助于云原生用户选择可信任的云原生安全产品。中关村信息安全测评联盟编制了《网络安全等级保护容器安全要求》，从等级保护的视角，提出了容器应具备的安全能力。

表 3.4 各机构制定的云计算安全相关标准

机构	名称
中国信通院	《容器平台安全能力要求》
	《云原生安全配置基线规范 2.0》
	《云原生能力成熟度模型》
	《云原生安全能力要求第 1 部分：API 安全治理》
	《云原生应用保护平台（CNAPP）能力要求》
	《云工作负载保护平台能力要求》
	《云原生数据安全能力要求》
	《云原生安全安全托管服务（MSS）能力要求》
CSA 大中华区	《云原生安全技术规范》
	《实现安全应用容器架构的最佳实践》
中关村信息安全测评联盟	《网络安全等级保护容器安全要求》

2024 年 7 月，信通院联合各家单位联合发布了《云原生安全配置基线 V2.0》，相较于 2023 年发布的 1.0 版本主要提升点在于对 Kubernetes 组件的安全配置进行了更为严格的规范要求，包括 Kubernetes 的核心组件 APIServer、控制管理器、调度器、Etcd，节点组件 kubelet、kube-proxy，以及其 CRI 插件和网络策略的相关安全配置。其次提升了基线范围，考虑了容器运行时以及提升了基线的自动化水平。

2025 年，信通院在 7 月发布了《云计算蓝皮书》，提出了“智算云”和“云原生 AI 安全”的概念。指出云原生安全不仅要保护容器，还要保护运行在云原生底座上的大模型和 AI 应用。

3.6.2.4 技术发展观察

趋势一：到 2026 年，随着 AI 应用加速向云端迁移，新兴 AI 组件的引入及供应链复杂度的提升，将显著增加配置缺陷与漏洞利用的风险。这导致模型参数、模型聊天记录、AI 密钥等核心资产面临严重的泄露威胁。因此，精准识别并有效收敛 AI 资产的互联网暴露面，已成为云上 AI 数据安全的首要防线。

趋势二：2025 年 AI 与云原生架构融合导致新的攻击面。攻击者将大模型漏洞作为核心突破口，利用云环境的脆弱性突破容器隔离边界，进而获取云基础设施的核心控制权。为应对此威胁链路，防守方亟需构建云 AI 靶场，通过实战化的全链路演练，有效验证“AI+云”复杂场景下的安全能力和防御闭环。

2025 年，构建云上 AI 靶场将成为防守方建设的核心趋势。与传统靶场不同，云 AI 靶场需要高保真地复现“大模型+云原生”的融合架构，能够模拟针对 AI 智能体的逻辑攻击、针对容器的逃逸攻击、以及针对容器编排平台的权限滥用等全链路攻击行为。通过引入云自动化红队工具和 AI 靶场，企业可以进行高频次、全要素的实战演练。能够验证云原生安全策略在 AI 场景下的有效性。

3.7 供应链安全

3.7.1 热点安全事件

近年来，软件供应链安全已成为全球数字安全领域的核心挑战。2024 至 2025 年间，软件供应链攻击呈现规模化、精细化、跨国化的发展趋势，攻击手法不断创新，影响范围持续扩大。行业监测数据显示，软件供应链攻击事件持续显著增长，其中针

对开源生态和关键基础设施的攻击尤为突出。这些事件不仅导致巨额经济损失，更暴露了全球软件供应链在安全管理、技术防护和跨组织协同等方面的深层次漏洞。

3.7.1.1 开源生态供应链攻击事件

开源软件作为现代应用的基石，已成为供应链攻击的主要目标。研究表明，高达70%的企业正在采用开源软件，但 90%的组织在管理其安全方面存在困难。2024 至2025 年间，针对主流开源生态系统的攻击事件频发，攻击手法从单纯的恶意包上传发展到利用社会工程学和自动化攻击工具链的复杂攻击。

2025 年，卡巴斯基实验室发现一个名为“Tsundere”的新型僵尸网络。该威胁行为者最初在 2024 年 10 月通过创建 287 个恶意 Node.js 包，使用仿冒流行包名称的方式在 NPM 上分发恶意软件，影响了 Windows、Linux 和 macOS 用户。在 2025 年 7 月左右，攻击者升级战术，该恶意软件具备动态执行远程代码的能力并实现持久化，展示了供应链攻击与新兴技术结合的复杂趋势。

表 3.5 2024-2025 年主要开源生态供应链攻击事件对比

攻击事件	目标生态系统	主要攻击手法	影响范围	行业影响
NPM 恶意包投毒	NPM	自动化创建海量恶意包、滥用区块链奖励机制	超 15 万个恶意包	污染整个开源生态，消耗基础设施资源
Tsundere 僵尸网络	NPM、Node.js	仿冒包、区块链 C2、动态代码执行	287 个恶意包，影响多平台用户	IT 基础设施、加密货币
XZ Utils 后门	Linux 生态系统	社会工程、长期渗透、代码隐匿	近 SSH 的广泛 Linux 系统	IT 基础设施、关键行业
攻击事件	目标生态系统	主要攻击手法	影响范围	行业影响

3.7.1.2 工业控制系统与物联网供应链攻击

随着工业 4.0 和物联网技术的普及，针对物理世界的软件供应链攻击呈现出破坏性更强、影响更直接的特点。2024 至 2025 年间，工业控制系统和物联网设备成为供应链攻击的高价值目标。

1. BadBox 2.0 僵尸网络全球爆发事件

2025 年初，美国 FBI 联邦调查局发布紧急警告，披露名为 BadBox 2.0 的大规模物联网僵尸网络已感染全球数百万台设备。被感染的设备会被劫持为“住宅代理节点”，攻击者借此伪装正常家庭网络流量，实施广告欺诈、分布式拒绝服务（DDoS）攻击、账户劫持和金融诈骗等非法活动。该事件暴露了物联网供应链中从硬件制造到软件分发环节的多重脆弱性，凸显出全球化供应链中安全监管缺位可能引发的跨域系统性风险。

2. 物联网设备批量高危漏洞

2025 年 3 月，国家信息安全漏洞共享平台(CNVD)披露多款主流物联网设备存在高危供应链漏洞，涉及路由器、工业控制器、医疗辅助设备等多个品类。TOTOLINK X18 系列路由器的命令注入与缓冲区溢出漏洞(CVE-2025-1340)影响了国内大量设备，而工业控制器的协议处理漏洞可能引发重大不可恢复故障。这些漏洞的共同特征是源于供应链上游的设计缺陷或安全测试不足，且由于 IoT 设备部署广泛、固件更新滞后，平均修复率较低，大量设备长期处于高风险状态。

3.7.1.3 商业软件与云服务供应链攻击

商业软件和云服务在现代企业 IT 架构中占据核心地位，针对这些目标的供应链攻击往往能造成大规模数据泄露和业务连续性中断。调研显示，68%的安全负责人担忧其组织技术栈中由第三方引入的软件工具或组件所带来的风险。

2025 年 8 月，攻击团伙 GRUB1 通过入侵 Salesloft 的 Drift 应用程序，窃取 OAuth 令牌，进而获取到与 Drift 连接的 Salesforce 实例的访问权限。攻击者声称从 760 家公司窃取了超过 1.5 亿条 Salesforce 记录，受害者包括 Palo Alto Networks、Zscaler 和 Cloudflare 等网络安全行业的领军企业。这一事件凸显了 SaaS 生态系统中的互信关系如何被利用来扩大攻击影响，一个边缘服务的沦陷可能导致整个关联生态的数据泄露。

3.7.2 国内外发展现状

软件供应链安全已成为全球网络空间安全治理的核心议题之一。随着软件产业分工全球化、组件开源化、交付链条复杂化，供应链攻击已成为网络攻击的主要形式。据 Verizon《2023 年数据泄露调查报告》显示，软件供应链环节的漏洞或恶意植入在网络安全事件中占据重要比例。各国政府、国际组织、行业联盟及科技企业纷纷通过政策立法、标准制定、技术研发、产业协同等手段构建安全防护体系。

3.7.2.1 国际发展现状

国际社会在软件供应链安全领域的探索起步较早，已形成“政策强制约束、标准体系支撑、产业协同实践”的三位一体发展格局，其核心特征表现为以“设计安全（Security by Design）”和“零信任（Zero Trust）”为核心理念，以软件物料清单（SBOM）为关键抓手，覆盖软件全生命周期的安全管控体系。

1. 政策法规：强制化与精细化管控并行

美国通过行政命令、联邦法规、备忘录等形式构建了层级分明的软件供应链安全政策体系，呈现强制化与精细化管控并行特征。2021 年美国第 14028 号总统令《改善国家网络安全》，将软件供应链安全提升至国家战略层面，明确联邦政府采购软件需提供 SBOM、供应商需实施安全软件开发实践并接受第三方审计。2023 年《国家网络安全战略》进一步提出“市场激励+政府监管”双轮驱动模式，推动全行业采用安全开发生命周期并建立攻击快速响应机制。

欧盟则通过区域协同立法强化管控。欧盟《网络弹性法案》要求进入欧盟市场的带有数字元素的产品（含开源软件）需提供 SBOM、实施漏洞管理及产品追溯，未达标产品禁止销售。同时，欧盟《数字运营韧性法案》（DORA）亦将金融领域软件供应链安全纳入监管范畴，要求金融机构对第三方供应商进行严格的安全风险评估。

2. 技术标准：体系化与实操性并重

国际标准化组织、行业联盟及技术机构已构建覆盖软件供应链全生命周期的技术标准体系，呈现体系化与实操性并重特征，为政策落地提供坚实技术支撑。

ISO/IEC 27036 系列标准作为供应链信息安全管理核心框架，其中 ISO/IEC 27036-4《供应商关系管理》明确软件供应商安全资质、合同安全条款及持续监控机制，ISO/IEC 27036-5《供应链安全指南》针对软件组件采购、集成、交付等环节提出风险管控措施。

NIST 发布《安全软件开发框架（SSDF）1.1 版》（SP 800-218）将安全开发划分为“准备、保护、生产、响应、改进”五大环节，提供多项核心实践和细化任务，成为全球软件企业安全开发参考模板。《软件供应链安全指南》（SP 800-161 Rev.1）构建包含风险评估、供应商选择、持续监控、事件响应等模块的风险管理框架。《软件物料清单（SBOM）最佳实践》（NIST IR 8363）则明确 SBOM 核心要素、格式标准及应用场景。

行业联盟标准形成有效补充。开放源代码安全基金会（OpenSSF）发布的《供应链安全指南》提出“SLSA（软件供应链层级与保证）框架”，将软件供应链安全等级划分为多个级别，为开源项目和企业应用提供可量化安全评估标准。云安全联盟（CSA）发布的《软件供应链安全指南》则聚焦云原生环境下的供应链安全风险管控。

3. 产业实践：技术创新与生态协同并进

国际科技企业与开源社区通过技术研发、工具推广、生态合作等多元路径推动软件供应链安全实践落地，呈现技术创新与生态协同并进的鲜明特征。

企业端聚焦安全工具研发，GitHub 推出集成依赖项扫描、秘密扫描、代码扫描等功能的“Advanced Security”套件，成为企业供应链安全管理核心工具。微软发布包含 SBOM 生成工具、第三方软件安全评估平台、供应链攻击检测系统的供应链安全套件，其 Azure 云平台已实现 SBOM 自动生成与漏洞关联分析。谷歌则通过“Software Supply Chain Shield”的数字签名、透明日志等技术，构建软件从构建到部署的全链路可追溯体系。

开源社区强化安全治理，OpenSSF 发起的“Scorecard”项目从依赖项安全、代码审查、签名验证等多个维度量化评估开源项目安全状况，为企业选型提供参考。Linux 基金会推动 SBOM Everywhere 倡议，联合众多企业和机构制定开源软件 SBOM 生成标准，主流开源仓库也在逐步要求项目提供 SBOM 并集成漏洞扫描功能。

产业协同机制持续完善。美国成立由政府、企业、科研机构组成的软件供应链安全联盟（S2C2），开展信息共享、威胁情报联动、技术标准验证等工作。欧盟则通过网络安全应急响应协调中心（ENISA）建立供应链攻击跨境响应机制，形成多方联动的安全实践生态。

3.7.2.2 国内发展现状

我国软件供应链安全发展始于“十三五”时期，近年来在国家网络安全战略引领下，呈现政策法规加速完善、国家标准体系成型、产业实践逐步深化的发展态势，核心特征表现为“政策驱动、标准引领、应用导向”，重点聚焦关键信息基础设施、重要行业领域的供应链安全保障。

1. 法律法规：从框架性规定到专项化治理

我国已构建以《网络安全法》《数据安全法》《个人信息保护法》为核心，配套行政法规、部门规章的软件供应链安全法律体系，呈现从框架性规定到专项化治理的演进特征。

2017 年实施的《网络安全法》明确网络产品和服务提供者的漏洞告知、补救等义务，为供应链安全奠定法律基础。2021 年实施的《关键信息基础设施安全保护条例》和修订的《网络安全审查办法》，分别要求关键信息基础设施运营者采购产品和服务时审查供应链安全风险、对影响国家安全的产品和服务实行禁入管控。工信部 2023 年发布的《关于加强工业领域软件供应链安全管理的通知》也针对工业软件提出清单管理、分级管控、全生命周期防护要求。

在电信行业监管方面，工信部持续完善网络安全考核体系。各地方通信管理局依据工信部年度工作要点，对基础电信企业开展网络与信息安全考核，重点检查供应链安全管理情况，对违规行为依法依规予以处置，形成了强有力的监管驱动机制。

2. 国家标准：从单点规范到体系化覆盖

我国软件供应链安全国家标准体系已基本成型，呈现从单点规范到体系化覆盖的发展特征，涵盖基础通用、安全要求、评价方法、技术工具等多个维度，为行业实践提供统一依据。

2024《网络安全技术 软件供应链安全要求》（GB/T 43698-2024）和《网络安全技术 软件产品开源代码安全评价方法》（GB/T 43848-2024）两项核心国家标准正

式实施。其中《软件供应链安全要求》构建“目标-原则-要求-评估”完整框架，明确供应链全生命周期风险管理及供需双方组织与供应活动管理要求。《软件产品开源代码安全评价方法》提出开源代码安全评价“要素-流程-等级”三维模型，为开源代码安全自评与第三方评价提供依据。

同时，《网络安全技术 软件安全开发生命周期管理规范》（GB/T 39414-2020）、《网络安全技术 软件产品安全评价方法》（GB/T 20945-2023）、《信息安全技术 供应链安全风险管理指南》（GB/T 36643-2018）等配套标准已陆续发布，形成覆盖“开发-采购-使用-运维”全流程的标准支撑体系。此外，《网络安全技术 软件物料清单（SBOM）生成与管理规范》已进入征求意见阶段，将进一步明确我国 SBOM 的格式要求、生成方法、应用场景及管理流程。

3. 产业实践：政策驱动与技术创新协同发展

国内安全厂商、互联网企业、行业用户及开源社区围绕政策要求与市场需求，推动软件供应链安全实践呈现政策驱动与技术创新协同发展的态势。

2025 年 7 月，中国信通院联合产业各方共同启动软件供应链供需赋能“清链”计划，旨在提升企业软件供应链管理的韧性与透明度。同时，中国信通院积极推进《人工智能物料清单（AIBOM）数据格式要求》等新兴领域标准编写工作。

安全厂商方面，绿盟科技推出软件供应链安全测评服务与相关平台，提供 SCA 软件成分分析、安全开发全生命周期管理、安全情报预警等核心能力。其方案通过运营数字供应链安全情报平台，对全球范围内的投毒情报、漏洞情报、停服断供情报进行实时监测与分析，实现精准预警。

主流安全厂商纷纷推出集成 SBOM 管理、开源组件扫描（SCA）、安全开发集成（DevSecOps）、供应链威胁检测等核心功能的解决方案。主要云服务商在其云原生产品中集成 SBOM 生成与漏洞管理功能。

关键信息基础设施运营者建立软件供应链安全管理体系，实施供应商安全评估、产品安全检测及持续监控等措施。部分大型企业还试点 DevSecOps 实践，将安全测试融入软件开发流程以强化源头安全。

通信行业的软件供应链安全建设，在工信部的网络安全合规考核驱动下，运营商普遍建立了严格的第三方供应商安全管理制度，动态维护供应商、产品及人员清单，积极部署 SCA、SBOM 管理等技术工具，构建供应链安全管控平台，有效规避了因第三方安全事件引发的运营风险。

3.7.2.3 国内外发展现状对比分析

通过从政策导向、标准体系、技术创新、产业生态、应用场景五个维度进行对比（见表 3.6），可以发现国内外软件供应链安全发展既存在共性特征，也存在显著差异，国内在政策推进速度、标准体系完整性上已逐步追赶国际水平，但在技术创新深度、产业生态成熟度等方面仍有提升空间。

表 3.6 国内外软件供应链安全发展现状对比表

对比维度	国际发展特征	国内发展特征
政策导向	以"市场激励+政府监管"双轮驱动，强调强制化要求与市场化机制结合，覆盖全球供应链	以"政策驱动+行政监管"为主，强调关键领域管控，聚焦国内供应链安全保障
标准体系	体系成熟、国际通用，SBOM、SDLC 等标准已形成行业共识，具有强实操性与兼容性	体系逐步完善，核心标准已发布，与国际标准衔接紧密，但部分细分领域标准仍需补充
技术创新	聚焦 DevSecOps、零信任、区块链追溯等前沿技术，产品化程度高，生态协同性强	技术路线与国际接轨，重点突破 SCA、SBOM 等核心技术，但高端产品市场占比

		仍较低
产业生态	开源社区主导、企业协同参与，形成"标准-工具-服务"完整产业链，市场成熟度高	政策驱动型生态，安全厂商、行业用户为核心，开源社区影响力逐步提升，产业链尚在完善
应用场景	覆盖全行业，重点聚焦政府、金融、科技等领域，全球化应用特征明显	重点聚焦关键信息基础设施、政务、金融等领域，行业应用深度逐步提升

共性特征

国内外软件供应链安全发展呈现四大共性特征，共同构成全球治理的核心方向：

◆ 以政策立法明确供应链安全管控要求，以技术标准规范实践落地，形成"政策引领-标准支撑-实践落地"的闭环机制；

◆ SBOM 作为供应链透明化的核心工具已成为共识；

◆ 通过建立开源项目安全评估机制、推广开源组件漏洞扫描工具、规范开源许可证合规性等方式，强化开源软件在供应链中的安全管控；

◆ 全生命周期防护已成为主流趋势，无论是国际还是国内的框架标准，均强调覆盖软件设计、开发、采购、部署到运维的全流程安全管控，逐步替代传统单点防护模式。

差异与差距

◆ 国内外软件供应链安全发展在政策执行、标准推广、技术创新及产业生态层面存在显著差异与差距：国际政策以“强制要求+市场激励”为核心，设定明确处罚机制，市场化主体主动参与治理的动力较强。而国内政策以行政监管为主，市场化激励机制尚不完善，部分企业实践仍停留在“合规达标”层面；

◆ 国际标准具有强国际通用性和实操性，已被全球企业广泛采纳。国内标准虽与国际衔接紧密，但在企业落地指引方面仍需加强，部分标准实操性有待市场检验；

◆ 技术创新上，微软、谷歌等国际厂商在 DevSecOps 工具链、供应链威胁情报、零信任架构融合等领域具备领先优势，产品化程度高，国内厂商虽在 SCA、SBOM 等领域实现技术突破，但在核心算法、威胁情报积累、跨平台兼容性等方面仍有差距；

◆ 产业生态方面，国际已形成“开源社区-企业-政府”协同格局，OpenSSF、Linux 基金会等组织发挥核心作用。而国内生态以政策驱动为主，开源社区影响力有限，企业间协同不足，尚未构建规模化的供应链安全信息共享与威胁联动机制。

发展趋势

未来，国内外软件供应链安全发展将呈现“趋同与差异化并存”的趋势：趋同方面，SBOM 的普及应用、全生命周期防护、开源安全治理将成为全球共识，国际标准与国内标准的衔接将更加紧密；差异化方面，国际将继续强化全球供应链安全管控，推动供应链安全规则国际化，国内将聚焦关键领域自主可控与安全可信，加快核心技术突破与产业生态完善。随着国内政策落地深化、技术创新加速、产业生态成熟，国内外在软件供应链安全领域的差距将逐步缩小，形成具有中国特色的供应链安全治理体系。

3.7.3 技术发展观察

3.7.3.1 SBOM 全生命周期管理技术

软件物料清单(SBOM) 是包含组件名称、供应商、版本、依赖关系、许可证等核心数据的结构化清单，是实现软件供应链透明化的基础。SBOM 已从单纯的合规文件，演进为供应链安全的核心基础设施，其全生命周期管理技术聚焦“标准化、自动化、动态化、场景化”，成为实现供应链资产可视化、风险精准管控的关键支撑。

1. 多源 SBOM 自动化生成技术

SBOM 生成技术已突破传统源代码分析的局限，实现对多类型资产的全面覆盖。当前主流工具支持源代码、二进制包、容器镜像、IoT 固件等多类型资产的成分自动

识别，覆盖 Java、Python、Go 等 20 余种主流开发语言及 NPM、Maven、PyPI 等包管理器。NIST SP 800-218 报告指出，自动化 SBOM 生成可降低 60%以上的人工错误率，相较于传统人工编制方式，效率显著提升。

在国内实践方面，中国信息通信研究院推动的“清链”计划将 SBOM 作为核心基础设施，通过标准化格式促进供应链上下游信息透明。国家电网等关键信息基础设施运营者在信创环境下创新设计专用 SBOM 格式，新增国产化运行环境、兼容性、自主可控等级等字段，解决传统 SBOM 不适应信创技术栈的问题。

2. 动态化 SBOM 管理与更新技术

为克服传统静态 SBOM 存在的“信息滞后”缺陷，动态化 SBOM 管理技术已成为发展核心，旨在适应组件频繁更新与依赖关系动态变化的真实场景。

该技术首先通过与 CI/CD 流水线深度集成，实现在代码提交、组件升级等关键节点自动触发 SBOM 更新，确保资产信息与部署状态实时同步，例如 Jenkins 的 SBOM 插件即可实现“构建即生成”；同时，通过建立资产属性自动关联机制，能够实时同步组件版本、漏洞状态及许可证信息等关键数据，如 Microsoft Azure 的 SBOM 管理平台可每小时同步漏洞状态，保障风险情报的时效性；此外，其版本追溯功能完整记录每次变更历史，支持跨版本对比以快速定位组件变更引发的风险，Amazon AWS 的 SBOM Explorer 工具便能追溯长达 12 个月的变更记录，为风险根源排查提供有力支撑。

3. SBOM 实战化应用技术

SBOM 的应用已超越传统的合规文件范畴，深度融合到漏洞预警、影响范围研判与供应链协同等核心实战场景，成为软件供应链安全管控的枢纽。

在漏洞关联分析层面，SBOM 能够与 CVE、CNNVD 等漏洞库实时联动，基于组件信息自动发现高危漏洞，并结合资产关键性生成优先修复清单。在影响范围研判方

面，依托 SBOM 构建的依赖关系图谱可快速定位受漏洞影响的系统与节点，大幅提升应急响应效率。在 Log4Shell 漏洞爆发期间，拥有准确 SBOM 的企业能够在几小时内定位受影响系统，而无 SBOM 的企业则需要数周时间进行评估，这充分证明了 SBOM 在应急响应中的关键价值。

在供应链协同中，SBOM 更成为上下游信息同步的桥梁。如某大型金融机构采用 SBOM 管理后，将供应商合规评估周期从 30 天压缩至 10 天，全面提升生态协同效率。

3.7.3.2 AI 与智能体技术

AI 技术已成为供应链安全效率提升的核心引擎，从风险预测、漏洞检测到应急响应，全面重塑供应链安全防御模式。Gartner 预测，到 2026 年，40%的企业将使用 AI 增强的应用程序安全测试工具来识别安全漏洞，比 2023 年的不到 5%大幅增长。

1. AI 驱动的风险预测与漏洞检测技术

AI 驱动的风险预测与漏洞检测技术正推动软件供应链安全从被动响应向主动防御转变。在风险预测领域，基于 AI 的分析系统通过综合分析组件漏洞历史、供应商安全评级及多源威胁情报，能够精准预测组件风险。例如，某电商平台采用 LSTM 模型分析超百万条组件漏洞数据，实现 89%的风险预测准确率；Microsoft 的 AI 平台则通过分析开源平台代码提交和维护者行为数据，可提前 30 天识别恶意包风险，误报率低于 3%。

在漏洞检测方面，基于深度学习的静态应用安全测试工具展现出卓越能力。绿盟科技采用 Transformer 模型对缓冲区溢出、SQL 注入等漏洞的检出率达 99.2%，误报率低于 2.8%，显著增强了供应链安全的主动防护能力。

2. AI 智能体协同防御技术

AI 智能体协同防御技术正推动供应链安全响应从“人工主导”向“智能体协同”的革命性转变。通过构建多智能体协作机制，该系统实现了跨平台、跨场景的智能威胁狩猎与自动化处置。

在技术架构层面，绿盟科技风云卫安全大模型集成了 20 余个专业安全智能体，覆盖漏洞分析、恶意代码检测、攻击溯源等关键场景，各智能体通过 A2A 协议实时共享记忆与推理结果，形成协同作战能力。在易用性层面，策略智能体通过自然语言交互与拓扑图解析实现了对话式策略配置，使用者只需用自然语言描述安全需求（如：禁止来自未知 IP 的组件下载），系统即可自动生成并部署对应安全策略，使配置效率提升 80%，极大降低了非专业人员参与复杂安全运维的门槛。

3.7.3.3 统一安全平台与集成化解决方案

面对软件供应链安全的复杂性和多维性，统一安全平台与集成化解决方案成为行业的重要发展方向。研究表明，现代软件交付的复杂性导致开发团队需要同时应对多种应用安全工具，存在形成工具孤岛的风险。

1. 应用安全态势管理(ASPM)

应用安全态势管理（ASPM）作为近年来快速成熟的技术范畴，通过统一聚合和关联来自 SAST、DAST、SCA、容器安全等多样化工具的安全数据，构建供应链安全的整体视图并实现自动化修复工作流。该平台有效消除工具孤岛，提供统一风险视图。据 Gartner 预测，到 2027 年将有 50%的企业采用 ASPM 统一管理应用安全风险，较当前不足 15%的比例实现跨越式增长。

基于上下文的风险优先级评估机制，综合考虑漏洞可利用性、资产关键性等维度，使企业能将修复资源集中于真正存在风险的 5-10%漏洞，修复效率提升 3-5 倍。同时，ASPM 支持安全策略自动执行与修复工作流智能管理，当发现严重风险时可自动创建

工单、分派任务甚至阻断流水线，将平均修复时间（MTTR）从数周显著缩短至数天，全面提升了软件供应链的安全运营效率。

2. 安全厂商供应链安全治理实践

主流安全厂商将其对供应链安全的理解与实践，固化为具体的平台、工具与服务，为企业提供关键支撑。以绿盟科技为例，其推出的软件供应链安全治理平台提供覆盖软件研发、检测、防护和运营全生命周期的综合解决方案。该平台的核心能力体系全面覆盖数字化安全的关键领域：首先，通过 SCA 软件成分分析精准绘制开源组件图谱，有效识别漏洞、投毒风险及许可证合规问题，其二进制分析能力突破性地覆盖工控固件、车端镜像等特殊场景。其次，将安全开发全生命周期管理深度融入应用开发流程，借助知识库与考评机制实现安全左移；同时依托运营数字供应链安全情报平台，对全球投毒情报、漏洞情报实现实时监测与精准预警。

3.7.3.4 新兴技术与未来趋势

软件供应链安全技术仍在快速发展中，多项新兴技术有望在未来几年内重塑供应链安全防护模式。

1. 区块链与量子加密技术

区块链与量子加密技术正共同构筑软件供应链安全的底层信任基石。区块链凭借其不可篡改、可追溯的技术特性，为供应链溯源提供了革命性解决方案。

与此同时，量子加密技术通过量子密钥分发等机制为供应链数据传输提供面向未来的安全保障，NIST 标准化的 SPHINCS+、CRYSTALS-Kyber 等后量子密码算法在保障安全性的同时显著优化性能，为供应链安全迎接量子计算时代做好了技术储备。

2. 软件供应链安全与 LLM 集成

随着大语言模型在软件开发中的广泛应用，软件供应链安全技术与 LLM 的深度融合已成为重要趋势。OWASP 2024 年发布的《OWASP Top 10 for Large Language Model Applications》中特别强调了软件供应链失效与 LLM 特定威胁的双重挑战。

LLM 引入了一系列新型供应链风险，包括训练数据污染、模型权重篡改以及提示注入攻击等。为应对这些挑战，针对 LLM 的供应链安全防护体系正在形成，涵盖训练数据验证、模型完整性校验、输出内容过滤与持续监控等关键环节。

绿盟科技在此领域的探索表明，利用大语言模型在代码理解、生成和漏洞模式识别方面的潜力，可将其应用于软件供应链的漏洞检测。同时，公司发布的大模型安全评估系统等产品，用于评估和扫描模型组件漏洞等风险，为 AI 自身供应链安全提供保障。

3. 轻量化与边缘计算场景适配

随着软件供应链安全向物联网和边缘计算场景快速扩展，轻量化技术正成为满足资源受限环境下安全需求的重要发展方向。

为应对物联网设备存储和计算能力的限制，业界正积极开发轻量级 SBOM 格式，在保留组件名称、版本、许可证等关键信息的同时大幅减小文件体积，预计到 2027 年此类轻量化 SBOM 将在 80%的 IoT 设备中得到规模化应用。

同时，专门为边缘计算环境设计的轻量级安全代理也日益成熟，能够在有限的硬件资源上执行基本的软件成分分析、漏洞检测等核心安全功能，有效支持离线或间歇性连接环境下的供应链安全管控需求，为构建端到端的软件供应链安全体系提供了关键的技术支撑。

3.8 蓝军建设

3.8.1 威胁热点观察

2025 年，AI 已从攻击活动的辅助效率提升工具，变成融入网络攻击执行阶段的攻击工具组件，未来有可能成为攻击活动的重要支撑能力。

在攻击初始阶段，借助 AI 技术显著降低了社会工程学攻破目标心理防线的难度。2025 年 11 月，UNC1069（MASAN）针对加密行业，利用 AI 生成技术伪造高管照片与视频，通过实时视频会议伪装成企业技术负责人，成功诱骗企业员工执行恶意操作。其欺骗效果之强，使受害者无法通过常规的视觉和语言特征辨识危险，彻底突破依赖人工经验的信任机制。

除了伪造视频欺骗视觉，AI 生成文本的能力也极大地加快了鱼叉攻击的速度。APT42 利用 AI 生成钓鱼邮件，通过模型调整文风、语法、文化细节甚至专业术语，大幅提高了诱饵内容的专业性与可信力，受害者在阅读时难以察觉。

在边界突破到内网驻留阶段，AI 技术被用于生成恶意代码和决策执行控制。2025 年出现在野样本使用 AI 动态生成指令。攻击者将 LLM 作为恶意代码的“实时命令生成器”，在执行过程中根据当前环境与攻击意图生成所需指令，使静态分析几乎无法定位关键行为。更进一步，攻击者尝试使用本地私有模型生成恶意样本核心功能代码，在动态生成核心功能的同时规避云端模型厂商的安全检测。

AI 也被用于高隐蔽性的执行决策控制，AI-Gated loader 则体现了“智能决策型恶意软件”的雏形。该类样本在执行 shellcode 前先收集系统遥测数据，包括运行进程、硬件特征与环境指标等，并将这些数据交由模型判断当前环境是否存在沙箱、EDR、调试器等分析迹象。当模型认为环境可疑时，恶意代码会选择推迟执行或完全拒绝执

行。这种通过模型做出执行决策的方式远比传统反沙箱技术更加灵活，不再依赖固定规则，而是呈现出基于上下文理解的环境适配性，提高恶意软件的隐蔽性与存活能力。

在数据窃取阶段，攻击者武器化本地合法 AI 工具，实现智能数据窃取。QUIETVAULT 是这一趋势的代表性样本。它借助受害者已安装的合法 AI CLI 工具，例如 Gemini CLI 或 Claude Code，通过自然语言描述让模型主动查找敏感文件的位置。这些 AI 工具可以迅速定位私钥、SSH 配置或云服务凭据等关键文件。整个过程与开发者日常使用 AI 搜索代码的行为极度相似，形成极高的隐蔽性。这种“利用合法 AI 工具的本地智能窃取”模式（LOLAI）正在成为一个新兴的强隐蔽攻击途径。

同时，AI 也在用于对抗安全分析本身，和实现自我变异逃避安全分析。随着越来越多安全厂商将 AI 用于样本分析与行为判断，攻击者通过提示操控来干扰分析结果。攻击者预先植入固定的提示词注入内容，试图误导基于 AI 的自动化分析系统，使其给出错误结果或安全评级。今年观察的恶意样本已经开始展示“自我变异恶意代码”的雏形。该样本会在运行时调用大模型 API，将自身的代码重写为功能一致但结构不同的版本，从而在每次执行前生成新的代码变体。一旦此类技术成熟，将使恶意软件具备无限多态性，现有基于签名、模式或固定 TTP 的检测体系将失去效果。

从 Deepfake 驱动的可信社工，到生成式 AI 完成文本级欺骗；从动态代码生成，到 AI-Gated 的环境感知执行；从滥用合法 AI 工具的本地智能搜索，到具有自我变异能力的未来恶意代码，可以清晰地看到：AI 已从辅助角色演进为网络攻击体系中的核心武器与重要部分，而整体攻击生态也正围绕 AI 的能力边界持续扩张。

3.8.2 国内外发展现状

3.8.2.1 从传统 AEV 到智能渗透

Gartner 将对抗性暴露验证（Adversarial Exposure Validation，AEV）定义为能够持续、连贯且自动化提供攻击可行性证据的技术体系。该技术是一种以结果为导向的评估技术，其通过模拟攻击场景，并根据实施结果证明潜在攻击手法如何成功突破企业防护体系、规避现有安全防护与检测机制，从而验证安全漏洞的真实存在性与可利用性。

根据 2025 年发布的《Gartner® Hype Cycle™ for Security Operations》报告，对抗暴露验证在技术成熟度曲线上处于“技术萌芽期”阶段，但相较于 2024 年定位发生上移，技术关注度与成熟度持续提升，仍作为安全运营领域持续威胁暴露管理（CTEM）战略的关键支柱。

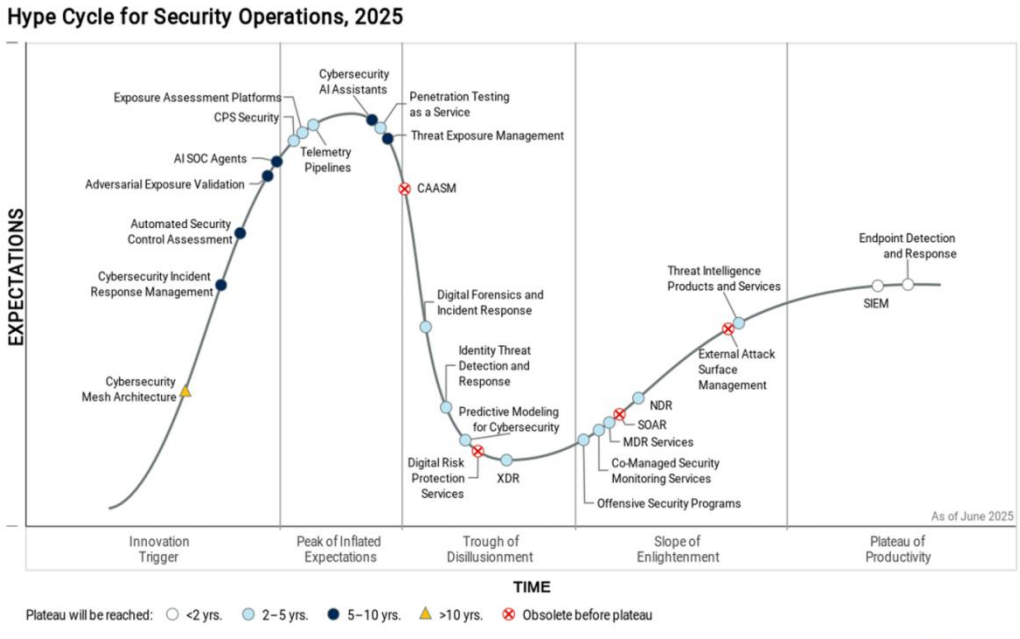


图 3.23 Gartner® Hype Cycle™ for Security Operations

3.8.2.2 政策、法规与标准对 AEV 的规范与驱动

剖析全球主要法规与标准，可发现其核心要求与 AEV 的“自动化、持续性、证据化”核心理念高度契合，从而在根本上推动了该技术的发展。

欧盟《数字运营韧性法案》（DORA）对安全运营韧性、渗透测试和网络攻击模拟提出了高要求，直接推动了金融服务业对 AEV 的需求。美国网络安全与基础设施安全局（CISA）要求美国联邦机构为云租户部署安全配置基线自动化评估工具，并开始持续报告合规要求。欧盟网络安全局 2025 年发布的技术实施指南建议，实体应确保网络和信息系统接受持续测试，尤其是在采用持续集成/持续部署（CI/CD）实践的环境中。

无论是 DORA 法案强化的“模拟测试”要求，还是 CISA 指令明确的“自动化评估”与“持续报告”任务，其本质都是要求组织从孤立、手动的安全检查，转向集成、自动化的攻击可行性验证。这一转变使得传统工具难以胜任，从而为能够提供持续性证据链的 AEV 解决方案创造了结构性市场机遇。

3.8.2.3 智能渗透是 AEV 演进的必然方向

“智能渗透”正是实现 AEV 的重要形式，代表了 AEV 技术从规则驱动的自动化向 AI 驱动的智能化迈进。智能渗透平台不再是简单的脚本执行器，而是利用 AI/ML 技术实现自动化决策和自适应攻击路径规划。

智能渗透的关键优势在于其“情报驱动”的决策引擎和运行时验证机制。一方面，平台利用上下文推理和威胁情报，确定通往关键资产的最隐蔽攻击路径。另一方面，智能渗透工具通过实际的漏洞利用来验证问题，而非仅仅根据系统响应猜测漏洞是否存在。这种方法极大地降低了误报率，使安全团队能够专注于经过验证的、真正的风险。

3.8.2.4 全球厂商布局与差异化竞争

全球 AEV/BAS 市场在 2025 年正加速整合，竞争的焦点集中在 AI 驱动的自主性和平台生态闭环能力。

大型安全能力提供商正将 AEV 能力作为暴露面管理和 XDR 平台的关键反馈层和验证层。例如 CrowdStrike 将其 AEV 能力深度集成到 Falcon Exposure Management 平台中，利用 AI 进行威胁驱动的风险优先级排序，并通过攻击路径分析来可视化横向移动的潜在路径。

以 Picus Security 和 Cymulate 为代表的传统 BAS 领导者，继续提供安全验证和防御优化的端到端解决方案。XBOW 等 AI-Native 智能渗透平台强调其 AI 驱动的渗透测试能力，旨在提供人类水平的安全测试。这些平台的核心优势在于其高度的自主性和高精度验证。

3.8.3 技术发展观察

3.8.3.1 自主性跃升：从工具辅助到全流程 AI 决策的质变

传统渗透测试高度依赖人工经验，需安全专家逐环节执行“信息收集-漏洞探测-利用验证-报告输出”的长流程，耗时久、成本高且难以规模化的内生缺陷。随着大模型（LLM）与 Agentic AI 技术的深度融合，自动化工具率先在 Web 应用渗透与全流程后渗透两大高频场景实现突破，标志着安全攻防正从“自动化工具辅助”向“全流程自主决策”演进。

Web 自动化：AI 代理模拟人类黑客的高效验证

AI Web 渗透是以大模型为核心驱动的智能化 Web 安全测试方式。它通过对 Web 应用的页面结构、数据流向、参数关系、鉴权流程以及业务语义进行深入理解，从中推理出潜在的攻击入口，并自动构造可执行的 Payload、绕过策略与利用链，形成针对 Web 应用的全方位智能渗透行为。其目标是在 Web 和 API 层面提供比传统扫描器更深度、更具语义理解、更贴近真实攻击者行为的渗透能力，尤其是在包含复杂逻辑、动态内容或强耦合前端框架的现代 Web 体系中弥补自动化测试的盲区。

全流程渗透：从单点利用到端到端自动化

AI 全流程渗透是以大模型为策略中心，自主规划执行从外部入口到内网渗透、权限提升、横向移动、持久化、数据影响分析的完整攻击链。其目标不是单点式漏洞验证，而是让 AI 具备完整攻击者视角，通过自主资产建模、漏洞链路关联、攻击路径规划、后渗透操作选择以及多工具编排来模拟真实威胁行为，实现对组织整体安全性的系统级验证。AI 全流程渗透覆盖 Web、系统、网络、身份与权限等多个维度，是对真实攻击者跨阶段、跨系统、跨协议行为的深度复刻，强调的是从入口开始到业务影响为止的全链路渗透能力。

3.8.3.2 范式转型：人机协同/纯 AI 驱动的未来形态

当前，渗透测试技术正经历从“自动化执行”向“全流程自主决策”的跨越式发展。**“数字渗透专家”模式以全流程自动化为核心，通过标准化基础安全验证流程实现规模化覆盖，成为基础安全检测的普适性解决方案。** 凭借自动化扫描与实时验证能力，无需人工调度即可每日执行测试，大幅缩短系统暴露窗口，在效率与覆盖范围上远超传统人工测试，显著降低安全运营成本。尽管这类工具受限于复杂业务逻辑理解、伦理判断及创造性思维（如针对定制化防御的绕过策略）能力不足，尚无法完全取代高级渗透

测试专家。但该模式通过标准化基础安全验证流程，使常态化、可持续的安全检测成为可能。

“AI 增强型高级渗透专家”模式以人机协同为基石，通过 AI 赋能与人类专家智慧的深度结合，成为攻克复杂安全挑战的核心手段。 对于需要深度攻击路径规划、定制化攻击策略或对抗高级防御体系的复杂任务，AI 转型为“超级副驾驶”，承担数据预处理、初步侦察、攻击面映射等耗时环节；人类专家则聚焦策略制定、创造性绕过及伦理决策等需要深度思考的环节。AI 可快速生成数百个潜在攻击路径供筛选优化，辅助识别逻辑漏洞并验证其实际影响。这种“AI 负责广度与速度，人类专注深度与伦理”的协同模式，既保留了人类专家的创造性优势，又通过技术放大了执行效能，成为应对未来高级威胁的核心解决方案。

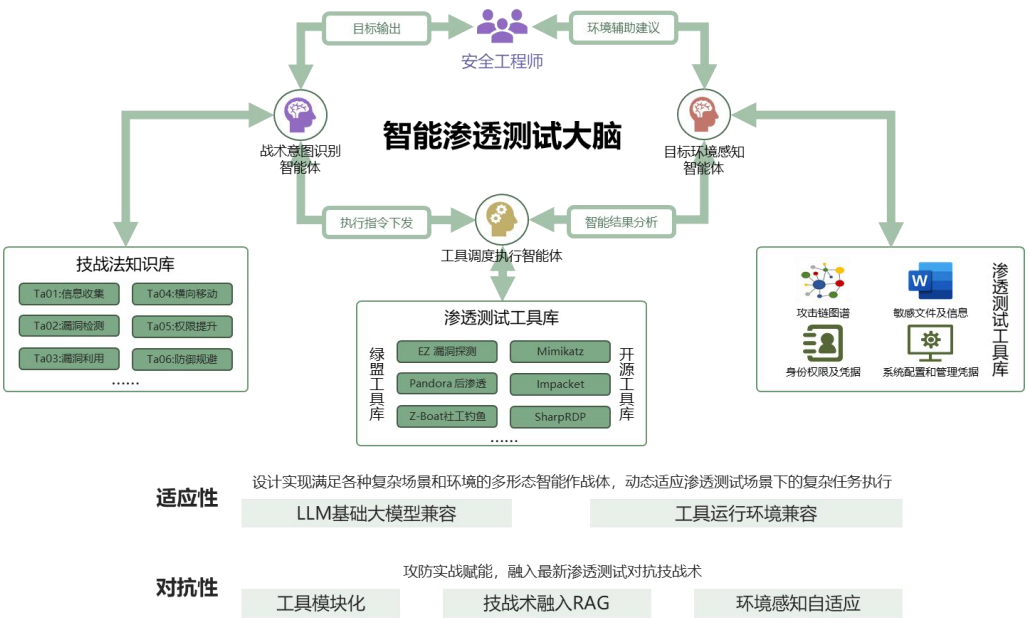


图 3.24 人机协同的 AI 智能渗透测试大脑

3.8.3.3 生态进化：低门槛 AI 调用推动智能渗透普惠化

近一年来，随着 MCP（模型上下文协议）的广泛应用和 AI 调用门槛的降低，智能渗透技术正以前所未有的速度向普惠化方向发展。传统上，高级渗透测试需要安全专家具备深厚的技术积累和丰富的实战经验。而现在借助 MCP 协议，即使是初级工程师也能通过 AI 的辅助，完成复杂的渗透测试任务。这种技术普惠化的趋势极大地提升了网络安全防护的效率和覆盖面。红队可以 AI 辅助更快地发现系统漏洞，蓝队也能借助相同的技术加强防御能力。安全测试的门槛降低，意味着更多的组织能够负担得起专业级安全评估，对网络安全防护水平提升有积极意义。

3.9 万物互联的安全

近几年，随着互联网、5G、传感器、智能芯片等技术的发展，越来越多的非传统 IT 设备连上了网络，有了“智慧”，形成了勃勃生机、万物互联的繁荣场景。而不同于计算机设备，这些联网的“新设备”就像刚刚接触互联网的懵懂少年一样，很容易遭受网络上邪恶黑手的攻击影响。因此绿盟科技持续关注万物互联的安全技术研究。今年我们针对车路互联、低空经济、卫星互联三个热点领域，汇总了我们的观察与研究成果，希望帮助这些新设备筑起新的防线。

3.9.1 车路互联

纵观网络安全行业，通常是由网络安全事件，萌生网络安全需求，推动法规、标准落地，催生网络安全产品，监管网络安全生态。车联网安全行业也同样如此。本章首先介绍两起网络安全事件，然后从法规视角介绍国内外发展现状，最后总结车联网安全三大关键技术。

3.9.1.1 国内外热点事件

近两年，有两起值得关注的汽车安全事件值得关注。

第一个事件发生于 2024 年 5 月，四川省成都市一辆渣土车发生交通事故，导致 1 人死亡 7 辆车受损。交警在检查该肇事渣土车卫星定位装置时发现，动态监控系统记录数据存在异常。最终专案组成功破获全国首例篡改货车车辆卫星定位装置数据系列案，共对 65 人采取刑事强制措施，查处涉事企业 109 家，扣押涉案卫星定位装置终端 1703 台。

第二个事件是起始于 11 月 28 日，俄罗斯多个地区陆续传出保时捷瘫痪的消息。官方表示，这是出厂时安装的 VTS 卫星安防系统出现问题，误判车辆有被盗风险，直接启动了最高级别锁死程序。

这两起事件给我们两个经验教训，第一，汽车数据安全治理是重中之重，车企、监管机构需要联合对汽车数据进行全面监管，防止不法分子从中破坏和牟利。第二，车企的防盗与驾驶的设计需要结合用户驾驶，先保证用户的驾驶安全，再保证车辆自身的安全。

3.9.1.2 国内外发展现状

2015 年前后，全球发生了多起智能网联汽车被远程攻击的事件，以及车联网、物联网安全漏洞频发。随后，汽车安全相关的法规和标准陆续出台，使汽车网络安全从边缘议题变成国际与中国的立法重点与媒体焦点。

国际方面，美国在 2016 年发布 SAEJ3061 和 NHTSA 网络安全指南，首次提出系统性安全框架。2020 年，联合国欧洲经济委员会出台 R155（整车网络安全）和 R156（软件更新安全），并在 2024 年全面实施。2021 年，ISO/SAE21434 发布，为

汽车电子电气系统全生命周期提供统一的网络安全工程要求，成为全球合规的核心技术标准。

国内在 2024 年，发布了《汽车整车信息安全技术要求》（GB44495-2024），这是国内首个整车信息安全强制性标准，并配合《汽车数据通用要求》(GB/T44464-2024)及各地实施细则，标志着中国汽车信息安全进入法规化阶段。

表 3.7　近年发布的车联网相关法规标准

时间	法规/标准	要点
2016	SAE J3061	首个系统性推荐实践标准，提出生命周期安全框架
2016	NHTSA 发布网络安全最佳实践指南	针对现代车辆的网络安全风险，提出政策指导与合规参考
2020	UNECE R155（整车网络安全）、R156（软件更新安全）	2024 年 7 月起全面实施，成为全球强制法规
2021	ISO/SAE 21434:2021 《道路车辆—网络安全工程》	取代 J3061，提供统一的工程要求，与 UNECE R155/R156 对接
2024	《汽车整车信息安全技术要求》 《汽车数据通用要求》	这是国内首个整车信息安全强制性标准

随着中国 GB44495-2024 等法规落地，整车厂必须建立覆盖全生命周期的信息安全管理体系。具体需求包括风险评估与威胁建模、安全编码与 AUTOSAR 安全模块集成、合规文档与型式认证准备，以及 OTA 升级的签名验证、版本控制和日志管理。这些工作与企业内部架构和流程紧密相关，必须由 OEM 主导完成。

与此同时，法规也催生了需要专业安全厂商支持的需求，例如车载 IDS/IPS 系统、SOC 平台与威胁情报服务、整车渗透测试与红队演练、密码学与密钥管理解决方案，以及数据加密与隐私保护工具。这些需求涉及专业安全技术和持续监控能力，通常由

网络安全厂商提供更高效的解决方案，与 OEM 形成分工协作，共同满足法规和标准的合规要求。

表 3.8 网络安全厂商与整厂/OEM 的安全分工协作

类别	由整车厂/OEM 实现	适合网络安全厂商配合
管理体系	全生命周期安全管理体系	安全运营中心 (SOC)
软件开发	安全编码、AUTOSAR 安全模块集成	提供安全库、加密算法支持
测试验证	HIL/SIL 安全测试、合规文档	渗透测试、红队演练
通信安全	ECU 内部通信加密、防篡改	IDS/IPS、车载防火墙
OTA 升级	签名验证、版本控制、日志	OTA 安全网关、流量监控
数据安全	本地数据保护、隐私合规	数据加密、脱敏、访问控制
威胁情报	内部风险评估	外部威胁情报订阅

3.9.1.3 技术发展观察

在网络安全的法规要求之下，不论是整车厂、科研机构还是安全厂商，都应该重视以下三个关键技术。

第一是安全测试，可以分为合规检测和渗透测试两大部分。合规检测是汽车上线之前必须要通过的测试过程，如果无法拿到合规检测相关的证书，则影响汽车上线售卖。渗透测试是高于合规检测要求的测试过程，在对网络安全要求较高的车辆（如运营车辆、特种车辆、中高端的民用车辆）中，整车厂愿意对车辆进行深度的渗透测试，以保障更高级别的车辆和企业的网络安全。

第二是入侵检测，由于车端的 ECU 种类多样，导致软件、网络的复杂度高，专业领域特征显著。若要覆盖所有的攻击入口，稳定检测攻击行为，那车载入侵检测系统的设计与实现会变得尤为艰难，这对车载入侵检测系统提出了以下细化的要求，其中前三个方面的功能，对车载入侵检测系统的要求更为严格。

表 3.9 车载入侵检测系统技术要求

能力要求	描述
入侵检测能力	实时检测攻击，支持基于规则和行为异常的可选告警模式。
软件环境适应	兼容嵌入式 OS，使用 C/C++，极小资源占用。
覆盖范围	监控 OS、CAN、Ethernet 和无线接口，检测内部/外部威胁。
认证与授权	集成车辆身份验证，防止未经授权访问。
自身安全性	使用 HSM/TPM 保护，防篡改。
合规性	符合 UN R155、ISO/SAE 21434，确保数据隐私。
实时性	处理高频数据流，支持低功耗。
鲁棒性	在嵌入式软件/复杂网络环境下稳定运行，适应动态拓扑。
可扩展性	模块化设计，便于添加规则或集成 ECU。
容错性	冗余机制，防止单点故障，支持自检和更新。
数据保护	加密通信/存储，最小化数据收集。
独立性	独立模块部署，避免依赖单一 ECU。
网络环境复杂性	处理异构协议、时钟同步和 EMI，支持 OTA 更新。
软件兼容性	兼容 AUTOSAR 等 OS；使用 C/C++，支持 OTA，避免冲突。
网络兼容性	兼容 CAN FD 等协议。
测试与验证兼容性	跨平台验证，包括虚拟汽车前端模拟。

第三是虚拟汽车技术。当前汽车环境大多以实车、实物零部件的形式提供，如果需要应对班级化教学或者竞赛，这种无并发性的物理环境将大大影响教学效率和比赛的公平。虚拟化成为这一难题的唯一解决途径。绿盟将汽车虚拟化细化到操作系统和 CAN 总线通信的级别，提供灵活的电子电气架构设计方案，可将不同品牌的汽车电子电气架构虚拟化，供教学、竞赛、科研使用。从虚拟化和实物两方面的对比来看，各方面优势如下所示。

表 3.10 虚拟汽车环境优势

方面	实物环境	虚拟环境
成本	高昂（硬件、场地）；一次性投资大	较低；可扩展，无物理损耗
安全性	潜在风险（如碰撞、故障）	无物理风险；模拟无实际伤害
真实性	高真实度（实际车辆、传感器）	嵌入式操作系统的电子电气架构仿真
可重复性	难以精确重复（环境影响，如天气）	高可重复性（相同输入相同输出）
灵活性	受物理限制（场地、设备可用性）	高灵活性（易修改场景、参数）
维护	需要定期保养；故障修复耗时	软件更新即可；无硬件磨损
复杂度	受限于物理资源；测试大型网络复杂	可模拟大规模系统；易添加复杂性
实时性能	实时，但受物理延迟影响	实时模拟；可加速/减速时间
数据收集	实际传感器数据；可能受干扰	精确控制数据；易生成大量测试数据

虚拟环境相比实物环境来看，虚拟环境具备无可比拟的大规模、高并发、高灵活性，以及超低成本适合科研、教学以及竞赛的场景，结合嵌入式操作系统级别的电子电气架构仿真的实现，在软件架构上也向实车看齐，与真车差异越来越低。虚拟化的汽车环境，将成为未来研究机构进行汽车安全研究的重要形态。

3.9.1.4 总结

纵观车联网法规发展史，我们看到车联网安全的三大需求，分别是安全测试、入侵检测以及虚拟汽车技术。车联网整车安全标准落地，主机厂面临合规压力，合规、渗透和车载 IDPS 等需求随之涌现。另外，高校尤其是传统网络安全和汽车工程这两大学院在车联网安全的教学与科研体系也亟需建立。

3.9.2 低空经济

3.9.2.1 热点安全事件

2025 年，低空经济作为国家战略级新兴产业加速崛起，无人机物流、应急救援、农业植保等应用场景持续拓展，成为数字经济与实体经济深度融合的重要载体。伴随 5G、AI、卫星通信等技术深度融入，低空领域安全风险同步进入高发期，网络攻击、系统漏洞、非法改装、运行碰撞等安全事件呈现多样化、复杂化特征，既威胁公共安全与生命财产安全，也对产业监管体系和技术防护能力提出严峻挑战，为低空经济高质量发展敲响警钟。

针对无人机固件分发系统的网络打击

2025 年 7 月，为俄罗斯军方提供定制化无人机固件的开发团队证实，其“1001”固件专属分发基础设施遭黑客入侵。安全专家 Oleg Shakirov 分析，攻击者大概率通过远程命令注入或 API 接口突破服务器防线，攻击手法专业、目标精准，疑似情报级别的对手所为。

“1001”固件采用闭环分发模式，通过前线部署的数百个加密认证“无人机服务终端”，为每日 500 至 560 架无人机提供“武器化”改装更新服务。此次攻击通过侵入远程服务器系统，在无人机更新终端设备推送伪造信息，最终导致整个固件更新体系瘫痪。前线无人机无法获取软件升级，直接影响装备交付节奏与战场生存能力。该事件暴露了中心化架构在关键基础设施中的脆弱性，也为低空装备的安全防护提供了重要警示。

非法破解无人机飞行控制系统

2025 年 9 月，成都公安公布“净网 2025”专项行动破获的一起非法破解无人机飞行控制系统案件。犯罪嫌疑人陈某利用系统漏洞，擅自解除无人机限高、限速等安全防护设置。该行为违反《中华人民共和国刑法》第 286 条规定，涉嫌构成破坏计算机信息系统罪，已移送检察机关审查起诉。

根据《无人驾驶航空器飞行管理暂行条例》，无人机地理围栏、限高及禁飞区限制属国家强制性安全措施。非法修改系统功能需依法追究刑事责任。该案反映出低空装备安全基线面临的现实挑战，也体现执法部门维护空域秩序的决心。

3.9.2.2 国内外发展现状

作为新质生产力的典型代表，低空经济正成为我国各地争相布局的新兴产业新赛道，在政策红利与技术创新的双重赋能下，产业发展全面提速，市场规模迎来爆发式增长的黄金期。从国家战略规划到地方实践落地，低空经济的发展框架不断清晰，应用场景持续丰富，一个规模化、规范化的产业生态正在加速形成。

2025 年政府工作报告：从“培育”到“壮大”的关键跨越

低空经济以低空飞行活动为核心，融合无人驾驶飞行、低空智联网等前沿技术，带动低空基础设施、飞行器制造、运营服务等全产业链发展。据中国民航局预测，这一新兴市场的规模 2025 年预计达到 1.5 万亿元，到 2035 年更有望突破 3.5 万亿元。

2025 年 3 月 5 日，十四届全国人大三次会议政府工作报告中明确提出“培育壮大新兴产业、未来产业，开展新技术新产品新场景大规模应用示范行动，推动商业航天、低空经济等新兴产业安全健康发展”，这是低空经济继 2024 年后再次被写入政府工作报告，表述从“积极培育”升级为“培育壮大”，标志着该产业已从概念探索阶段迈入规模化、规范化发展的新阶段，而“安全健康发展”作为关键词，既呼应了 2024 年末低空经济司组建后聚焦“安全监管”的工作重点，也凸显了国家在推动产业扩张的同时坚守风险防控底线、实现“飞起来”与“管得住”并重的发展思路。

国家发改委专项部署：安全发展路径的细化落地

政策导向的细化落地紧随其后。国家发改委政策研究室副主任李超就低空经济发展做出专项回应，将“安全是低空经济发展的首要前提”明确为核心原则，为产业发展

划定了清晰的实践框架。针对当前行业存在的“黑飞”等安全隐患，发布会明确提出“坚持管得住才能放得开”，严厉打击驾驶员无证飞行、航空器未取得适航证、飞行活动未报批等违法违规行为。

美国：技术引领与市场化驱动并行

在政策构建中，美国通过顶层立法、专项监管与安全防御的多维度协同，为产业发展筑牢保障网络。其中，《先进空中交通协调及领导力法案》作为重要立法基础，明确要求交通部统筹安全监管、基础设施建设、网络安全及联邦投资等跨领域事务，核心目标在于系统性构建载客 AAM 飞机的研发与运营生态。

针对无人机滥用可能带来的风险，《恢复美国空域主权》行政命令形成了有效的安全保障补充。该命令通过设立专项任务小组、强化关键区域飞行限制、加重违法飞行法律责任等一系列防御性管控措施，构建起全方位的空域安全保障体系，为产业健康发展规避风险。

3.9.2.3 技术发展观察

无人机发展的系统之困与反制之殇

无人机作为低空运行体系中的核心载体，无人机系统的安全防护能力尚存不足，尤其体现在其核心组件本身的脆弱性，以及反制技术发展所带来的“双刃剑”效应上。

从系统构成来看，无人机已从单一航空器演进为集感知、决策、通信与云服务于一体的智能网络节点。这一转变在提升作业能力的同时，也扩展了其潜在的攻击面。部分无人机厂商搭建的云服务平台承载着飞行计划管理、调度与实时数据传输等关键功能，若安全措施不到位，不仅可能导致飞行计划被恶意篡改，影响任务执行。更可能在云服务被攻陷后，使攻击者获得对无人机的批量干扰甚至非法控制能力，突破电子围栏闯入敏感区域。

在软件开发与维护层面，厂商的源代码若保护不足，一旦被窃取，可能暴露系统深层的逻辑缺陷或加密机制，为后续定向攻击提供信息基础。此外，用于固件升级的OTA 服务器若与代码服务器同被攻陷，极易形成连续的攻击链路。攻击者既可获取关键代码，又能向终端设备批量推送恶意固件，从而实现对目标机群的规模化干扰与控制。这些风险点相互关联，使得无人机在单机运行与集群协同中均面临数据泄露或控制权丢失的隐患。

另一方面，为应对“黑飞”“乱飞”等威胁，现代低空防御系统通常集成雷达、无线电侦测及导航诱骗等功能，实现区域级无人机动态感知与管控。然而这类反制系统若因配置缺陷或通信协议漏洞遭到入侵，攻击者就可能篡改其识别逻辑、伪造目标信息，甚至滥用干扰功能影响合法无人机的正常运行。更值得警惕的是，被操控的反制系统可能被用作“掩护工具”，为恶意无人机规避监管提供通道。

低空经济网络安全正处在一个风险与防御相互竞逐的发展阶段。无论是无人机自身在软硬件、通信及云端服务中存在的薄弱点，还是反制系统网络化之后带来的双向风险，都呼吁更加系统化、前置化的安全设计。未来，有必要从多维度构建纵深防御体系，同时推动反制系统自身的安全准入与行为审计机制，在护航低空经济创新应用的同时，筑牢其健康发展的安全底座。

低空经济场景持续扩展，新旧融合下的安全隐忧

与传统高空民航不同，低空经济以城际载人运输和高效物流为主要运力，致力于构建灵活、短距、高频次的空中交通体系，成为连接地面交通网络与高空航空运输的重要补充。这种空域结构与运营模式的拓展，不仅丰富了低空经济的应用场景，也提升了空域资源的利用效率。

在低空经济体系快速发展的同时，其中网络安全问题已成为制约低空经济健康发展的关键因素。为适应低空运营环境的特殊需求，新型航空器普遍集成自动驾驶、5G-ATG 通信等先进技术，并通过与人工智能、大数据等前沿科技的深度融合，实现

精准调度与高效运行。这些数字化、智能化功能的引入也扩大了系统的网络攻击面。与此同时，为维持与现有航空体系的兼容性，航空器仍保留了传统民用航空通信功能。老旧功能与新型技术的混合应用，形成了复杂且异构的技术生态，带来了前所未有的网络安全挑战。当前，低空经济网络安全技术虽已在加密通信、身份认证等领域取得一定进展，但仍需在智能化防护、跨领域协同、安全标准构建等方面持续突破，以夯实产业规模化发展的安全基础。

3.9.3 卫星互联网

卫星互联网正从“新基建”跃升为全球关键信息通道，其在军事、通信、定位等领域中的支撑作用日益凸显。因此，必须把技术防护、供应链韧性和跨国监管三位一体的安全体系嵌入卫星互联网的全生命周期，确保其战略价值不被破坏性攻击所削弱。

3.9.3.1 热点安全事件

2025 年以来，全球范围内卫星互联网相关安全事件频发，攻击手段呈现多元化、精准化特征，涵盖电子干扰、网络入侵、恶意软件植入等多种形式，对卫星通信链路、导航服务连续性及数据安全构成严重威胁。

俄乌冲突背景下卫星通信定向电子干扰事件

2025 年初，乌克兰多家卫星互联网服务提供商遭遇大规模电子干扰，导致部分区域服务完全瘫痪。此次干扰被怀疑源自俄罗斯，受影响卫星主要承担军事通信与关键基础设施数据传输功能，直接削弱了乌克兰战场指挥控制能力。作为现代战争中电子战的典型应用，该事件印证了卫星通信链路在对抗环境下的易受攻击性。

伊朗油轮卫星终端弱口令致大规模网络入侵事件

2025 年 3 月，黑客组织 Lab Dookhtegan 针对伊朗国家油轮公司及伊斯兰共和国

航运公司的 116 艘油轮发起网络攻击，造成卫星通信系统全面瘫痪。攻击者利用默认弱口令获取 iDirect 卫星通信终端 root 权限，部署自动脚本擦除设备存储器。此次攻击疑似有国家行为体支持，其攻击手法与 2022 年 Viasat 卫星网络攻击高度同源，暴露了卫星终端设备运维管理漏洞带来的规模化安全风险。

美国军用卫星系统遭“轨道阴影”恶意软件植入攻击

2025 年 5 月，美国多套军用卫星系统遭遇针对性网络攻击，疑似国家级黑客组织利用卫星老式指令接口漏洞，在系统更新阶段植入“轨道阴影”恶意软件。攻击者通过劫持星地通信链路搭建后门，企图干扰卫星定位与时间同步核心功能，且攻击波及美军依赖的商业卫星基础设施。美国国防部迅速采取隔离感染站点、升级加密协议等应急处置措施，未造成卫星失控，但该事件直接凸显了军用卫星系统接口安全隐患与供应链安全风险。

3.9.3.2 国内外发展现状

2025 年，在全球政策驱动下，卫星互联网产业正从单极主导向多极竞争演变。美、中、欧三足鼎立格局形成，在卫星互联网领域的立法与监管活动显著增强，核心聚焦于网络安全、数据主权、供应链韧性及关键基础设施保护。这些政策不仅深刻影响着卫星运营商、设备制造商和服务提供商的商业运营模式，也反映了在地缘政治竞争加剧和技术快速迭代的背景下，各国将太空安全提升至国家战略高度的共同趋势。

中国：将卫星互联网打造为国家“新基建”战略重要支柱

2025 年 3 月，工信部印发了《卫星网络国内协调管理办法（暂行）》。该办法的出台，旨在建立一个清晰、高效的国内协调机制，以解决不同卫星网络之间可能产生的干扰问题，确保国家卫星通信系统的安全、稳定和可靠运行。此举不仅是对国内频谱资源管理的一次重要完善，更是明确了卫星互联网作为国家“新基建”核心组成部分

的战略地位。为后续更大规模的卫星星座部署，奠定了坚实的基础。

2025 年 9 月 5 日，工信部正式印发了《关于优化业务准入 促进卫星通信产业发展的指导意见》。该指导意见的发布，标志着中国卫星互联网建设进入了政策与市场双轮驱动的加速期，旨在激发市场活力，促进技术创新，并最终构建一个覆盖全球、服务高效、安全可靠的天地一体化通信网络体系。预计到 2030 年，中国卫星互联网市场规模将实现指数级增长，并在应急通信、海洋渔业、航空互联网、物联网等垂直领域实现广泛应用，有力支撑经济社会的高质量发展。

随着卫星通信技术，特别是手机直连卫星技术（D2D）的快速发展，国家互联网信息办公室于 2025 年 4 月正式发布了《终端设备直连卫星服务管理规定》。该规定是中国首部专门针对直连卫星终端设备管理的法规，目标是确保此类新兴服务的安全、有序发展，并将其全面纳入国家现有的网络安全与数据治理法律体系之下，标志着我国针对卫星互联网服务的监管进入了精细化、制度化的新阶段。

在产业发展方面，随着中国卫星互联网“国家队”——中国卫星网络集团有限公司的部署进程显著加速，国内三大运营商也开始积极布局，探索如何构建天地一体化网络的商业化生态。这一趋势表明，中国的卫星互联网发展正从单纯的技术验证和星座建设阶段，迈向与地面网络深度融合、商业化生态逐步构建的新阶段。

美国：降低企业准入门槛，聚焦商业太空系统与供应链安全

美国联邦通信委员会在 2025 年 10 月份提出的“Space Modernization for the 21st Century”改革，展现了其通过优化监管来保持产业领先地位的战略思路。其核心目标是简化卫星许可流程、强化美国在全球太空经济领域的竞争力，并将国家安全纳入监管改革的核心驱动力。通过创造一个更具吸引力和效率的监管环境，加速本国企业的星座部署，并吸引更多国际航天企业在美国开展业务，从而进一步集聚全球顶尖的航天技术和人才，强化其在全球太空经济中的核心枢纽地位。

2025 年 9 月，SpaceX 向美国联邦通信委员会提交了一份申请，请求为其 Starlink 星座增加多达 15,000 颗新卫星，以支持其即将推出的直连设备服务。这一申请是 SpaceX 在卫星互联网领域迄今为止最大规模的扩张计划之一。

欧洲：战略自主计划推动产业链协同发展，提出统一安全框架

欧盟在 2025 年推出 IRIS²计划，明确表达了其追求“战略自主”的决心。长期以来，欧洲在卫星通信领域高度依赖美国，这不仅带来商业上的不确定性，更在数字主权和安全层面引发担忧。IRIS²计划的启动，标志着欧盟将构建一个由自己掌控的、安全可靠的卫星通信基础设施，确保其政府、军队和关键基础设施的通信安全。更重要的是，IRIS²计划将极大地推动欧洲本土航天产业链的发展，从而在全球卫星互联网的价值链中占据更有利的位置。

在政策方面， 2025 年 6 月欧盟委员会正式提出了《欧洲空间法案》的立法提案。《欧洲空间法案》在网络安全方面的规定与欧盟现有的《网络和信息系统安全指令》和《网络韧性法案》的目标保持一致，将空间系统明确视为欧盟关键基础设施的一部分，确保包括卫星导航、地球观测和安全通信在内的关键空间基础设施具备韧性、互操作性，并处于欧盟的主权控制之下。该法案特别强调了对关键空间系统的强制性安全要求，其中网络安全是重中之重。

3.9.3.3 技术发展观察

全球卫星互联网建设进入高速发展阶段，在这一进程中，网络安全技术的演进与整个产业的壮大相互呼应，既得益于基础设施的快速迭代，也面临着由此催生的新型挑战。从顶层设计到具体实现，卫星互联网正处于一场深刻变革之中，其网络安全体系的构建也必须在动态发展中持续调整与加强。

火箭回收、模块化卫星设计、批量制造等关键技术取得突破，大幅降低了单星成本和发射门槛，推动大规模星座部署从构想走向现实。这种规模化、网络化的发展趋

势，客观上要求网络安全防护不单要考虑“单星防护”，还需要转向覆盖整个星座的系统性安全架构设计。大量在轨节点构成一个复杂空间网络，其内生安全、星间链路安全以及星地一体化安全传输成为新的技术焦点。

与此同时，国际电信联盟（ITU）所遵循的“先到先得”原则导致轨道与频率资源的国际竞争日趋激烈，这不仅涉及商业利益，更牵涉国家空间安全与网络主权。对我国而言，加快推进自主星座计划，提升在轨卫星的数量，是确保我国在未来空间网络治理中拥有话语权和规则制定参与度的关键。从网络安全角度看，掌握自主可控的轨道与频率资源，是构建安全可靠卫星互联网服务的根本前提。

然而，产业高速发展的背后，研发周期的大幅压缩与技术迭代的加速也带来了严峻的网络安全挑战。当前商业航天为抢占市场，将研发周期压缩到数年甚至更短，这可能使得安全设计和测试验证环节被削弱。软件定义卫星和通用化平台在提升系统灵活性的同时，也引入了更多通用软硬件漏洞风险。大规模星座的自动化运维与管控高度依赖地面站和运控系统，一旦这些系统被攻破，可能引发连锁反应，威胁整个星座的稳定运行。此外，低轨卫星通信信道的开放性以及星上处理能力有限等特点，使其更容易遭受无线注入攻击、数据窃取、协议欺骗或拒绝服务等威胁。

随着空天地一体化发展，太空也成为重点的主权领域和攻防焦点。美军定义了太空轨道战、太空电磁战和太空网络战。开发了 SPATAR 框架进行太空战的建模框架，并举办了多次“黑掉卫星”比赛。

表 3.11　历年黑掉卫星比赛内容

赛季	比赛内容	队伍数量
HSA2020	恢复与失控卫星的正常通信；修复卫星，阻止卫星自转；恢复与有效载荷模块的通信；控制成像器，拍下月球图像。	8
HSA2021	控制失控卫星；开启卫星主要通信载荷；利用漏洞向其他队伍发送命令注入的攻击型数据包。	8

HSA2022	解密、FTP 服务器、加密关键数据、自定义网络服务器、杂项；系统漏洞、逆向分析、逆向令牌。	8
HSA2023	卫星操作、射频通信、漏洞研究和逆向工程等。	5

图 3.25 HAS2023 年比赛目标：MoonLighter 卫星

面向未来，必须始终坚持发展与安全并重的原则。有必要前瞻布局星载安全芯片、星间加密通信、人工智能驱动的威胁感知、自动化应急响应等关键安全技术，推动安全能力实现从“事后补救”到“原生内置”、再到“全程伴随”的深刻转变。

同时，为保护我国新一代卫星互联网安全，构建太空靶场并进行太空网络攻防演训、防御技术开发研究和人员培养也势在必行。建设太空网络蓝军依托太空靶场以攻促防，对卫星用户终端、卫星服务端、卫星运维、卫星测控、卫星网络和卫星本体进行威胁建模，针对太空网络体系的脆弱性和威胁采用 AI 智能体等新技术开发先进有效的防御手段。守护太空网络安全。

04

总结

2025 年 10 月二十届四中全会审议通过了《“十五·五”规划建议》，全文用 15 个部分 61 条内容，细致勾勒了 2026-2030 年的发展路径。纵观了《“十五·五”规划建议》，文件强调“统筹发展和安全”，“安全”二字已经取代“增长”，成为十五·五规划的灵魂之一。以前的规划，“安全”都只是附带的，今天的文件，经济安全、科技安全、能源安全、粮食安全，甚至文化安全。过去那种“发展是硬道理”的逻辑，已经悄然让位于“安全是硬前提”。

2026 年是国家“十五·五”的开局之年，我们在立足当前网络安全市场发展现状之外，还要关注网络安全行业发展的未来。而这除了要求我们网络安全企业奋进自强以外，也需要全社会各个领域的企事业单位共同探讨，共同进步。而这就是绿盟科技连续九年坚持编写年度网络安全综合报告的初衷。

本篇报告的三个篇章，在宏观视角篇我们看到了国内外网络安全政策的发展与变化。在安全态势篇，我们透过对网络资产、攻击动态等指标的动态观察，总结了当前网络安全现状。在技术发展篇，我们对十余个网络安全技术领域的现状及发展趋势进行了总结与预测。

没有交流就不会有进步，没有合作就不会有共赢。网络安全的技术发展与能力提升，离不开政府机构、广大企事业单位、行业企业、高校研究机构的共同合作努力。我们希望这篇年度报告能够为企事业单位安全建设提供参考，略尽绵薄之力；也希望能够抛砖引玉，与有识之士开展交流合作，百尺竿头更进一步。

欢迎领导专家、各界同仁继续对我们给予关怀和指导。让我们共同努力，启航十五·五。

NSFOCUS

THE EXPERT BEHIND GIANTS

巨 人 背 后 的 专 家

多年以来，绿盟科技致力于安全攻防的研究，
为政府、金融、运营商、能源、交通、科教文卫等行业用户和各类型企业用户，
提供具有核心竞争力的安全产品及解决方案，帮助客户实现业务的安全顺畅运行。在这些巨人的背后，他们是备受信赖的专家。

网络安全2026

扫描绿盟科技官方二维码
可在手机端直接观看报告电子书