数说安全

2025网络安全十大创新方向

Top Ten Innovative Technology Trends in Cybersecurity for 2025

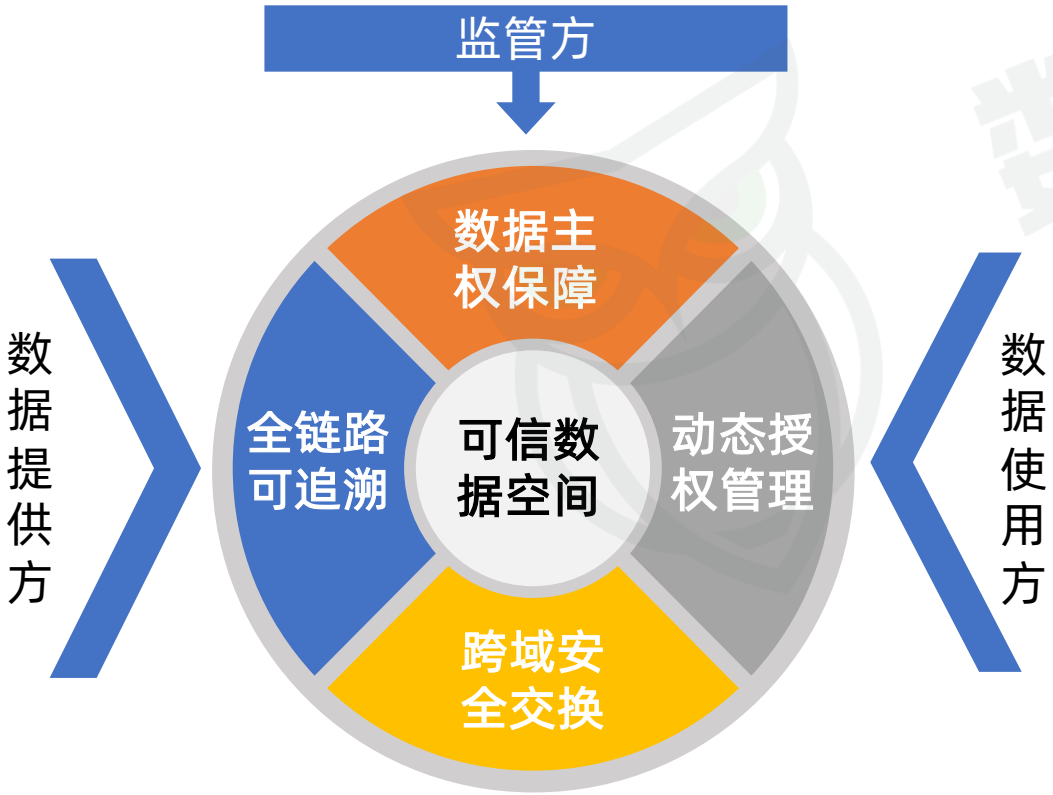
2025

数说安全

目录

可信数据空间

为加快构建以数据为关键要素的数字经济，国家数据局于2024年11月印发《可信数据空间发展行动计划（2024-2028年）》，提出了到 2028 年，可信数据空间运营、技术、生态、在“标准、安全等体系取得突破，建成 100 个以上可信数据空间的目标。可信数据空间（Trusted Data Space，TDS）是基于共识规则，联接多方主体，实现数据资源共享共用的一种数据流通利用基础设施，是数据要素价值共创的应用生态，是支撑构建全国一体化数据市场的重要载体。数据可信空间借鉴 IDS-RAM 与 Gaia-X Trust Framework，在连接器中固化身份、加密、使用控制和审计规则，让数据可控、可计量、可追溯”的条件下跨组织流通；同时与中国国家数据基础设施标准（NDI-TR-2025）对接，形成“平台 + 连接器 + 策略引擎”三层架构，支持数据目录发布、契约协商和联邦计算；其核心目标是打破“数据孤岛”，破解数据“不敢共享、不愿共享”的困局，构建“数据即资源、数据即能力”的协同创新平台。



核心能力

1、 可信身份与连接器管理
2、 数据使用控制与合约执行
3、 端到端加密与可信计算
4、 数据/算法可追溯与可审计
5、 语义互操作与目录编目
6、 跨空间协同与治理编排

关键挑战

1、 信任根跨域扩展难度大
2、 高强度加密产生性能瓶颈
3、 数据主权与责任划分不清楚
4、 技术标准缺失
5、 数据定价与价值分配机制不健全
6、 安全管理与持续运维难度大

应用场景

1、 可信制造业供应链协同质检与预测性维护
2、 跨医院医学影像及医疗数据联合建模
3、 电网数据与气象/需求侧信息的可信融合
4、 物流全链路运输数据实时共享与风险预警
5、 政务数据与社会数据的合规融通

典型厂商

火山引擎 · 中移互联网 · TAIJI 太极 · 数鑫 · 蚂蚁密算 · 清雁科技 · 一知安全 · 奇安信 · DualPi 双湃智安 · 易安联 · 阿里云 · TRANSWARP · Unicom Digital Tech 联通数科 · 锘崴科技 · 華控清交

火山引擎 Jeddak AICC

数说安全

火山引擎

火山引擎是字节跳动旗下云和AI服务平台，将字节跳动快速发展过程中积累的增长方法、技术能力和应用工具开放给外部企业，通过云和智能技术帮助企业构建体验创新、数据驱动和敏捷迭代等能力，推进企业AI转型，激发增长潜能。

火山引擎云安全依托字节跳动在安全技术上的实践沉淀，面向互联网、金融、汽车、大消费等行业输出云上安全能力。重点布局大模型安全、数据隐私安全、AI安全智能体等领域，致力于在AI时代，为企业大模型应用提供最全面的云上安全防护方案。

方案概况

方案介绍：

火山引擎Jeddak AICC 是基于机密计算、密码学应用以及信息流安全等隐私保护创新技术，面向云环境下敏感数据流转和应用安全的通用技术产品。

Jeddak AICC旨在帮助客户构建一套用户信任的安全计算服务，为端上用户提供安全可靠的云上运行环境，保障端云协作全链路的安全。

典型应用场景：可信AI推理

方案优势和用户价值

方案优势：

1、数据安全：基于硬件芯片级加密技术，实现全链路100%加密，保障用户数据安全。
2、模型安全：基于用户自主设置密钥，可对私有模型加密后完成在机密容器的部署。
3、能效平衡：相比明文环境，端到端推理性能损失10%以内，端侧用户感知可控。
4、透明可信：面向开发者和端侧用户提供云上服务的透明可信验证，实现自证清白。

用户价值：

基于火山引擎硬件级加密技术等核心能力，为客户构筑端云混合的安全防护体系，为个人提供全链路加密的可信私密云服务。

典型客户：

Lenovo 联想　nubia

中移互联网有限公司-中移数盾

中移互联网有限公司是中国移动面向互联网领域设立的专业子公司，2022年3月正式入围国资委“科改企业”名单。公司聚焦“信息+互联网”，利用中国移动规模、能力、体系优势，与中国移动的网络、技术、数据相融合，面向算力、大数据、人工智能等“9+6”科创战新领域布局，推动“号、卡、云、消息、通话”快速跃迁为新型信息基础设施，融合打造个人云服务基座、信创安全基座、超大AI应用基座，不断延展生态新边界，拓展信息服务新空间，激发CHBN价值增长新动能，为两个“做强做优做大”贡献更大力量。

方案概况

方案介绍：

中移数盾由中移互联网有限公司自主研发，针对个人数据要素流通场景提供身份认证、数据资产汇聚、授权管理、存证溯源的个人数据空间解决方案。产品采用实名号卡、证书签名、量子国密资源池与区块链技术，为实现政府、企业的个人数据可信流通保驾护航。

方案示意图/拓扑图：

方案优势和用户价值

方案优势与特点：

（1）高安全：采用国密算法与量子密钥技术，实现数据流转全程高强度加密，筑牢安全防线

（2）强合规：基于实名号卡与数字签名，确保用户身份及授权行为可验证、可追溯、符合法规。

（3）可监管：依托区块链技术实现数据流通全链路存证溯源，满足多方透明监管与审计需求。

（4）更便捷：提供移动认证能力，通过统一账号实现业务场景快速、灵活接入。

用户价值：

（1）保障用户主权和数据安全，确保数据在授权、流通、使用全链路精细化授权管控、安全传输和存证溯源。

（2）平衡数据利用效率与隐私安全保护的关系，激活个人数据要素潜在价值。

典型客户或目标客户：

属地数据局、数据集团、政企客户、公共企事业单位、金融医疗机构等。

AI赋能数据安全

数说安全

2025年，AI技术深度融入数据安全领域，推动防护模式从“被动响应”向“主动智能”跃迁。生成式 AI 与嵌入式机器学习 （ML） 已融入 数据安全态势管理 （DSPM--Data Security Posture Management）、数据泄露防护 （DLP）、数据目录和威胁检测等产品， AI赋能的数据安全体系实现了对海量数据的分类分级、自动化脱敏、敏感数据识别、泄露检测，威胁检测和实时响应，显著提升了对新型攻击的防御效率，例如在自动化敏感数据发现方面，基于自监督向量模型，跨本地、多云及SaaS环境自动识别敏感数据分布，消除数据盲区；智能语义分析，利用大模型语义理解能力，精准检测“影子数据集”与误标信息，提升数据分类分级准确性；实时行为关联，在运行期动态关联用户、数据内容与操作行为，实现“数据检测与响应”（DDR）的闭环防护。基于智能体的自动化分析系统可快速定位漏洞攻击链并生成防护策略，而隐私计算技术则保障数据流转中的机密性。当前，AI驱动的数据安全体系已成为企业数智化转型的核心支柱，通过降本增效与业务创新，构建新型生产力生态。

核心能力

1、 智能数据发现与分类
2、 动态阻断数据外泄及修复
3、 数据安全态势可视化
4、 隐私及主权合规映射能力
5、 生成式人工智能风险检测

关键挑战

1、 高质量训练预料缺乏
2、 算力配置与延迟成本
3、 模型幻觉与误分类
4、 对抗性攻击风险
5、 数据合规与隐私保护

应用场景

1、 数据智能分类分级
2、 敏感数据流入流出智能监测与脱敏
3、 数据泄露防护
4、 数据资产持续盘点测绘
5、 数据安全态势管理与合规检测

典型厂商

安恒信息 DAS-Security 安全中国
观安
火山引擎
腾讯云安全
领信数科 SECLEAD
SANGFOR 深信服科技
奇安信
360数字安全 数字安全的领导者
数安行 DataSecOps

观安信息-基于大模型的数据分类分级引擎

数说安全

观安

上海观安信息技术股份有限公司是一家提供AI+泛安全产品与服务的工信部专精特新“小巨人”企业。公司聚焦数据安全、人工智能安全、网络空间安全、工业互联网安全、5G安全及公共安全等核心方向。为运营商、政府、能源、金融、智能制造等行业提供全面的网络与数据安全解决方案。公司将人工智能和大数据模型算法应用于网络安全，提高网络威胁的识别率和准确度，保护数字经济下的数据安全和数字安全。公司与多个国际组织、监管机构、高等学府、科研院所建立了良好、紧密的合作关系。

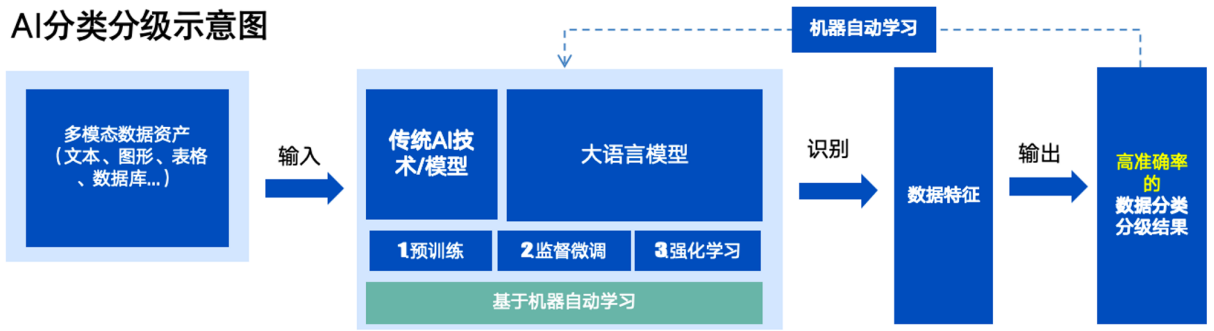
方案概况

方案介绍：

基于大模型的数据分类分级引擎，可以基于字段样例，融合语义关联、模式识别等技术，针对长文本关键字段，中英文夹杂字段，模糊不完整字段自动识别，补全、检出，对数据完成分类分级及原因分析。检出率超过90%+，100%提供字段分类决策依据，整体准确性超过90%，其中在个人敏感信息字段模型准确识别准确率达98.7%。支持多模态图片、文档处理，灵活配置行业规则，轻量化部署。

方案示意图/拓扑图：

AI分类分级示意图



方案优势和用户价值

方案优势与特点：

（1）多模态数据构建体系：首创“人工+模型”双引擎数据生成范式。

（2）轻量化专家模型架构：领域适配微调使训练速度明显提升；轻量化参数压缩显存占用显著降低；适配CPU，GPU，国产NPU等不同计算资源。

（3）对抗鲁棒性增强机制：设计输入层，语义层，训练层三级对抗防御体系。

用户价值：

（1）构建可信决策体系：提高分类分级检出率，准确率，引入大模型根因分析机制，理清责任归属问题。

（2）动态适配监管要求：能够动态适配各地不同的监管要求，适配不同行业分类分级标准。

（3）节约企业运营成本：缩短数据盘点周期，降低人工运维成本，降低GPU资源依赖。

典型客户或目标客户：

电信运营商，其他需要数据盘点、分类分级等安全防控的客户。

ADR

数说安全

ADR（Application Detection and Response，应用检测与响应）基于RASP技术，是区别于HIDS、EDR等传统安全设备的新型安全防护理念和产品形态。随着企业数字化转型和云原生架构的普及，传统以主机、网络为核心的检测与响应手段已难以应对复杂多变的应用层威胁。ADR聚焦于应用层安全，通过对应用流量、行为、数据及接口关键执行方法的监听，实现对应用层攻击的精准检测，并通过自动化响应与溯源分析，实现应用层安全运营闭环。ADR不依赖规则，通过感知上下文，深入理解应用逻辑，可以防护应用层0Day漏洞、内存马注入、反序列化等新型攻击，还可以全量测绘应用资产、发现影子API及陈旧API、治理组件库漏洞，通过联动WAF、SOC、SIEM等安全设备，提升整体安全防护能力。随着AI、自动化等技术的融合，ADR正成为企业应对应用层威胁、保障业务连续性的重要防线和兜底方案。

应用层0Day漏洞自免疫

通过实时行为监测，在未知漏洞利用阶段实现攻击拦截，有效应对0Day漏洞威胁。

内存马注入攻击防护及应急查杀

采用无文件攻击检测引擎，结合运行时内存扫描技术，精准识别并清除内存马等无实体文件攻击载荷。

组件资产全量测绘及漏洞治理

持续识别应用依赖组件并关联漏洞库，建立覆盖开发、测试、运行全周期的漏洞管理闭环。

API资产全量排查及风险预警

动态测绘API接口，实时发现僵尸接口、未授权访问等风险，构建API全生命周期监控体系。

核心能力

1、 应用层0Day漏洞自免疫
2、 内存马注入攻击防护及应急查杀
3、 组件资产全量测绘及漏洞治理
4、 API资产全量排查及风险预警
5、 应用弱密码检测

关键挑战

1、 对多样化应用的兼用适配能力
2、 防护效果和性能占用的平衡机制
3、 业务无感知的批量注入实现
4、 轻量化设计对业务影响的控制
5、 研判精准性与误报率控制能力

应用场景

1、 攻防演练高危未知威胁防护
2、 供应链安全风险治理与闭环
3、 云上应用防护
4、 API风险防护能力提升
5、 老旧业务风险治理

典型厂商

boundaryx 边界无限　奇安信　青藤云安全　亚信安全　昂锐

边界无限靖云甲ADR：防护应用0Day、内存马填补行业空白

数说安全

boundaryx

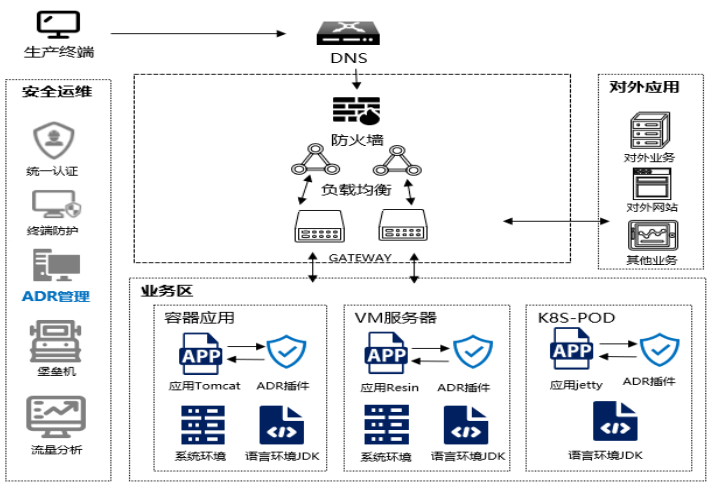
边界无限：守护无界，安全无限！作为国内RASP（应用运行时自防护）的领航者以及ADR概念的创领者，边界无限将国内Top10级别的攻防能力与技术创新基因全部倾注至核心产品**靖云甲ADR**，该产品基于RASP技术，突破传统安全设备的被动防御模式，可实时检测和阻断0Day漏洞利用、内存马注入、反序列化等高危威胁，并可以与WAF等传统应用安全设备协同联动，形成纵深防护的整体应用安全防护体系。 边界无限靖云甲ADR以业界首屈一指的防护效果、轻量化无感部署、兼容性、稳定性、研判精准性，帮助客户实现“入侵可视、攻击可防、漏洞可控”的安全目标，已成为金融、运营商、电网等标杆示范性客户在RASP/ADR领域的首选方案。

方案概况

方案介绍：

边界无限靖云甲ADR采用轻量级探针深度嵌入应用的运行环境，通过Hook关键函数调用和上下文行为分析，无需依赖规则即可实时阻断内存马注入、RCE攻击乃至0Day漏洞利用等新型高危未知威胁。靖云甲ADR融合了资产风险发现能力，可自动梳理应用组件依赖、API接口等资产，识别已知漏洞、API风险、应用弱密码等，达到“动态防御、内外兼顾”的一体化防护，真正实现战时可防+日常可用。

方案拓扑图：



方案优势和用户价值

方案优势与特点：

（1）“全”，应用运行时全生命周期防护覆盖，支持40余种漏洞攻击手法。
（2）“准”，基于应用执行上下文的智能威胁判定引擎，误报率低。
（3）“快”，无感批量部署，分钟级部署生效，秒级检测与响应。
（4）“深”，可穿透应用容器及中间件，实现内存马扫描与清除。
（5）“稳”，经10万+商用节点验证，兼容性高，运行平稳，性能资源占用低。

用户价值：

1. 面对令客户谈虎色变的0Day漏洞、内存马等新型高危未知威胁，高效、精准防护和查杀，保护客户核心应用，轻松应对攻防演练及可能发生的实网攻击。

2.面对日益严峻的外采供应链、老旧业务、云上应用及互联网暴露面风险，为客户应用提供兜底方案，实现供应链安全风险治理闭环，并排查“两高一弱” 等风险。

3.AI智能研判能力，帮助客户提高安全人效，显著降低运营成本。与多种安全设备联动，既提供防护亮点，又避免重复建设，提高精度与广度，达到1+1>2的效果。

典型客户：

交通银行、工银科技、渤海银行、中国农业发展银行、太平洋保险、中央国债、银联数据、浙江移动、越盛能源、华住集团

奇安信- ADR

数说安全

奇安信科技集团股份有限公司（以下简称奇安信，股票代码688561）成立于2014年，专注于网络空间安全市场，向政府、企业用户提供新一代企业级网络安全产品和服务，在人员规模、收入规模和产品覆盖度上均位居行业第一。2019年，中国电子战略入股奇安信，奇安信成为网络安全“国家队”，迄今已完成91次国家重要活动网络安全保障任务，参与近千场实战攻防演习。

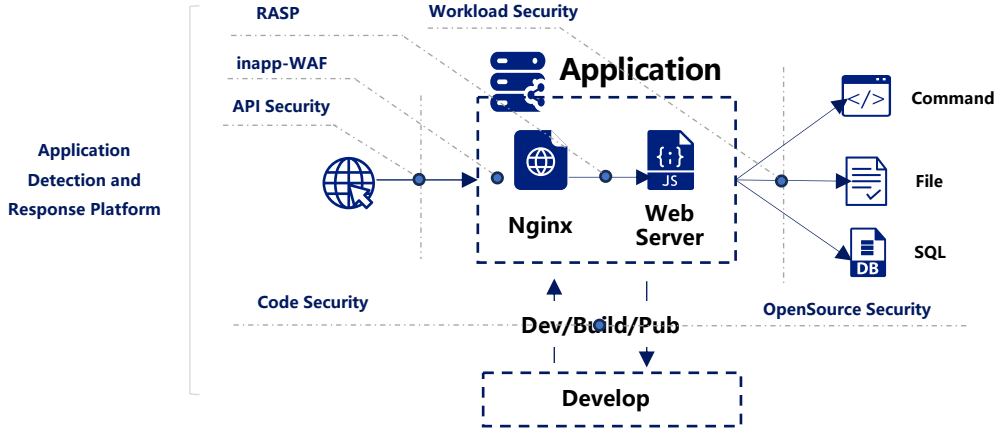
作为网络安全领军企业，奇安信将针对新技术下产生的新业态、新业务和新场景，继续为用户提供全面、有效的网络安全解决方案，向成为“全球第一的网络安全公司”的愿景目标不断奋进，助力实现“十五五”良好开局。

方案概况

方案介绍：

奇安信ADR是一款面向云原生环境的应用程序检测与响应解决方案，方案融合奇安信在云原生安全、源代码检测、API安全领域的优秀能力和最佳实践。从对应用进行全面的资产解构开始，深度识别应用存在的各种风险，多维实时监控应用上下文及内外部环境的恶意行为，实现对应用程序全生命周期的安全管理。方案具有安全能力丰富、DevSecOps集成简便、云原生适应性强等特点，帮助客户提升应用程序的安全。

方案示意图：



方案优势和用户价值

方案优势与特点：

1）**全面的应用资产梳理**。通过组件识别、流量分析等方式，完整解构应用框架、组件、API等应用资产。

2）**应用供应链风险检测是特色**。除应用资产存在的弱口令、漏洞、配置风险等问题发现，还提供对应用组件的SBOM源污染、开源漏洞、开源许可风险的检测。

3）**多维度的应用运行时威胁监测。**方案从应用内调用、应用外运行交互、应用流量侧交互多维度进行入侵检测。

4）**集成式响应能力**。提供应用内函数级拦截、应用外命令执行、文件操作、数据库操作等行为拦截、应用API级流量阻断、开发构建流程拦截等多种响应能力。集成Devsecops，左移至开发上线阶段实现检测响应。

客户价值：

1）以应用为中心，良好融入客户开发态和运行态，强化左移安全和运行时安全并重的应用全生命周期安全检测与响应。

2）适应快速发展的云原生进程，无需重复投入实现从ADR向CADR的平稳演进。

供应链安全

数说安全

供应链安全是指对软件、硬件从开发到交付全生命周期的安全管理与防护，旨在防止恶意篡改、漏洞植入和知识产权侵害。随着数字化进程加速，供应链攻击事件激增，供应链攻击已从“代码注入”升级为“全链路渗透”，统计显示，2020-2023 年开源仓库中的恶意组件大规模出现，2024 年又出现针对知名开源项目的编译链劫持和模型权重投毒。硬件侧同样告急，首个专门针对 Linux 的 UEFI 启动套件于 2024 年曝光，固件级植入进入常态化威胁。2025年，供应链安全聚焦于物联网设备硬件安全、开源组件治理、源码保护等核心领域，供应链安全产品正从单点扫描演进为“生成-签名-分发-运行”四阶段一体化防护：上游生成可验证软件物料清单，下游通过硬件根信任实时度量与封装签名，辅以威胁情报与自动修复闭环，形成覆盖软件、硬件与服务供应链全景的数字护城河。

研发与设计：
源码安全、开源组件治理

运营维护：
动态监控与响应，
漏斗修复与
合规审查

硬件供应链溯源：
硬件物料清单收集、物联网设备安全

供应商协作管理：
多级供应商审计，
威胁情报共享

软件交付与集成：
CI/CD流程嵌入安全检测，云基础设施安全

核心能力

1、 动态HBOM、SBOM生成与差异监测
2、 自动化安全检测工具集成
3、 远程度量与固件健康证明
4、 开源组件安全管控
5、 源码安全与知识产权保护

关键挑战

1、 HBOM、SBOM 格式碎片化
2、 闭源固件与第三方IP黑盒
3、 多级供应商管理缺位
4、 开源组件漏洞泛滥
5、 物联网设备供应链复杂性

应用场景

1、 基础设施固件完整性监控
2、 开源软件治理
3、 软件开发与交付
4、 云服务供应链安全管理
5、 关键信息基础设施全链路防护

典型厂商

卡奇科技 CHAITIN
安全玻璃盒 孝道科技
昂锐 ANINMON
海云安 secidea
BlingSec 比瓴科技

安全玻璃盒-供应链安全AI检测智能体与威胁情报

数说安全

安全玻璃盒【杭州孝道科技】是一家专注于为用户提供软件供应链安全产品和解决方案的国家高新技术企业、专精特新企业。基于AI大模型和卷积神经网络，自主研发全链路智能动态污点分析、二进制函数级智能基因检测、路径可达性可触发性分析、自动化验证与溯源、智能检测一键式修复及运行时免疫防御等核心技术与产品，为用户提供DevSecOps安全解决方案及软件供应链安全一体化解决方案、上线即安全与免疫防御解决方案、供应链安全威胁情报与态势感知及软件供应链安全检查评估工具等。

方案概况

方案介绍：

供应链安全AI检测智能体与威胁情报基于多源资产探测与关键节点建模技术，动态构建覆盖供应链全生命周期的资产图谱与安全图谱，首创供应链安全检测智能体，可以精准识别技术漏洞、供应中断、知识产权侵权等多维度风险。平台结合实时威胁情报数据，构建供应链威胁情报联动机制，基于供应链图谱进行风险溯源，能够快速定位风险影响路径和影响范围。平台针对性解决软件漏洞、组件投毒、供应链断供等供应链安全威胁，实现风险评估、威胁告警与处置的全链路安全治理。

方案示意图/拓扑图：

方案优势和用户价值

方案优势与特点：

（1）供应链数据采集与标准化建模：运用智能采集助手整合多维度数据，形成高质量供应链数据资源池。

（2）供应链资产图谱构建与风险传播分析：动态可视化的供应链资产图谱，实现风险的可视化分析与预测。

（3）多模态AI驱动的供应链安全智能检测：实现从代码层到交付层的全栈风险感知与闭环管理，推动供应链安全从单点工具堆砌向智能化协同防御转型。

（4）供应链安全态势感知与威胁情报协同：整合多维度威胁情报源，构建实时动态的供应链安全态势感知体系，实现风险的全局监控与精准响应。

用户价值：

减少安全事件损失，优化安全管理成本：通过自动化的实时监测溯源分析，减少人工检测和管理成本，快速响应和处理安全事件，降低直接经济损失。

加速软件开发流程，提升供应链协同效率：在开发过程中可以及时发现安全风险，通过与供应商共享安全信息，协同解决安全问题。

目标客户：

政企事业单位、互联网、金融等行业；重视安全、关注新兴技术的企业；监管单位。

长亭科技-“慧鉴”智能静态应用程序安全测试系统

数说安全

长亭科技 CHAITIN

长亭科技成立于2014年7月，是国际顶尖的智能网络安全公司之一。长亭科技创始团队来自国际知名CTF战队清华“蓝莲花”，在近年来的全国网络安全竞赛以及国家和各地区各行业的实网攻防演练中，荣获冠军近百次，被评为“最了解攻击的防守队伍”。面向数字时代的新型安全风险，长亭科技秉持“知攻善防，智能安全”的理念，从攻击视角构建防御体系，以AI驱动重塑安全运营，通过新一代智能安全产品防护体系和专业的安全测试及咨询服务，帮助金融、通信、政府、能源、互联网等重要行业的4000多家用户，打造了先进可靠的实战化、体系化、常态化的智能安全运营能力。

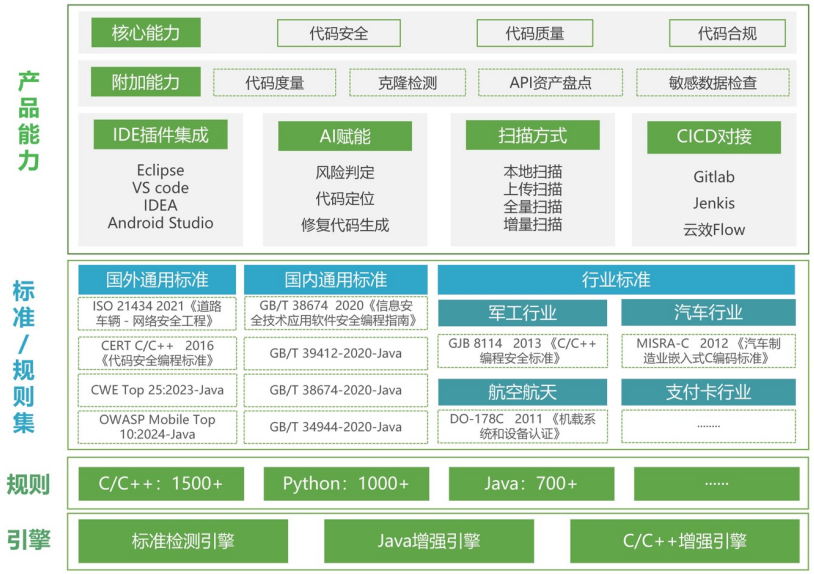
方案概况

方案介绍：

当前AI辅助开发工具显著提升生产效率，产品迭代进一步加快，安全审查和渗透测试工作量激增。长亭科技“慧鉴”是自主研发的新一代静态应用程序安全测试系统，专为AI编程时代打造，致力于解决代码安全、质量和合规三大核心问题。

产品采用数据流分析、控制流分析、污点分析、约束求解、符号执行、AI辅助分析等前沿静态检测技术，显著提升检测准确度和覆盖范围。系统支持20多种主流编程语言，内置超过3000条检测规则，覆盖OWASP、CWE、SANS等国际标准以及国内外合规要求，支持对接常见的开发IDE、CICD流水线，为现代软件开发提供全方位的安全保障。

方案示意图/拓扑图：



方案优势和用户价值

方案优势与特点：

- **全面**：一站式全维度代码智能分析，单次深度扫描即可实现代码安全、质量提升、合规达标、度量分析、克隆优化、资产盘点和敏感数据保护七大目标
- **精确**：领先的C/C++与Java分析能力；完善的嵌入式编译环境支持；完善的Java开发框架支持；低至10%的实战误报率
- **易用**：模块化设计，快速接入任意CI/CD环境；详细的结果展示，帮助用户快速定位成因；丰富的API接口，实现DevSecOps流转
- **智能**：利用AI实现代码语义的理解，提升逻辑相关问题的发现能力；大模型赋能，基于程序分析过程构建RAG方案，实现智能误报分析、智能二次降噪及生成代码修复方案

用户价值：

- **自身安全需求**：通过深度集成到CI/CD开发流程中，实现代码安全前置，显著提升软件产品整体安全性和代码质量水平，从源头保障软件安全。
- **测试阶段安全验证**：为安全团队提供专业的代码审计工具和完整的安全评估体系，确保软件产品在上线前通过严格的安全验证，构建企业安全治理能力。
- **监管合规审查**：为监管机构提供标准化、权威性的代码安全检测工具，支撑行业监管和第三方认证工作，提升整个行业的安全水平。

AI应用防火墙

伴随生成式AI在金融、政务、客服等场景迅速普及，传统WAF只能看流量而无法理解提示词-输出链的语义风险：提示注入、越权调用、机密泄漏与有害内容生成频繁出现。
AI应用防火墙（AIFW）应运而生，专为AI系统量身打造，通过对AI模型、数据流、接口调用等环节进行实时监控，智能识别并阻断恶意攻击，有效防范模型窃取、越权访问、敏感信息泄露等威胁。
与传统安全产品相比，AIFW能够深入理解AI交互的上下文和复杂语义，识别潜藏在提示词设计、模型调用链路中的隐蔽风险。同时具备动态学习与自适应能力，根据新型攻击手法及时调整防护策略。未来，AI应用防火墙将成为AI系统安全的"第一道防线"，助力企业在智能化转型中构建稳固安全屏障。

核心能力

1、Prompt 语义检测与越权拦截
2、数据流与接口安全监控
3、多语言 / 多模态上下文关联分析
4、访问控制与身份鉴别
5、自适应安全策略更新

关键挑战

1、攻击手法快速演变
2、模型与数据复杂性高
3、误报与漏报风险
4、高精度语义分析耗时耗算
5、对抗 Prompt 与变形语句，可能绕过语义特征匹配

应用场景

1、SaaS 平台与插件市场的 AI API 入口防护
2、金融、医疗等高风险行业生成内容合规过滤
3、开发者平台 Prompt 安全测试与沙箱
4、智能客服与对话机器人安全防护
5、企业级AI平台的接口与数据安全防护

典型厂商

火山引擎
安泉数智 新一代数智安全引领者
奇安信
亚信安全
SANGFOR 深信服科技
启明星辰 www.venustech.com.cn

奇安信-大模型卫士（GPT-Guard）

奇安信科技集团股份有限公司（以下简称奇安信，股票代码688561）成立于2014年，专注于网络空间安全市场，向政府、企业用户提供新一代企业级网络安全产品和服务，在人员规模、收入规模和产品覆盖度上均位居行业第一。2019年，中国电子战略入股奇安信，奇安信成为网络安全“国家队”，迄今已完成91次国家重要活动网络安全保障任务，参与近千场实战攻防演习。

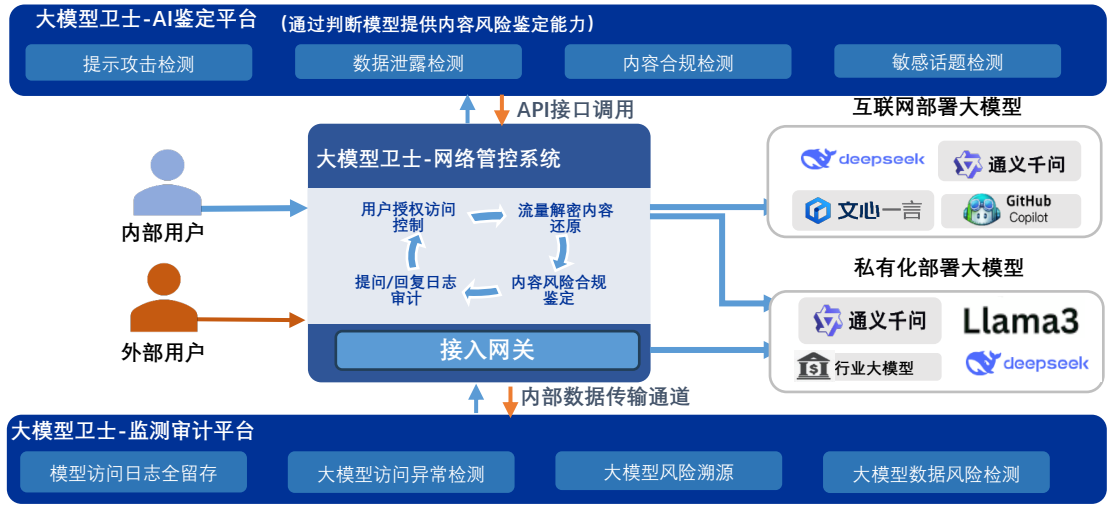
作为网络安全领军企业，奇安信将针对新技术下产生的新业态、新业务和新场景，继续为用户提供全面、有效的网络安全解决方案，向成为“全球第一的网络安全公司”的愿景目标不断奋进，助力实现“十五五”良好开局。

方案概况

方案介绍：

奇安信网神大模型卫士（GPT-Guard）是奇安信推出的业内首款大模型应用安全防护产品，该产品专注于解决企业在大模型使用过程中的模型安全、数据安全、内容合规及网络安全问题。通过提示词攻击防护、算力消耗防护、数据泄露防护等能力，对输入、输出的内容进行安全防护，保护模型业务不受 OWASP LLM Top10 攻击，防止模型中的敏感数据的泄露，确保企业在大模型时的应用安全、数据安全与合规性。

方案示意图/拓扑图：



方案优势和用户价值

方案优势与特点：

（1）轻落地部署，零改造实现大模型安全升级。

（2）实时内容检测及风险事中鉴定、自动拦截，有效防护模型新型攻击。

（3）自研安全对抗与防御引擎，可实时拦截Prompt注入攻击、模型对抗攻击等多种攻击手法。具备对抗训练数据学习泛化能力，可识别新型大模型攻击变体。

（4）集成集团多项数据检测和保护能力，可以有效防护模型数据泄露。

（5）具备软硬件一体机与软件交付两种形态，成熟网关产品灵活适配企业环境。

用户价值：

（1）帮助用户抵御AI攻击风险，提升大模型在企业的使用效率。

（2）帮助用户规避大模型运行服务中的数据泄露风险。

（3）帮助用户解决大模型服务中的合规和隐私保护问题。

（4）为用户提供精细化管控和大模型使用审计与事件追溯手段。

典型客户或目标客户：

政务，医疗，金融行业，大型企业、制造业、外企等。

亚信安全-AI大模型防火墙

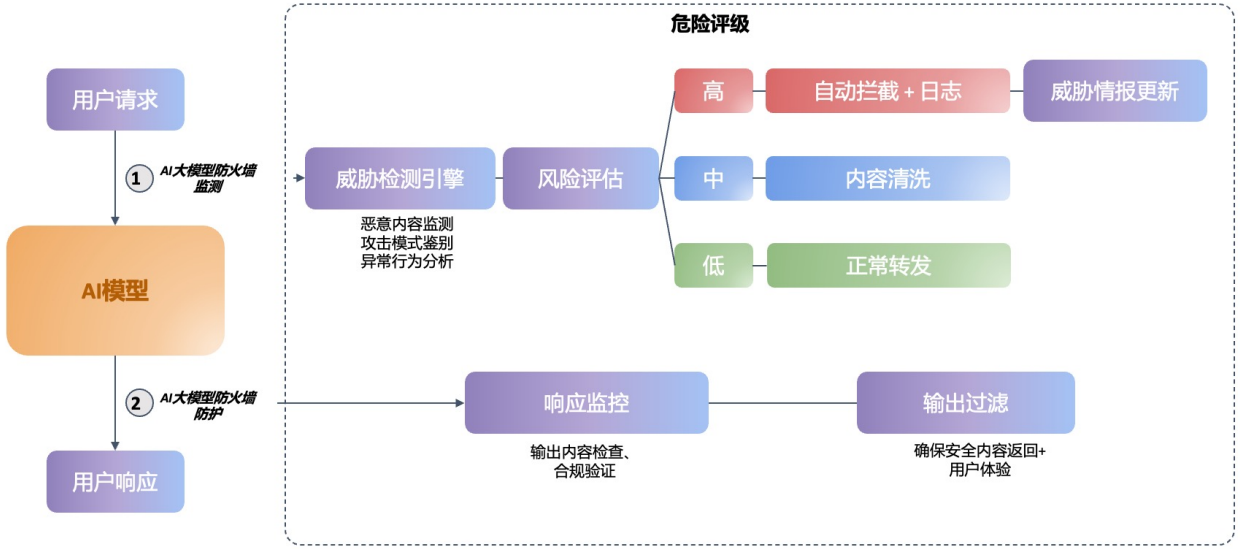
亚信安全科技股份有限公司（简称亚信安全，股票代码688225）是安全+数智一体化领导者。亚信成立于1993年，承继30余年互联网建设经验，以护航产业互联为使命，安全数字世界为愿景，亚信安全2015年正式启航。多年耕耘，已成为国家网络安全建设重要力量。为护航数智化未来，2024年底亚信安全完成对亚信科技（股票代码01675.HK）的并购，以“懂网、懂云、懂安全”为优势基因，打造“云网安”一体的能力体系，成为中国首家近百亿规模的“安全+数智”一体化领导者，赋能产业数智化发展与变革。

方案概况

方案介绍：

产品旨在为大语言模型(LLM)应用提供输入、输出的语义级内容安全防护，包括检测和防护针对LLM的提示词注入攻击，各类大模型攻击防护，防止模型中的敏感数据的泄露等。

方案示意图/拓扑图：



方案优势和用户价值

方案优势与特点：

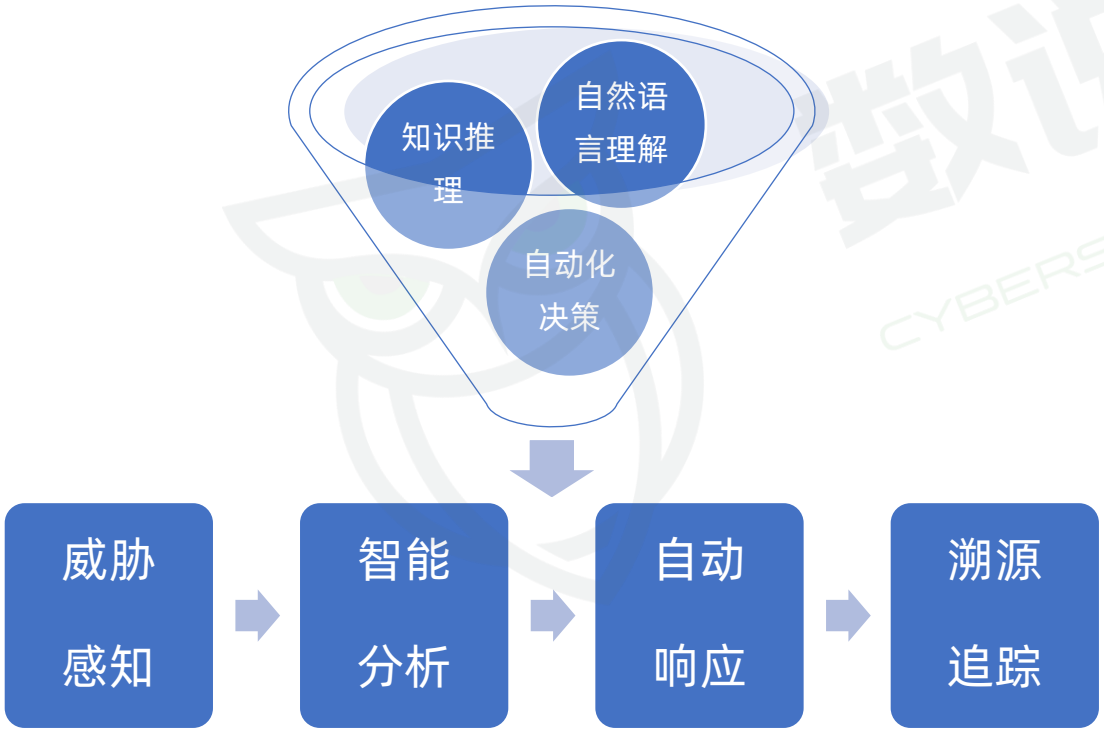
1. 支持多模态检测，包含文本、图片、音频、视频；
2. 基于预训练的安全模型对AI大模型应用进行防护，做到以模制模；
3. 对于后门攻击、数据污染、梯度替换、模型操控、供应链投毒等攻击防护成功率可达98%以上。对事实性评测、模型幻觉、逻辑推理、安全合规、隐私保护等检测准确率可达98%以上；
4. 集成300+前沿模型，支持2000+系统漏洞评估，可检测50+攻击手段，涵盖30+评估标准；
5. 轻量化部署，节省算力资源，消费级显卡即可运行。

用户价值：

1. 增强AI大模型与应用的安全性，保障数据隐私与合规性；
2. 智能化安全运营，提升安全运营效率。亚信安全AI大模型防火墙能够自动学习和适应新的威胁模式，自动更新防护策略，自动生成风险报告，提升了安全运营效率。

安全运营智能体

网络安全智能体是指基于人工智能大模型，专为网络安全领域设计的智能化软件体或服务系统。随着网络威胁日益复杂，传统安全产品和防护体系越来越依赖资深专家进行策略配置、威胁分析和事件响应，导致人力成本高、响应速度慢、创新能力受限。网络安全智能体通过自然语言理解、知识推理、自动化决策等AI能力，将专家经验与大数据智能相结合，实现威胁感知、分析、响应、溯源等环节的高度自动化和智能化。它能够自主学习新型攻击手法，自动适应复杂多变的网络环境，大幅提升安全运营效率和防护能力，成为网络安全产业智能化升级的重要方向。



核心能力

1、多模态威胁情报检索与推理
2、智能告警关联与优先级判定
3、自主安全策略生成与优化
4、可解释决策与合规映射
5、持续学习与基线更新
6、多 Agent 协同分工

关键挑战

1、大模型幻觉或误判
2、权限边界模糊
3、企业环境高度异构
4、安全工具 API 碎片化
5、成本：私域部署算力与数据安全成本高

应用场景

1. 7×24 SOC 自动化监测与响应
2. MSSP 托管安全服务的工单编排
3. 零信任策略实时微分段与动态访问控制
4. 云原生环境入侵检测与容器自愈
5. 数字取证中的事件溯源脚本与时间线重建
6. DevSecOps 流水线中的自动漏洞修复建议

典型厂商

安恒信息 DAS-Security 安全中国
SANGFOR 深信服科技
奇安信
360数字安全 数字安全的领导者
金睛云华 AI驱动安全
火山引擎
启明星辰 www.venustech.com.cn
NSFOCUS
天融信 TOPSEC
未来智安 XDR SEC

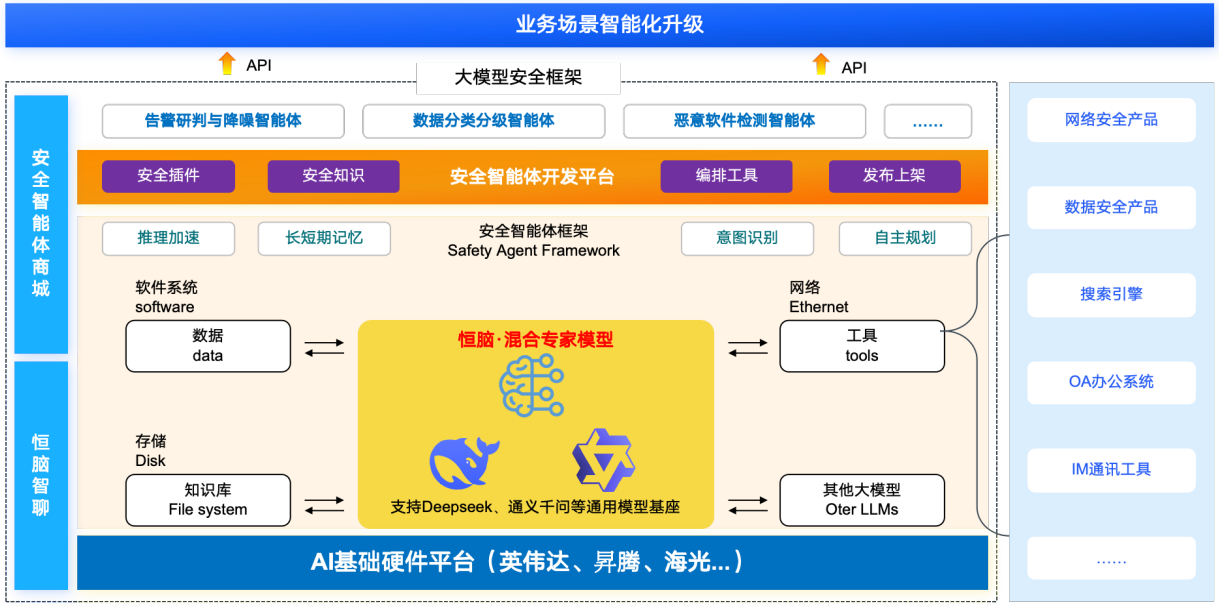
安恒信息-恒脑安全智能体开发平台

安恒信息技术股份有限公司（股票代码：688023）成立于2007年，于2019年科创板上市，是国内网络安全、数据安全和数据要素领军企业之一，科创板创新30强中唯一的数字安全企业，也是国内积极推动以AI赋能传统网络安全、数据安全转型的典型企业代表。安恒信息以AI驱动产品服务革新，于业内首发恒脑安全垂域大模型，并推出国内首个安全AI智能体恒脑3.0，自主研发了安全岛隐私计算平台、数由空间、数由器等明星产品，为数据安全流通、数据可信接入等数据基础设施建设场景中的难题提供了解决思路。）

方案概况

方案介绍：

恒脑安全智能体平台是基于恒脑安全垂域大模型，结合自研安全智能体框架，具备零代码和低代码创建智能体能力，用户可基于业务场景，结合安全运营最佳实践编排智能体，灵活调度安全工具、补充安全知识库，实现从环境感知到决策处置闭环。



方案优势和用户价值

方案优势与特点：

（1）支持多种参数规模（包括7B、32B和72B等），也可基于客户算力资源条件，采用int4、int8等不同程度量化

（2）支持零代码、低代码创建智能体，官方智能体超100个，已覆盖90%日常运营和80%重保场景

（3）意图识别精准，可实现任务的自规划、自执行和自反馈

（4）能力开放，提升现有安全产品“秒变”智能化

（5）支持昇腾910b、海光DCU（K100_AI）等国产算力平台

用户价值：

1、省人力：智能体大幅提升安全运营效率，运营人力成本直降50%

2、提效率：告警研判准确率超99%，处置时间从小时级缩至分钟级

3、降风险：经亚运会、亚冬会等“零事故”验证，重大活动安全保障无忧

典型客户或目标客户：

亚奥理事会、杭州大数据局、金华大数据局、浙江移动、中兴通讯、国能集团、海量集团、太仓市网信办、

深信服科技-AI安全运营方案

数说安全

SANGFOR 深信服科技

深信服科技股份有限公司，成立于2000年，总部位于深圳，是一家企业级网络安全、云计算、AI及物联网领域的创新高科技企业。公司拥有超7000名员工，在全球设有多个研发中心，业务覆盖30个国家和地区，为超过10万家企业级用户提供数字化转型的支持。深信服旗下拥有深信服智安全和信服云两大业务品牌，致力于让用户的数字化过程更加简单和安全。

方案概况

方案介绍：

深信服AI安全运营方案，通过从“安全运营效率高、线上服务省心”，依托生成式AI技术实现自主安全运营。该方案能够自动化完成安全数据的开放接入、前置对抗、全量告警研判与威胁闭环，结合专家服务实现AI智能值守，显著提升安全运营的实战能力。

方案示意图/拓扑图：

方案优势和用户价值

方案优势与用户价值：

（1）智能对抗，防御领先一步：基于生成式AI实现自动化对抗，有效提升威胁防御水平和响应速度；

（2）全量告警研判，少量精准地提取关键风险：从海量告警中自动研判并精准识别关键威胁，输出高可信安全事件，大幅降低误报与释放人力以应对高价值事件；

（3）服务闭环，保障用好AI：通过线上+客户成功服务体系，确保AI能力与实战场景深度融合，保障运营效果持续优化，真正实现省心可靠的安全闭环。

典型客户或目标客户：

该方案已广泛应用于某头部部委、某省级气象局、某能源龙头企业、某大型光能企业、某头部家电制造业、某头部金融证券、某空管单位等数百家大型用户，覆盖央国企、部委、制造业、金融、能源、教育、医疗、交通等重点行业。

通过AI安全运营可日均分析数百万告警，实战场景自动化运营达成率高达96%以上，显著提升用户安全运营效率与体系化防入侵能力。

奇安信-智能安全运营AISOC

奇安信科技集团股份有限公司（以下简称奇安信，股票代码688561）成立于2014年，专注于网络空间安全市场，向政府、企业用户提供新一代企业级网络安全产品和服务，在人员规模、收入规模和产品覆盖度上均位居行业第一。2019年，中国电子战略入股奇安信，奇安信成为网络安全“国家队”，迄今已完成91次国家重要活动网络安全保障任务，参与近千场实战攻防演习。

作为网络安全领军企业，奇安信将针对新技术下产生的新业态、新业务和新场景，继续为用户提供全面、有效的网络安全解决方案，向成为“全球第一的网络安全公司”的愿景目标不断奋进，助力实现“十五五”良好开局。

方案概况

方案介绍：

奇安信 AISOC 是奇安信集团推出的AI 驱动的安全运营中心，由**下一代安全运营中心（NGSOC）**与**奇安信安全大模型（QAX-GPT）**深度融合而成，代表了网络安全运营领域的革命性突破。其核心目标是通过 AI 技术重构安全运营流程，实现威胁检测、调查与响应（TDIR）的全自动化和效率跃升。

AI赋能运营全流程：

方案优势和用户价值

方案优势与特点：

（1）“**智筛**”，整合190+设备告警，98%无效告警秒级过滤，锁定 1% 高价值威胁

（2）“**快判**”，智能体 7×24 值守，单条告警处理≤9 秒，研判效率超人类 60 倍

（3）“**秒溯**”，攻击路径秒级溯源，100% 可视化重建

（4）“**闭环**”，联动协同190+设备，编排式自动处置响应

（5）“**进化**”，智能体自主感知识别，反哺迭代，从 “被动响应” 转向 “主动狩猎”

用户价值：

（1）效率革命：人力成本直降70%。传统SOC需3人日处理的攻击事件，AISOC单分析师7分钟闭环，年节省专家工时超2千小时，响应从 “小时级” 迈入 “秒级”。

（2）精准防御：关键威胁0漏报。通过攻击链完整还原，攻击行为无遗漏，避免 “过度告警” 与 “漏报风险” 的双重困境。

（3）能力升维：从 “应急” 到 “预测”。依赖于AISOC 构建的威胁预测模型，分析师从 80% 重复性工作中解放，专注高阶威胁狩猎，安全团队效能向 “战略防御” 转型。

典型客户：

JAGUAR LAND ROVER　CEC 中国电子 CHINA ELECTRONICS　生态环境部　广西壮族自治区 yjglt.gxzf.gov.cn　应急管理厅　+更多

360-360安全大模型夯实智能体发展基石

数说安全

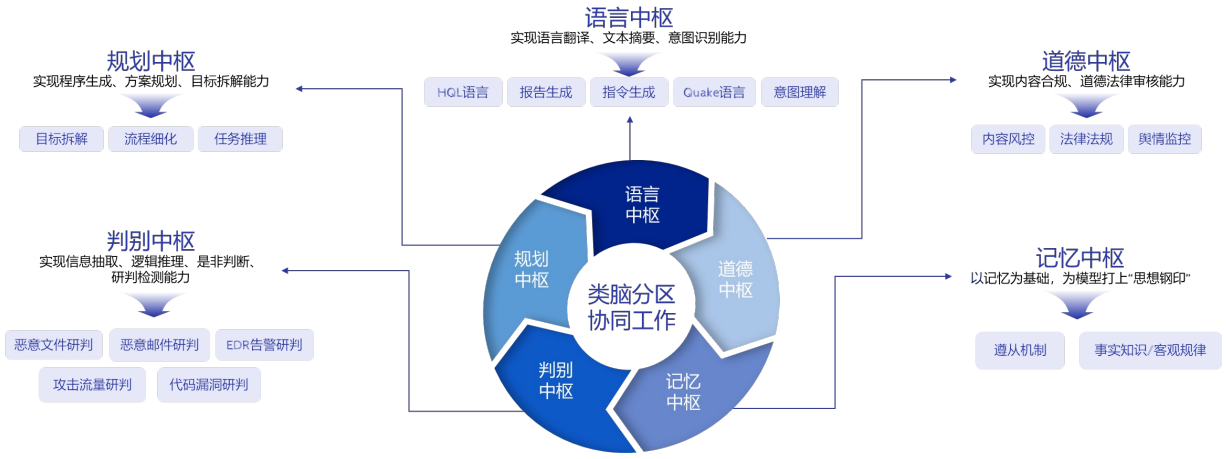
360数字安全科技集团有限公司（360数字安全集团，360 Digital Security Group）是数字安全的领导者，秉持“上山下海助小微”企业使命，确立安全和AI发展双主线。在安全领域，为解决国家“看见”高级威胁的卡脖子问题探索形成一套以“看见”为核心的数字安全中国方案，基于安全即服务理念，将这套方案通过360安全云赋能城市、大型企业、中小微企业。在人工智能领域，推出360安全大模型并率先实现AI实战应用，基于“以模制模”理念，推出数字专家智能体，打造大模型安全卫士解决传统安全和大模型安全问题。

方案概况

方案介绍：

依托360海量安全数据，以CoE（多专家协同）安全大模型架构为基础，通过安全场景专项化训练，形成安全垂类大模型，完成终端行为归因、网络告警研判、钓鱼邮件深度解析等专项安全任务，夯实安全智能体发展基石。

方案示意图/拓扑图：



方案优势和用户价值

方案优势与特点：

（1）参数共享，模型紧凑：360安全大模型基于一个模型基座专项训练专家分区，每次安全任务无需激活所有参数，根据任务类型和能力灵活组合调用。

（2）多“专家”独立，无干扰：相比于多种能力训练在同一个模型中，360安全大模型采用多“专家”增量训练的方式，有效避免了多任务冲突问题。

（3）“专家”紧密协同，好扩展：安全任务上多专家协同，多个专家之间可以共享kv-cache, 协同过程更紧密且更高效。

用户价值：

（1）终端行为研判归因能力：展现出通用大模型不具备的终端行为研判归因能力；

（2）网络日志检测能力：百亿参数专项训练达到或超越DeepSeek满血版的效果。

安全威胁检测智能体

数说安全

安全威胁检测智能体融合大语言模型、多模态感知与自主决策，持续摄取并语义化融合网络、端点、身份、云与OT等多源遥测数据，通过时序行为建模、图谱关联分析与因果推理自动生成可解释的攻击故事。

安全威胁检测智能体实现从传统"规则匹配+静态模型"向"语义理解+动态推理+自主验证"转型。技术上向实时流式+边缘协同、图神经网络+概率风险量化、假设生成+主动验证自动化等方向演进；商业上从技术指标转向业务风险量化，推动"检测即服务"模式创新。最终实现从被动告警向主动威胁猎捕跃迁，解决多源数据孤岛、告警冗余、复杂威胁响应滞后等核心痛点。

多源数据语义融合
统一摄取日志、网络、端点等异构数据并语义标准化

智能威胁推理
运用因果推理识别真实攻击路径并量化风险

攻击链路自动重构
基于时序分析与图谱关联自动生成完整攻击故事

主动假设验证
自主生成威胁假设并通过沙箱等手段验证可行性

自适应响应编排
自动生成处置建议并编排跨系统响应流程

核心能力

1. 多域数据语义融合
2. 攻击故事自动生成
3. 因果推理与风险量化
4. 主动验证与假设检验
5. 自适应学习与进化与智能响应编排

关键挑战

1. 数据质量与标准化难题
2. 模型可解释性与信任度
3. 实时性能与资源消耗
4. 对抗攻击与模型鲁棒性
5. 隐私保护与数据合规

应用场景

1. APT高级持续威胁检测
2. 内部威胁与身份滥用监控
3. 云原生与混合环境安全
4. 工业控制系统（OT）安全
5. 金融交易欺诈检测

典型厂商

SANGFOR 深信服科技
金睛云华 AI 驱动安全
微步在线
奇安信
安恒信息 DAS-security 安全中国
启明星辰 www.venustech.com.cn

深信服-全量威胁检测智能体

数说安全

SANGFOR 深信服科技

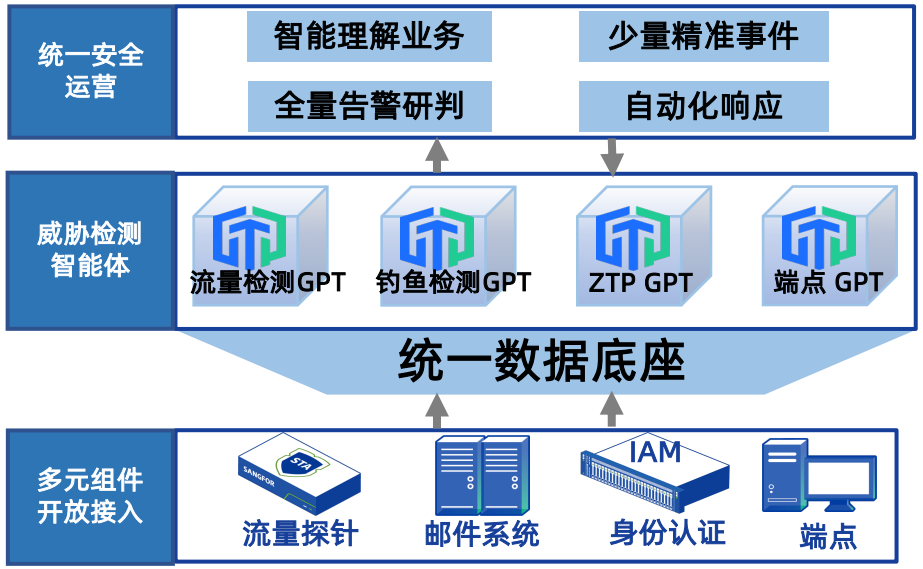
深信服科技股份有限公司，成立于2000年，总部位于深圳，是一家企业级网络安全、云计算、AI及物联网领域的创新高科技企业。公司拥有超7000名员工，在全球设有多个研发中心，业务覆盖30个国家和地区，为超过10万家企业级用户提供数字化转型的支持。深信服旗下拥有深信服智安全和信服云两大业务品牌，致力于让用户的数字化过程更加简单和安全

方案概况

方案介绍：

全量威胁检测智能体是融合大模型技术的新一代检测系统，从流量、钓鱼、身份、端点四个维度实现AI对抗威胁，具备"感知-理解-推理-研判"全链路威胁检测对抗能力，精准应对0day、加密威胁、无文件攻击、高级钓鱼、身份滥用等传统方案难以应对的核心威胁。

方案示意图/拓扑图：



方案优势和用户价值

方案优势与特点：

（1）流量威胁检测：精准识别0day攻击、加密通道威胁、业务逻辑漏洞等传统NDR无法检测的攻击行为；

（2）钓鱼威胁检测： 有效检测加密附件、二维码跳转、社工欺诈等高隐蔽性钓鱼手法，提升邮件安全防护等级；

（3）安全风险治理：以身份、资产、权限和策略为中心，针对潜在安全风险进行持续的感知、评估与动态控制，解决身份滥用、异常访问行为等风险；

（4）端点威胁检测：解决无文件攻击/凭证窃取/横向移动/银狐等恶意程序威胁。

用户价值：

（1）第一时间发现高级威胁：无论是0day漏洞利用、加密Webshell上传还是高仿钓鱼邮件，均可在第一时间精准告警，不再错过关键攻击信号；

（2）智能化自动响应处置：90%以上的攻击可实现一键自动遏制，配合攻击解读与处置建议，极大提升事件响应效率，完成安全事件闭环，降低人工处置负荷。

典型客户或目标客户：

某金融客户在实战攻防中，依托该方案部署成功检出多起传统引擎漏判的真实攻击，包括0day利用、钓鱼攻击和加密webshell，对高混淆对抗攻击检出率超过96%。

目标客群：适用于政府，金融、能源、交通、运营商、医疗，教育及大型企业，助力构建新一代智能安全防御体系。

金睛云华-大模型赋能的自动化威胁检测&安全运营解决方案

数说安全

金睛云华
AI 驱动安全

北京金睛云华科技有限公司成立于2016年，是国家级专精特新“小巨人”企业，国家信息安全漏洞库(CNNVD)一级技术支撑单位。公司以“AI驱动安全”为核心理念，致力于成为以AI技术为核心的新一代安全产品及解决方案提供商。

金睛云华核心团队主要来自清华大学KEG实验室、华为、启明星辰、东软等专业机构，在人工智能和网络安全领域拥有超过二十年的专业经验和技术积累。至今已获得了近100项人工智能和网络安全相关的发明专利；

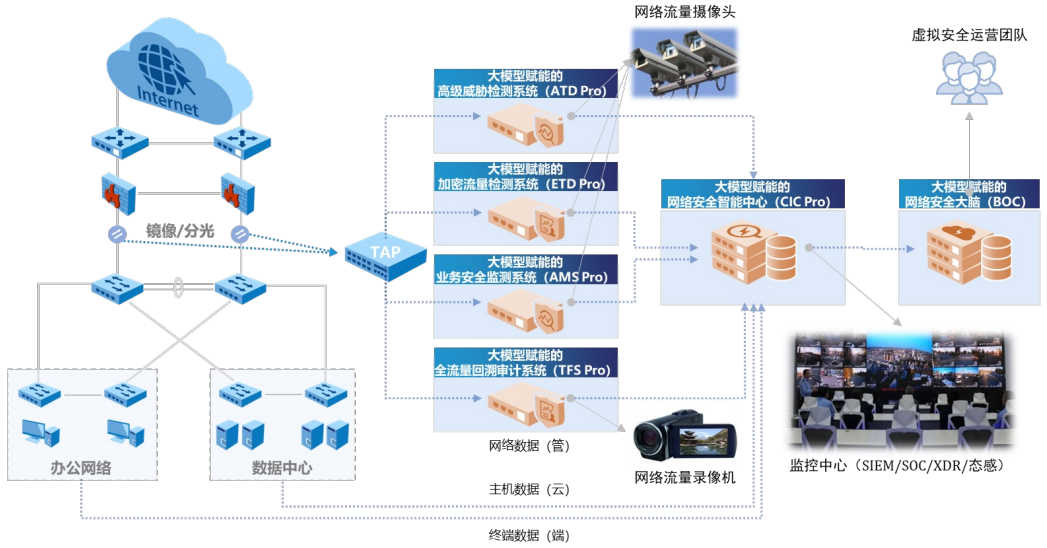
金睛云华AI赋能的NDR/EDR/XDR系列产品与解决方案至今已服务的客户超过1200家，现网部署的系统超过6500套。

方案概况

方案介绍：

安全检测和事件处理一直是网络安全领域的重要课题，难度大，对人才的能力和数量要求高。基于大模型的检测技术，对加密流量、各种变形，免杀都有非常好的检测能力，基于大语言模型的运营能力，对安全事件的解读，建议，预处理等都能做到智能化，有效降低对人员的能力和数量依赖。

方案示意图/拓扑图：



方案优势和用户价值

方案优势与特点：

（1）支持基于大模型&智能体的未知威胁&APT检测，支持恶意代码变种检测；
（2）支持基于大模型&智能体的加密流量检测（无需解密）；
（3）支持基于大模型&智能体的API业务安全监测及API资产智能识别；
（4）支持基于大模型&智能体的全流量回溯分析取证，PB级数据秒内检索；
（5）支持基于大模型&智能体的自动化安全运营；
（6）算力要求低（2块RTX 4090），处理性能高（≥1000eps），极具性价比。

用户价值：

（1）海量告警日志降噪&降误报率最高可达99.9%以上；
（2）有效事件检出率可提高20%以上；
（3）运营成本可降低90%以上，运营效率可提高数十倍以上。

目标客户：

对威胁检测及安全运营要求高，期望通过大模型实现降本增效的关基行业客户。目前已在电力、能源、军队、运营商、金融、政府、安全监管等行业成功应用。

微步在线-XGPT Agent 威胁分析与安全运营智能体

数说安全

微步在线®

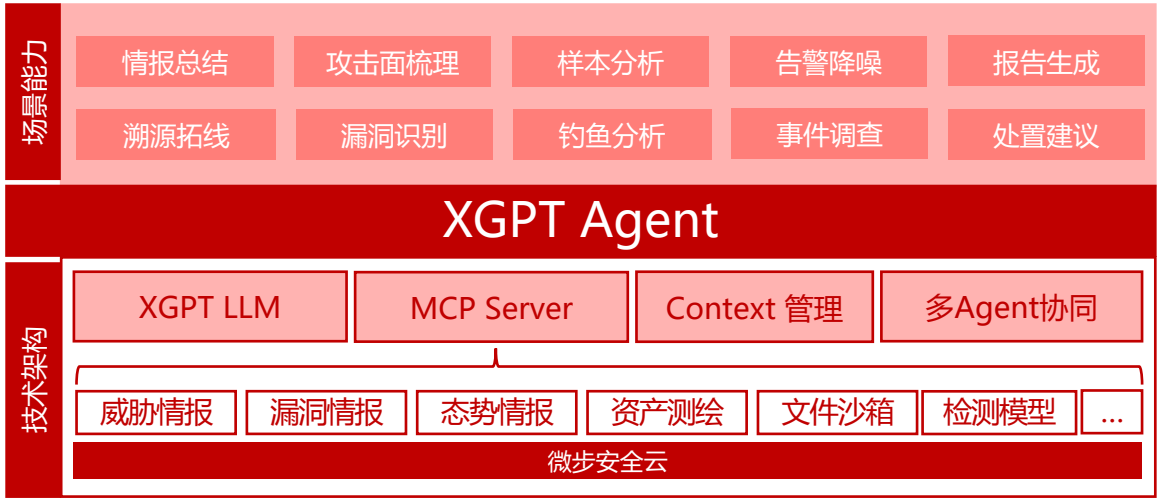
微步成立于2015年，是网络安全技术创新型企业，专注于精准、高效、智能的网络威胁发现和响应，开创并引领中国威胁情报行业的发展，以威胁情报（TI）和人工智能（AI）为技术内核，提供TI+AI驱动的“云+流量+边界+端点”新一代智慧安全运营产品及服务，帮助客户建立全生命周期的威胁监控体系和安全响应能力。微步旗下安全大模型XGPT是国内首个通过中央网信办双备案的安全大模型，打造专属于安全从业者的智能AI助手，深度融合了微步业界领先的威胁情报，加速分析研判和处置，提升安全运营效率。

方案概况

方案介绍：

微步在线XGPT Agent是依托威胁情报（TI）与人工智能（AI）打造的威胁分析与安全运营智能体。XGPT Agent深度融合微步安全云中的情报数据和分析引擎，并通过多Agent协同架构，训练吸收安全运营最佳实践，能自主化地完成复杂的威胁分析、事件调查、响应处置等工作，全面驱动安全运营智能化与自动化。

方案功能与架构：



方案优势和用户价值

方案优势与特点：

- **深度融合威胁情报，分析结果精准可靠：**XGPT Agent 深度集成了业界领先的微步在线威胁情报和威胁分析工具。确保每一次分析结论都准确、丰富、且有理有据。
- **实战化智能体集群，场景覆盖全面：**XGPT Agent吸取了行业内的最佳实践，打造了一个强大的实战化智能体集群。无论是情报分析、告警研判、漏洞运营，还是攻击面梳理、钓鱼分析，XGPT Agent 都能提供专业的支持，有效应对各类任务。
- **智能协同，高度自动化：**XGPT Agent具备高度自动化能力。通过多智能体协同机制，能自主对齐目标，制定并执行计划，从而独立完成多步骤的复杂安全任务，极大地提升了安全人员的工作效率，使其能够专注于更具战略性的工作。

用户价值：

（1）分析精准度：分析结果准确率99.99%，覆盖99%关联威胁实体，单次威胁分析平均耗时分钟级；

（2）场景自动化：85%标准化工作无需人工介入，减少70%重复性操作；

（3）协同响应效能：多步骤调查任务提速80%。

深度伪造检测

数说安全

随着生成式人工智能的快速发展，深度伪造（DeepFake）技术在文字、图片、音频和视频等多模态内容生成领域取得了突破性进展。与此同时，深度伪造内容在网络谣言、虚假信息传播、身份冒用、网络诈骗等场景中带来了严峻的安全挑战。深伪鉴别技术应运而生，旨在通过算法和工具对由AI生成的伪造内容进行精准识别与溯源。该技术综合利用多模态数据分析、机器学习、特征提取和行为建模等手段，能够自动检测和判别伪造内容的真实性，有效遏制深度伪造带来的风险。随着应用需求的不断提升，深伪鉴别技术正逐步向高精度、实时化和自动化方向演进，成为网络安全防护体系中不可或缺的重要组成部分。

数据输入与预处理

- 收集待检测的图像、音频或视频数据
- 进行格式转换、剪裁、降噪等预处理操作
- 选择性地进行人脸、语音等目标区域的提取

特征提取与多模态分析

- 提取空间域、频域、时序等多维特征
- 利用卷积神经网络（CNN）、时序神经网络（RNN）等模型进行特征编码
- 多模态融合：结合图像、音频、文本等多源信息提升检测准确率

伪造检测与结果输出

- 使用分类模型（如深度学习、集成学习等）判断内容真伪
- 输出检测结果（真/伪、置信度分数等）

核心能力

1. 多模态微特征提取与融合
2. 自监督生成检测模型
3. 可信溯源水印
4. 可解释性与可追溯性分析能力
5. 基于大数据的伪造样本训练与判别能力

关键挑战

1. 生成模型迭代迅速
2. 伪造内容与真实内容的高度相似性
3. 多模态内容的复杂性与异构性
4. 高质量训练数据稀缺且标注成本高
5. 实时场景算力与带宽开销大

应用场景

1. 媒体内容真实性审核
2. 金融反欺诈与身份认证
3. 政府及司法证据鉴定
4. 社交平台内容安全监控
5. 企业舆情与品牌保护
6. 重要会议与远程通信安全

典型厂商

中科睿鉴 让世界更可信

RealAI

中科睿鉴-端云协同多模态伪造检测方案

数说安全

中科睿鉴
让 世 界 更 可 信

中科睿鉴以“用技术让世界更可信”为使命，致力于用AI技术守护AI安全，围绕数字内容伪造检测、多模态内容安全、大模型安全等核心技术，打造集数据、算力、模型为一体的AI安全能力底座，面向国家监管、行业应用和个人服务等场景，推出独家端云协同治理方案，提供音视图文伪造检测与安全防护技术服务。现已和工信、公安、网信等多个国家部委，新华网、人民网、移动、联通、电信、华为、讯飞等头部企业，手机、笔记本、服务器头部品牌厂商展开合作，全力打造互联网3.0时代的可信数据底座。

方案概况

方案介绍：

中科睿鉴针对音视图文全类型数据，构建“真伪检测-痕迹提取-伪造溯源”全流程技术体系，结合端云协同式治理框架，面向国家安全、企业安全、个人安全打造专业产品与方案。

技术体系上，真伪检测环节全面覆盖深度伪造、AI生成、PS篡改、活体攻击等检测场景；伪造痕迹提取环节挖掘纹理、光照、频谱等多维度特征，突破模型可解释性难题；伪造溯源环节基于指纹提取技术，结合动态进化的伪造技术特征库，定位伪造工具及算法。

治理模式上，面向国家安全场景，采用云端部署睿鉴图灵鉴伪大模型方式，实现公域流量批量检测；面向个人安全场景，在终端设备上部署终端AI鉴伪大师，兼顾安全防护和隐私保护；针对金融、教育、通讯、公安司法等多行业需求，提供专用检测产品及服务，可与现有业务系统无缝对接。

方案示意图/拓扑图：

方案优势和用户价值

方案优势与特点：

- **云端鉴伪大规模化实战验证：**睿鉴图灵鉴伪大模型是首个经千亿次实战验证打磨的检测模型，解决现网数据高压缩、强对抗、伪造手段复杂等难题，实现开放复杂环境下快速精准检测，整体检出率超90%。
- **轻量化终端鉴伪商业化部署：**独家研发终端AI鉴伪产品，对鉴伪模型采用多芯异构适配、量化、软硬协同加速技术，解决检测延时、算力有限、隐私安全等难题，针对手机、电脑、平板等智能终端打造操作系统级别AI安全服务。
- **小样本快速迭代适配新模型新场景：**基于模型、数据、算力一体的体系化能力底座，可利用少量样本在一两周内实现新场景和新型伪造技术快速响应、精准检测。

用户价值：

- **国家安全可信防护：**国产软硬件全面适配，满足数据安全需求；持续迭代鉴伪能力，为未来潜在伪造风险提前布局。
- **企业安全一站式保障：**提供多行业、多形态、多模态保障方案，一站式满足用户多样防护需求，降低管理和防护成本。
- **个人用户终身陪伴：**为个人及家庭提供终身陪伴的AI安全保镖，全面保障学习、生活、工作安全。

典型客户或目标客户

运营商、终端厂商、银行金融机构、高校、个人用户等

大模型安全评估

数说安全

2025年，随着人工智能技术在金融、医疗、交通、政务等关键领域的广泛落地，人工智能系统的安全性问题日益突出。人工智能安全性评估旨在系统性识别、分析和量化AI系统在算法、数据、模型、部署和运行等各环节所面临的各类安全威胁，包括对抗攻击、数据投毒、模型窃取、隐私泄露、偏见与歧视、可解释性不足等。通过安全评估，可有效发现和修复AI系统的潜在风险，提升其在实际应用中的鲁棒性、可信度和合规性，保障关键业务和用户利益。人工智能安全性评估已成为构建AI可信体系、推动AI产业健康发展的重要基础设施和能力支撑。

人工智能面临的安全威胁

核心能力

1. 模型攻击面梳理与威胁建模
2. 对抗鲁棒性 / 越狱测试
3. 数据与模型偏见、毒化检测
4. 隐私与推理泄漏评估
5. 模型可解释性与决策透明性评估

关键挑战

1. 模型版本迭代极快，评估基线需高频更新
2. 权重与训练细节闭源，黑盒评估可信度受限
3. 高质量攻击语料匮乏
4. 多国法规差异化，合规映射与取证复杂度高
5. 评估与仿真需大量算力，中小企业成本压力
6. 供应链深度依赖第三方API，外部依赖风险

应用场景

1. 金融风控与反欺诈系统安全评估
2. 智能医疗辅助诊断系统安全审查
3. 自动驾驶与智能交通系统安全测试
4. 智能政务与公共服务系统安全保障
5. 智能客服与推荐系统风险排查
6. 关键基础设施智能化系统安全评估

典型厂商

君同未来
奇安信
360数字安全 数字安全的领导者
安恒信息 DAS-Security
卡泰科技 CHAITIN
火山引擎
SANGFOR 深信服科技
网宿科技 WANGSU.COM

君同未来-大模型安全评估

数说安全

君同未来

君同未来，专业的人工智能生态治理厂商，构建有一套全面覆盖人工智能政策合规、能力评估、风险防护、监测管控以及人才实训培养的生态治理体系，能力涵盖人工智能全流程模型合规验证、高质量模型对抗增强、多维度模型风险防护、全场景模型监测管控等。

方案概况

方案介绍：

君同生成式人工智能治理方案，主要由生成式人工智能防护管控系统（依托前沿分布式风险检测与托底式输出防护技术架构，在大模型训练、应用阶段持续进行性能优化与安全性强化，实现模型的自适应演进与持续更新）和生成式人工智能评测验证系统（基于海量样本、多元评价、红队测试、MoE专家评测等多样化技术基座，构建应用、安全、RAG及Agent四大评测体系，助力企业开展大模型选型、智能体安全评测和监管单位安全监管与能力摸排，为生成式人工智能的持续优化与进化提供坚实的科学依据和技术支持）组成，覆盖生成式人工智能从开发到应用的全生命周期治理。

方案示意图/拓扑图：

方案优势和用户价值

方案优势与特点：

1、权威的评测体系和智能评测引擎：内置国内外20+评测体系，个性化适配300+评测标准选择，100+场景定制化测评方案，千万量级智能题库，基于特定领域的评测题自动生成能力；实现“合规、能力、应用”多维度评测，灵活的定制化配置能力、交付形态与部署方式。

2、全方位人工智能模型防御：内置50 + 防御及检测方法，实现覆盖多模态的人工智能模型全方位防御，从设计开发、模型训练部署、到模型应用落地的全生命周期，保护系统和算法免受恶意攻击、滥用、数据泄露或不当访问等威胁。

用户价值：

1、为客户大模型应用可信、安全、合规落地提供关键技术支撑，有效满足客户在模型安全合规、风险可控方面的核心需求，显著降低应用风险，保障AI技术可信落地。

2、帮助行业客户完成模型迭代，确保生成内容合规率提升至99%以上，攻击抵御成功率达到97%以上，并通过备案审查，构建起“评测-整改-验证”的闭环备案支撑体系；帮助各行业客户构建覆盖"输入-推理-输出"全链路的智能风控体系，形成端到端安全闭环。

典型客户或目标客户：

产品方案覆盖国央企、互联网、科技、金融、教育、医疗等行业客户，包括网信办、公安部、工信部、人民银行等部委下属机构以及海康威视、字节跳动、蚂蚁集团、涂鸦智能等大型企业。

奇安信-AI大模型安全技术评估服务

数说安全

奇安信

奇安信科技集团股份有限公司（以下简称奇安信，股票代码688561）成立于2014年，专注于网络空间安全市场，向政府、企业用户提供新一代企业级网络安全产品和服务，在人员规模、收入规模和产品覆盖度上均位居行业第一。2019年，中国电子战略入股奇安信，奇安信成为网络安全“国家队”，迄今已完成91次国家重要活动网络安全保障任务，参与近千场实战攻防演习。

作为网络安全领军企业，奇安信将针对新技术下产生的新业态、新业务和新场景，继续为用户提供全面、有效的网络安全解决方案，向成为“全球第一的网络安全公司”的愿景目标不断奋进，助力实现“十五五”良好开局。

方案概况

方案介绍：

大模型安全技术评估服务，基于已发布的法律法规标准，以及大模型安全攻防研究成果，结合自研的大模型安全评估平台和对输出内容的人工对抗，对大模型进行多层次全面的技术安全评估。

方案示意图/拓扑图：

方案优势和用户价值

方案优势与特点：

（1）“全面覆盖”，服务评估内容从内容安全、智能体应用安全、大模型自身安全、大模型供应链应用、大模型运行的基础环境、大模型数据，覆盖多风险维度。

（2）“内容合规”，内容安全方面，结合国内合规环境，测试样本库覆盖国家标准GB 45438-2025（原TC260-003升级）要求的五大类，31种内容安全风险。

（3）“海量题库”，结合国内环境，目前标准题库有1.5W+，可以结合项目利用自研大模型生成针对性问题题库。

（4）“专业工具”，内部自研大模型安全评估平台、云安全评估平台、创新渗透测试平台，可在服务过程中提高效率，降低人天成本。

（5）“攻防能力”，奇安信观星实验室实施，业界一流攻防实战能力和漏洞挖掘能力。

用户价值：

（1）“合规备案”为用户大模型备案材料，发现不合规安全内容，提供大模型的安全评估报告。

（2）“发现风险”有效发现不合规安全内容；自研智能体的数据泄露，越权；大模型组件如Ollama、OpenWebUI等组件的安全问题；发现大模型的提示词注入、数据泄露、幻觉问题；大模型云环境中的安全漏洞。

360-大模型安全评测平台

数说安全

360数字安全科技集团有限公司（360数字安全集团，360 Digital Security Group）是数字安全的领导者，秉持“上山下海助小微”企业使命，确立安全和AI发展双主线。在安全领域，为解决国家“看见”高级威胁的卡脖子问题探索形成一套以“看见”为核心的数字安全中国方案，基于安全即服务理念，将这套方案通过360安全云赋能城市、大型企业、中小微企业。在人工智能领域，推出360安全大模型并率先实现AI实战应用，基于“以模制模”理念，推出数字专家智能体，打造大模型安全卫士解决传统安全和大模型安全问题。

方案概况

方案介绍：

360大模型内容安全评测平台是一款专注于大模型安全评测的自动化工具，面向企业和机构，提供一站式的模型接入、数据管理、安全评测、任务管理及结果分析能力。通过预置的海量安全评测数据集和大模型自动泛化生成的攻击数据集，结合自动化测试与智能判定机制，帮助企业高效、精准地评估大模型在内容安全、合规性、对抗攻击防御能力 等方面的表现。

方案示意图/拓扑图：

方案优势和用户价值

方案优势与特点：

（1）“快”：标准化API多模型接入，一键测试降门槛成本。
（2）“全”：海量安全数据集，自动生成样本保全面时效。
（3）“智”：智能算法自动评测，精准识风险提效率。
（4）“专”：专业报告可视化，趋势对比助决策优化。

用户价值：

本平台为企业、机构及模型商提供自动化安全评测，助其降低合规风险、优化模型安全、提升可信度，增强市场与合规优势。核心价值：
（1） 防风险： 精准检测敏感泄露、违规内容等风险，提供优化建议，规避上线前隐患。
（2）省成本： 自动化替代人工审核，高效完成大规模评测，降成本提效率。
（3）增信任： 提供标准化权威报告，助力备案审查、赢得客户信任，提升竞争力。

典型客户或目标客户：

天津网信办（已合作）

长亭科技-AIGC安全风险评估

数说安全

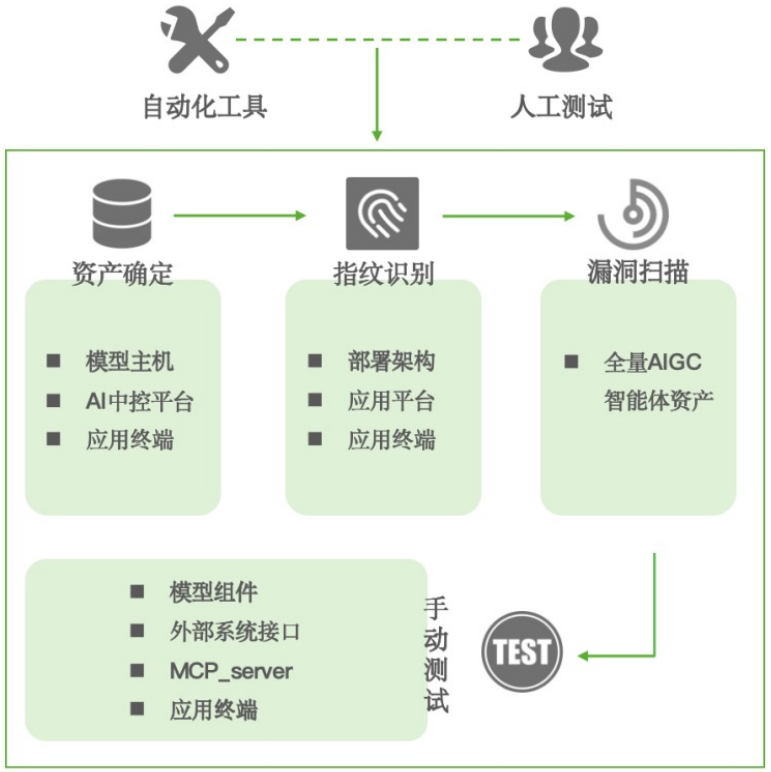
长亭科技 CHAITIN

长亭科技成立于2014年7月，是国际顶尖的智能网络安全公司之一。长亭科技创始团队来自国际知名CTF战队清华“蓝莲花”，在近年来的全国网络安全竞赛以及国家和各地区各行业的实网攻防演练中，荣获冠军近百次，被评为“最了解攻击的防守队伍”。面向数字时代的新型安全风险，长亭科技秉持“知攻善防，智能安全”的理念，从攻击视角构建防御体系，以AI驱动重塑安全运营，通过新一代智能安全产品防护体系和专业的安全测试及咨询服务，帮助金融、通信、政府、能源、互联网等重要行业的4000多家用户，打造了先进可靠的实战化、体系化、常态化的智能安全运营能力。

方案概况

方案介绍：

长亭科技基于在AI安全研究和基础攻防领域的深厚积累，结合自动化工具和人工深度对抗测试，对AIGC应用进行全面的安全风险评估。通过解构AIGC应用和资产明确保护对象，采用风险评估的基本思路开展工作。通过资产维度创建AIGC智能体资产梳理、威胁维度创建AI威胁建模、脆弱性维度创建AIGC安全评估，实现从智能体资产梳理到闭环咨询的全流程服务，帮助客户系统化地识别、管理和降低AIGC应用的安全风险，确保其安全合规、稳定运行。



方案优势和用户价值

方案优势与特点：

- **人机协同评估技术**：结合自动化工具的广度扫描和人工深度分析，模拟真实攻击者的创造性思维，突破传统工具扫描的规则限制，全面识别潜在安全威胁。
- **AI驱动的安全评估框架**：利用长亭自研的“问津”安全大模型，对AIGC应用进行多维度的安全评估，包括内容安全、模型越狱、基础设施安全等。
- **资产梳理与威胁建模技术**：系统化识别和管理AIGC智能体的核心资产信息，结合标准化威胁模型，精准识别攻击面和潜在威胁。

用户价值：

1、帮助AIGC应用安全合规：根据国家相关法律和标准，提供从安全评估到整改建议的全流程服务，帮助客户建立完善的制度、流程和技术框架，确保AIGC应用符合国家相关法律法规和监管要求。

2、降低数据和业务风险：通过提前识别和修复AIGC应用中的安全漏洞，减少因安全事件导致的数据泄露和业务中断风险，让AIGC应用能够更高效更安全地支持业务运营，提升服务效率，带来直接的经济效益。

3、提升系统稳定性：通过人工深度对抗测试、AI辅助判断分析测试和大模型系统专项渗透测试，全面评估相关的安全风险，确保AIGC应用在高负载和复杂攻击环境下的稳定运行。

合规管理&安全运营深度融合

数说安全

随着信息化、数字化技术的飞速发展、网络安全法律法规的密集发布，大部分政企客户的网络环境、网络应用愈发复杂和多样，在网络安全方面需要承担越来越多的工作和责任，而各个区域或行业在网络安全上的投入呈现出缩减的趋势，广大政企客户面临“预算缩减、事情增多、要求提升”的困扰，在应对合规检测/评估、网络或数据安全检查、重保/HW/攻防演练、日常的安全运营等各项工作上面临极大的挑战。

如何能在一定的安全投入情况下更好的将网络安全安全管理和安全技术兼顾支撑到位并越做越好，是大部分政企客户需要解决的问题。供应商应基于以上的情况为客户解决困扰和问题，满足“既要、又要、还要”的客户诉求，将合规管理工作与安全运营工作深度融合，实现管家式的安全服务提供。

核心能力

1. 合规与运营融合管理能力
2. 多维度合规数据管理能力
3. 全量资产梳理分析能力
4. 安全制度标签化管理能力
5. 安全迎检数据自动汇聚能力
6. 智能化合规自查自检能力

关键挑战

1. 服务标准化与客户定制化需求的平衡
2. 技术集成与兼容性
3. 持续运营能力要求高
4. 用户信任、透明度与市场教育

应用场景

1. 头部行业客户合规管理信息化数智化
2. 区域政企领域合规与运营一体化运营
3. 区域行业客户监管与运营一体化方案
4. 小微客户等级保护一站式运营服务

典型厂商

SANGFOR 深信服科技

阿里云

电信安全 China Telecom Cybersecurity Tech

小西牛 XIAOXINIU

以数据为基础的网络安全产业研究平台

ssaq@geniuscybertech.com

2025