AI大模型如何改变安全运营

让 数 字 世 界 更 安 全 可 靠

华青融天（北京）软件股份有限公司　　包彤

智慧城市安全风险分析

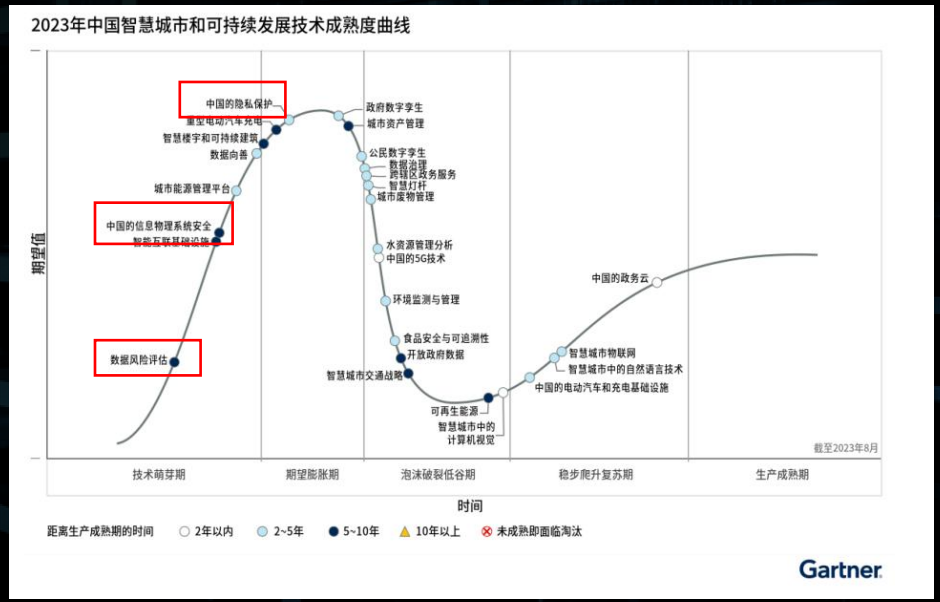
- 关键信息基础设施失效带来服务不可用风险
- 网络攻击或意外原因导致网络瘫痪风险
- 数据集中存储、边界模糊带来数据泄漏风险
- 政府、企业重要信息，个人敏感信息安全管理不严造成滥用风险
- 物联网设备被攻击风险

分层	主要安全风险
应用服务层	应用不可用风险 隐私泄露风险 网络欺诈风险
数字平台层	数据泄漏风险 平台被入侵风险
网络连接层	网络被攻击风险 网络失效不可用风险 无线网络被侵入风险
边缘计算层	设备被入侵风险 设备被攻击风险
终端感知层	终端设备被盗用风险 数据完整性被破坏风险 设备被仿冒风险 设备被干扰风险

2023年中国智慧城市和可持续发展技术成熟度曲线

- **关键城市基础设施和服务** — 利用技术赋能新建和改造市政基础设施，提高城市管理与服务效率，改善市民体验。
- **数字化赋能的可持续发展** — 借助数字化和综合性工具减轻环境影响和碳足迹，提升城市韧性。
- **赋能数字技术** — 这些技术要素为各个智慧城市所需的组合式创新赋能。
- **基于信任的数据驱动生态系统** — 通过一系列与数据相关的基本举措和活动，在智慧城市范围内，构建可信的协作式城市生态。

当务之急是建立一个系统的安全保护体系，提高信息物理系统的安全性，减少安全风险。



让数字世界安全可靠 - 以业务和数据为核心， 构筑四道防线

Fusionskye 融天

信息科技风险分类

- 1. 业务连续性风险
- 2. 业务可靠性风险
- 3. 网络和数据安全风险
- 4. 监管合规性风险

各级监管机构： 政策制定、数据报送、监管通报

国家监管机构、属地监管机构、行业监管机构等

内部审计部门： 检查审计、评价及整改建议

设立专门的信息安全风险审计岗位

风险管理部门： 持续风险评估、跟踪整改落实

风险管理部门、合规检查部门

信息科技部门： 准确识别并及时处置各种安全风险隐患

开发部门、运维部门、安全部门等

华青融天：基于大数据与AI，为数字世界构建纵横联动的防控体系

- 全面掌控：通过报送/直采方式掌握整体安全运行状况，风险监控识别
- 联动赋能：联防联控、共性分析及通报指挥、行业性能力沉淀

- 合规监测：行业规则嵌入系统，结合行为数据，实现合规管理
- 事中控制：通过实时数据监控行为和风险，及时风控，数据留痕

- 快速处置：快速发现并定位风险，并持续掌握和提升质效
- 实时感知：针对网络和信息安全风险，实时感知和处置闭环

AI开启数字世界新时代

Fusionskye 融天

1. 大语言模型在2023年进入爆发期

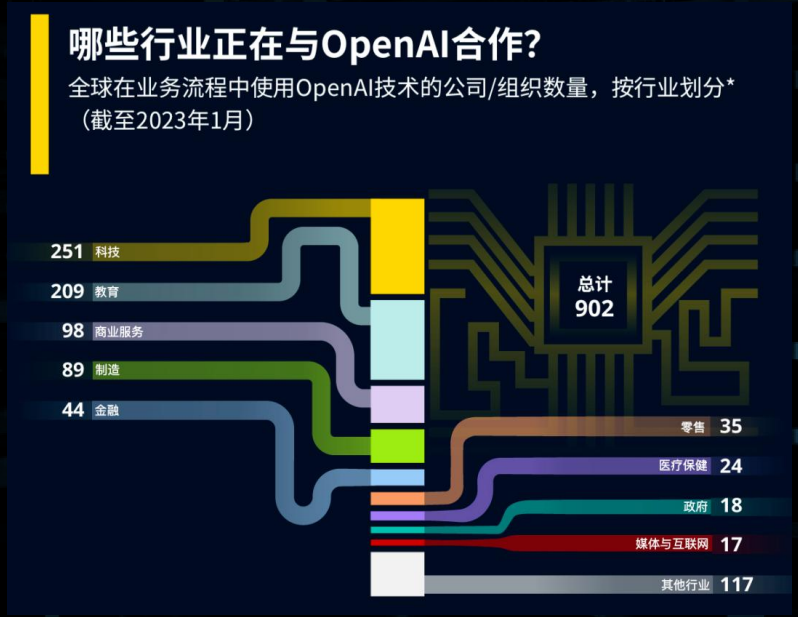
- 2023年3月， OpenAI 发布GPT4，参数量为1750亿，预训练数据45TB
- 国内外厂商普遍展开大语言模型的相关研究，包括谷歌、微软、百度、科大讯飞、华为等

2. 大语言模型逐步发展为多模态大模型

- GPT4提供丰富的扩展插件，并提供开放API

3. 大模型开始在垂直领域发力

- OpenAI的合作企业达到数百家，涉及十个以上行业



AI大模型在网络安全领域研究探索

Fusionskye 融天

1. 针对大模型在网络安全领域的研究逐步出现，并尝试应用在攻防两方面

- 基于ChatGPT的自动编程、程序分析工具
- 基于ChatGPT的安全情报分析工具
- 基于ChatGPT的渗透测试工具
- 基于ChatGPT的滥用（钓鱼邮件生成）

2. 大模型在网络安全领域仍缺乏系统研究

- 如何挖掘应用场景?
- 如何进行垂直领域优化?
- 如何进行能力和成熟度评估?
- 如何进一步评估和保障大模型本身的安全性和可信性?

ChatGPT安全工具	主要功能
ChatGPT Scanner	基于 ChatGPT 的扫描工具
GPT_Vulnanalyzer	使用 ChatGPT、Python-Nmap 和 DNS Recon 等模块的漏洞分析器
PentestGPT	基于 ChatGPT 的自动化渗透测试工具
IATelligence	基于 GPT-3 的恶意软件分析、恶意软件行为评估概念性 Python 工具
GPT-wpre	基于GPT-3 的全程序逆向工程原型系统
Falco-GPT	Falco 事件审计及 A1 自动化处置工具

AI大模型在城市安全运营领域应用能力分析

应用场景

Web攻击检测
钓鱼邮件检测
数据泄漏检测
恶意代码检测
漏洞利用检测
隐蔽隧道检测
智能渗透测试
智能漏洞修复
智能SOAR
智能安全问答
智能安全专家

AI大模型安全运营服务设计

Fusionskye 融天

训练数据集构建

漏洞库 | 知识库

规则库 | ……

通过迁移学习，实现大模型在垂直领域的性能优化

信息来源

IDS | SIEM

SoC | ……

对安全情报实施诊断和推理分析

工具接口

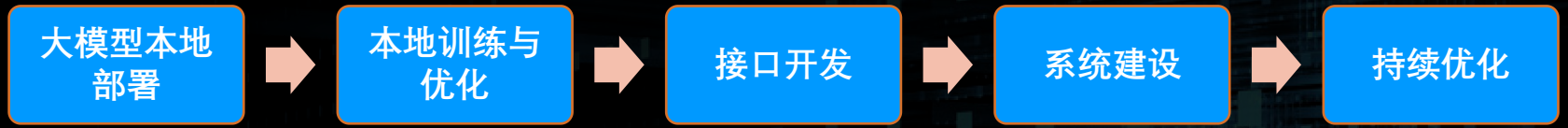
人机交互 | 渗透工具

防火墙/IDS | ……

实现自动（处置）流程和人机交互

大模型安全运营服务

AI大语言模型的构建与训练



典型应用场景


自动化渗透测试
自动化流程
推理与诊断能力
连续交互能力
自动化安全事件监测、追踪与分析
事前检测
事中分析
事后处置
威胁情报分析与处理
威胁情报管理
舆情分析

智慧城市一体化安全运营平台建设目标

平台架构

Fusionskye 融天

展示层

安全态势 | 安全工作台 | 安全监测 | 安全分析

资产管理 | 响应协调 | 运营管理 | 组件市场

应用层

资产管理	漏洞管理	预警管理	处置管理	情报管理	主题库
● 租户机构管理 ● 资产信息管理 ● 资产信息接入	● 漏洞信息分析 ● 漏洞信息管理 ● 漏洞信息接入 ● 漏洞趋势统计	● 安全预警通报 ● 预警模板管理 ● 预警通知管理	● 风险处置通报 ● 处置级别管理 ● 风险通知管理 ● 处置模板管理 ● 风险处置反馈 ● 处置统计分析	● 情报对接 ● 情报关联 ● 情报碰撞 ● 本地情报	● 组件日志库 ● 告警库 ● 白名单库 ● 模型库

分析层

高危资产分析模型 | 共性安全风险分析模型 | 高可信安全风险分析模型 | 场景建模 | 高风险安全事件分析模型 | 综合态势评分模型 | 资产脆弱性分析模型

分析建模中心

统计模型 | 规则模型 | 关联模型 | AI模型

数据层

实时分析引擎 | 离线分析引擎

数据处理 | 数据解析/富化 | 数据存储 | 数据管理

资产数据接口 | 安全组件告警数据接口 | 第三方数据采集接口

通报预警派发

通报预警反馈

大模型分析举例

Fusionskye 融天

Web攻击样本分析

请对下面代码进行逐项分析，给出潜在风险和安全问题。 针对这些潜在的风险和安全问题，需要采取哪些 措施进行处理?

<frame oncut=\"alert(1)\" contenteditable>test</frame>\n<ul class=\"simple\">\n<frameset onmouseover=\"alert(1)\">test</frameset>

XSS攻击

1. 跨站点脚本（XSS）攻击风险：这段代码中的 `<frame oncut="alert(1)" contenteditable>` 部分包含了一个 JavaScript 函数调用来触发一个警告弹出框。而这种设置可能导致未经授权的用户在该代码中注入恶意脚本，从而执行 XSS 攻击。建议在这段代码中添加适当的输入过滤和验证机制，以确保只允许合法的内容。
2. 浏览器支持性问题：这段代码使用了 HTML5 标签 `<` 属性 `.simple`，但这些标签和属性可能不兼容于某些旧版本的浏览器。如果目标用户使用的浏览器不支持 错误或者其他不可预期的行为。因此，可以考虑在代码中加入相应的浏览器兼容性解决方案，例如引入 P 。
3. 信息泄漏风险：这段代码没有明确指定任何敏感信 由于它涉及到网页元素的位置和样式，存在一些隐私敏感的信息可能会被暴露给第三方。为了减少信息泄露的 上启用 HTTPS 协议，并在发送数据时采用加密方式保护用户隐私。
4. 安全性测试需求：对于这样的代码片段，我们还需要对其进行全面的安全性测试，以发现潜在漏洞和弱点。这可以通过利用自动化工具如 Burp Suite、Nessus等进行扫描和渗透测试，以便及时修补潜在的安全问题。

大模型分析举例

Fusionskye 融天

告警事件分析

解读下这段威胁日志:

[{“attacker_city”:“Beijing”,“kill_chain”:“attempt-attack”,“dst_ip_geoloc”:“34.7732,113.722”,“victim_geoloc”:“34.7732,113.722”,“hostip”:“124.205.141.206”,“dst_ip_country_code”:“CN”,“rid”:“4021929”,“mat_off”:[73],“uuid”:“6f2ab6b3-838e- 4f7f-aab6- b3838e0f7f29”,“number”:1,“hostname”:“ATD”,“payload”:“”,“dst_ip_isHost”:0,“commid”:“1:DFap72vZJR7OIhIiZDxPwI7Rkmk=”,“attack_status”:“attempt”,“victim_city”:“”,“attacker_country_code”:“CN”,“dst_ip_country”:“China”,“victim_country_code”:“CN”,“attacker_isHost”:0,“proto”:“tcp”,“enrichments.src_ip.malicious”:“2”,“http”:{“resp_version”:“HTTP/1.0”,“status_code”:“200”,“method”:“GET”,“server_header_names”:“Server: FSM-Webs\r\nPragma: no-cache\r\nCache-control: no-cache\r\nConnection:close\r\nContent-Type: text/html;charset=gb2312\r\n\r\n”.....}]

大模型分析举例

Fusionskye 融天

分析结果和建议

威胁日志。从日志中可以看出，攻击者使用了一个名为"zgrab"的扫描工具，尝试向一个位于中国北京的主机发送恶
式可能是利用漏洞或者未经授权的方式获取到该服务器的信息。
析这份日志，我建议你参考以下几个步骤：

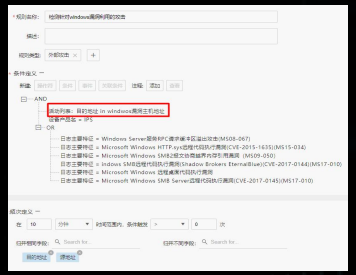
确认所涉及的主机是否已经安装了最新的补丁和防火墙规则。如果发现存在已知的漏洞，应及时修补或升级相关软件版本。

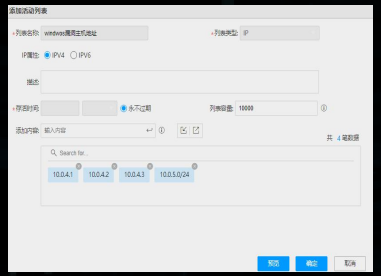
2. 通过检查日志中的源IP地址，判断是否为黑名单或可疑IP地址。如果是，可以考虑加强防护措施，如封锁相应的IP地址段等。
3. 根据日志记录的内容和行为，评估攻击者的意图和目的。比如，查看是否涉及到敏感数据或关键系统的访问，是否有明显的破坏企图等。
4. 对于类似的威胁事件，建立相应的响应计划和应急流程。一旦发生类似情况，可以按照事先设定好的预案快速反应，降低损失的可能性。
5. 最后，建议加强对网络安全培训和意识教育，提高员工和用户的安全素质，避免因为疏忽或错误导致信息泄露等问题的发生。

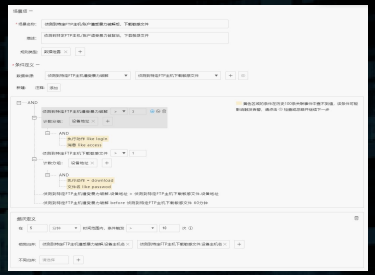
综上所述，综合考虑以上因素，可以有效提升整个组织的网络安全水平。如果有其他问题或疑问，欢迎随时咨询。

一体化运营平台 - 安全场景建模

Fusionskye 融天







事件特征分析 → 事件分类聚合 → 事件群体关系 → 事件时序/主体关联 (多维事件关联模型) → 通报/监测告警

勒索病毒事件

FW：大量终端的445端口异常连接

IPS：大量终端MS17-010 漏洞利用

流量探针：发现创建异常服务文件
遍历磁盘文件,异常DNS请求

中间结果

终端补丁检查：
存在MS17-010漏洞缺失

中间结果

NSA网络武器库被黑泄露，
勒索病毒有爆发流行的趋势

关基机构受到勒索
软件攻击

暴力破解事件

FW: 服务器22端口被扫描

FW：log in 登陆报文次数较多

IPS：发现爆破工具Hydra指纹

中间结果

目标服务器：
反馈多次登陆失败事件

目标服务器密码被暴力破解

让数字世界更可靠

Fusionskye 融天

发现看不见的价值

See what's invisible