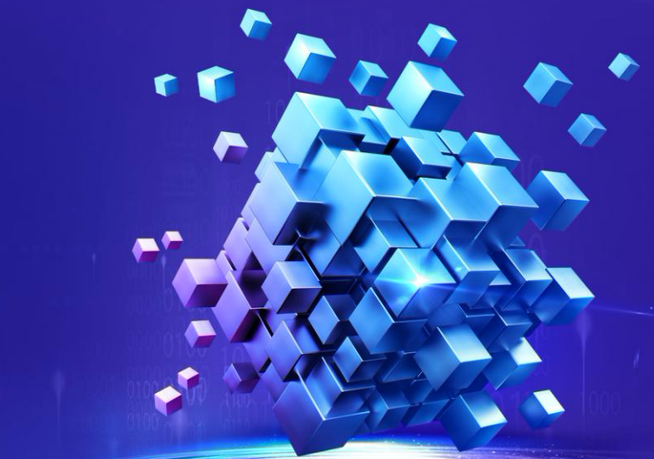
2021 WEST LAKE CYBERSECURITY CONFERENCE
西湖论剑·网络安全大会

2021 安全：数字化改革之根基

CYBERSECURITY :
THE FOUNDATION OF DIGITAL REFORM

西湖论剑·网络安全大会



2021 WEST LAKE CYBERSECURITY CONFERENCE
西湖论剑·网络安全大会

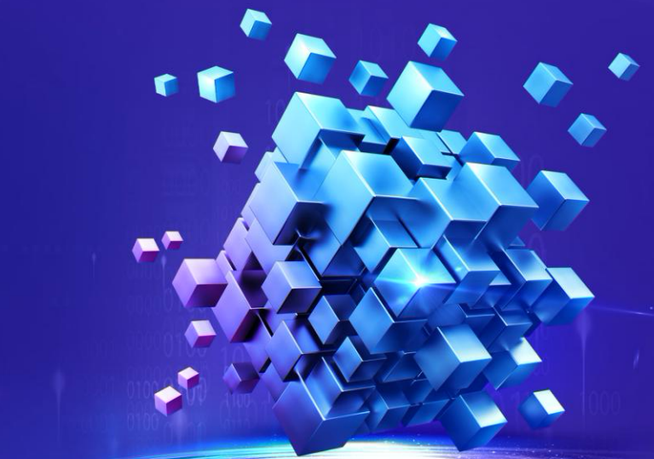
2021 CYBERSECURITY : THE FOUNDATION OF DIGITAL REFORM

安全：赋能数据开放、激活数据价值

演讲人：刘博

杭州安恒信息技术股份有限公司 首席科学家

之江实验室网络安全研究中心　副主任



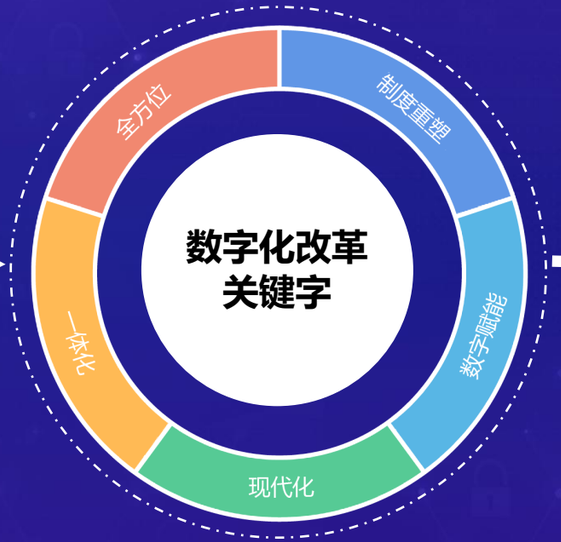
数字化改革时代的背景

“十四五”规划中，明确要求全面加强网络安全保障体系和能力建设，切实维护新型领域安全，网络安全、数据安全保障体系和能力是对国家安全的有力维护。

- 中央明确数据作为第五生产要素
- 打通数据孤岛，进入政府数字化转型时代
- 数字化成为主导国际竞争的关键力量
- 数字经济引领技术变革和产业升级
- 数据安全将上升至法律层面

2020年7月3日，《中华人民共和国数据安全法（草案）》
2020年10月21日，《中华人民共和国个人信息保护法(草案)》

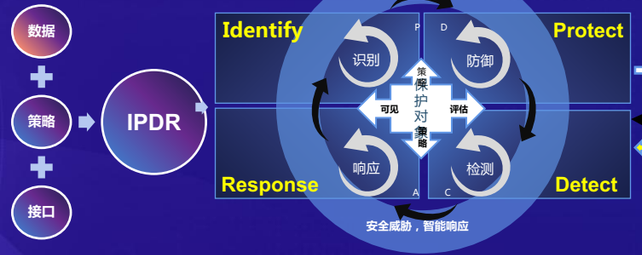
网络安全



数据安全

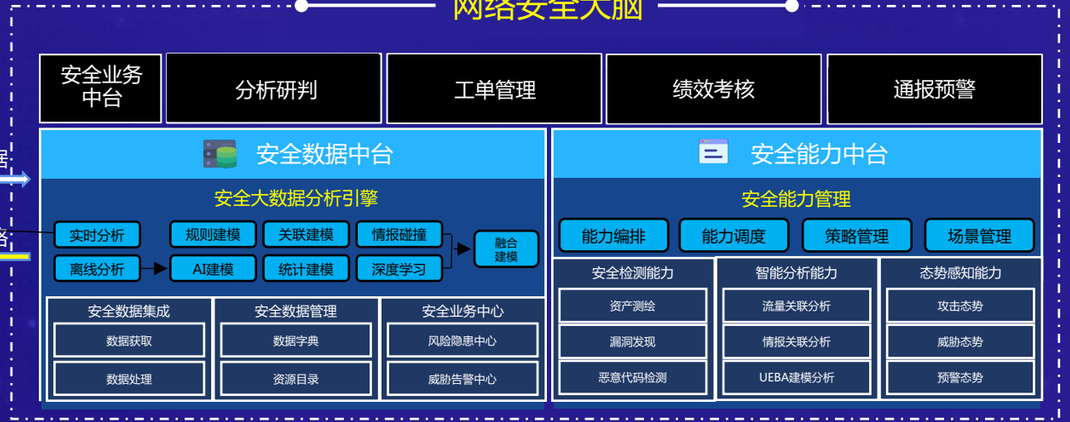
- 打造数据经济新优势
- 营造良好数字生态
- 提高数字政府建设水平

网络安全大脑



数据

策略



数据

动作

网络安全态势感知(监管）

安全服务
- 大会安保
- 攻防演习
- 漏洞评估
- 事件追溯

运维服务
- 基础保障
- 系统维护
- 培训服务
- 日常巡检

辅助工具
- 信息安全等级保护检查工具
- 工业控制系统安全检查工具
- 网络安全事件应急处置工具

安全体系
- 物理安全
- 主机安全
- 系统安全
- 业务安全
- 大数据安全
- 容灾备份
- 安全审计

监管应用层



业务系统：等级保护 | 实时监测 | 态势感知 | 重大安保 | 情报侦察 | 侦查调查 | 通报预警 | 指挥调度 | 行政执法 | 攻防演练

事前感知预警 → 事中防护响应 → 事后取证优化 → 持续监测分析

创新　　赋能

AiLPHA关键信息基础设施安全保卫平台

数据分析	事件分析处理	威胁感知建模	事件关联挖掘	事件追踪分析	样本深度分析
数据存储	原始库	资源库	主题库	知识库	业务库
数据处理	接入子系统	处理子系统	存查子系统	分析子系统	建模子系统

数据采集层

基础设施层：计算基础设施 | 存储基础设施 | 网络基础设施 | 其他基础设施

标准规范
- 采集探针
- 平台接口
- 数据规范
- 业务流程

业务机制
- 等保管理
- 通报预警
- 应急处置
- 指挥调度
- 护网演习

新监管态势感知：智能化、实战化

全息档案 + 挂图作战 + 实战化新监管

实战化新监管

- 智能协调指挥体系
- 全息档案
- 新业态监管
- 挂图作战

网络空间

AiLPHA政企智能安全大脑：纵深防御、协同联动

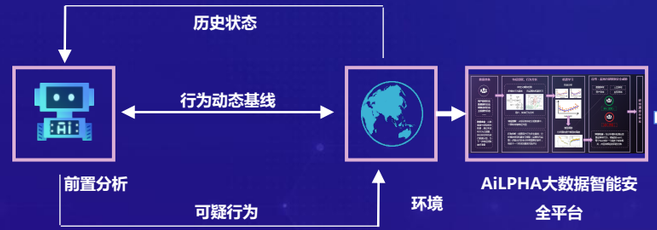
核心价值

- ✓ 整合现有安全资源，构建纵深防御体系
- ✓ 大数据高可用架构，稳定高效灵活扩展
- ✓ 场景化AI安全分析，威胁狩猎精准告警
- ✓ 智能编排决策响应，聚焦实战运营增效
- ✓ 级联管理策略协同，助力集团安管闭环
- ✓ 联动监管风险预警，构建城市安全生态

企业网络安全态势感知建设案例

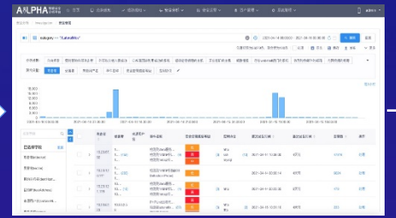
大数据智能分析识别0Day攻击

① 攻击者通过构造恶意的钓鱼链接；
② 企业员工打开该钓鱼链接；
③ 某电脑版聊天软件成为攻击目标；
④ 触发谷歌浏览器的0day漏洞；
⑤ 加载远程JS代码并执行；
⑥ 该员工PC被远程控制，渗透内网



AiLPHA大数据智能安全平台

级联协同汇聚威胁要素

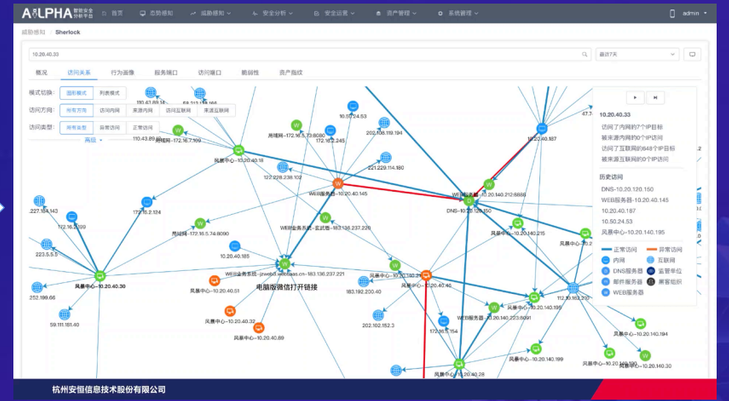


检测到内网主机下载外部JS文件行为

攻击过程推理溯源

溯源取证引擎

知识分享自动响应处置



企业网络安全态势感知建设案例

大数据智能分析识别0Day攻击

下级单位威胁汇聚

事件信息	事件类型
	情报IOC
	受害者信息
	处置建议
	……
补充信息	文件检测报告
	……

接入汇聚

级联协同汇聚威胁要素

数据预处理

日志格式兼容	固有属性补充
IP归属地补充	名称格式化
单位信息补充	属性存在性检查
格式合法性检查	格式转换处理
缺省值补充	……

本地威胁验证

云端威胁分析中心

威胁知识

行业威胁知识库

知识分享自动响应处置

城市网安监管中心

企业网络安全态势感知建设案例

大数据智能分析识别0Day攻击

行业威胁知识库

获取威胁知识

风险类型	某软件0Day漏洞
情报IOC	Jis****@163.com
检测策略	**********
处置建议	升级软件版本至最新
其他信息	……

级联协同汇聚威胁要素

安全策略同步

知识分享自动响应处置

触发SOAR剧本

自动对攻击行为进行阻断，查杀风险终端，结果告知安全管理员

企业网络安全态势感知建设成果

承担国家级、省级示范课题76项，国家级课题30项，参与制定国家标准13项，参与制定行业标准7项

- ✓ 2018年CCIA优秀网络安全创新产品
- ✓ Breakout Security Information Event
- ✓ Management (SIEM) InfoSec Award for 2019
- ✓ 2020网络安全金帽子奖--UEBA
- ✓ 2019年世界人工智能大会十大安全创新实践
- ✓ 数博会2019年领先科技成果、数博会2019大数据产品百佳案例
- ✓ 2019年大数据优秀产品和应用解决方案案例
- ✓ 2020年ICT中国创新奖“ICT中国创新应用特别贡献奖”

保障 Protect

网络安全

赋能 Enable

数据安全

数据经济时代数据安全事件频发

过去2年，数据泄露造成的经济损失增长了**31%**

《内部威胁成本全球报告》趋势对比

内部数据泄露平均成本（单位：万美元）

876（2018年） → 1145（2020年） 31%↑

内部数据泄露事件数量（单位：件）

3200（2018年） → 4716（2020年） 47%↑

2020年报告数据

事件平均处理周期

77天

增长最快行业

零售业

38.2%

损失最严重的三个行业

金融服务业： 1450万美元

服务业： 1231万美元

IT行业： 1230万美元

数据来源： Ponemon研究所2020 "COST OF INSIDER THREATS GLOBAL REPORT"

2018年7月 — 近5亿条华住旗下酒店信息被公开售卖

2018年8月 — 倒卖个人信息数百亿条（含cookie、URL等），新三板公司（市值21亿）停牌！业务取缔、员工入刑

2018年9月 — Google+API接口漏洞导致50万用户数据泄露。谷歌决定永久关闭这一服务

2018年10月 — 隐藏用户数据泄露事件拒不报告，与各州政府达成1.48亿美金和解（IPO前重大诉讼风险）

2018年11月 — 拉考征信、AdMaster、钱宝网、宝付支付、榕树贷款、有脉金控等多家金融大数据公司被调查，相关责任人被约谈

2018年第四季度 — 万豪酒店旗下喜达屋酒店被曝光遭到网络攻击，有5亿名客户信息泄露，被索赔125亿美元

2019年初 — "社保掌上通"通过不合理条款获得用户授权，在用户注册、查询服务，将用户的社会保障号、社保查询密码等个人敏感信息传送至第三方服务器

2019年3·15晚会 — MongoDB、ElasticSearch曝出未授权访问安全漏洞，超2亿中国用户简历泄露

2019年6月 — 普华永道因处理员工个人数据缺少合法依据被希腊数据保护执法机构HDPA处以15万欧元的罚款

2019年7月 — 今日头条被用户起诉，未明确告知和同意，擅自读取用户通讯录推荐好友

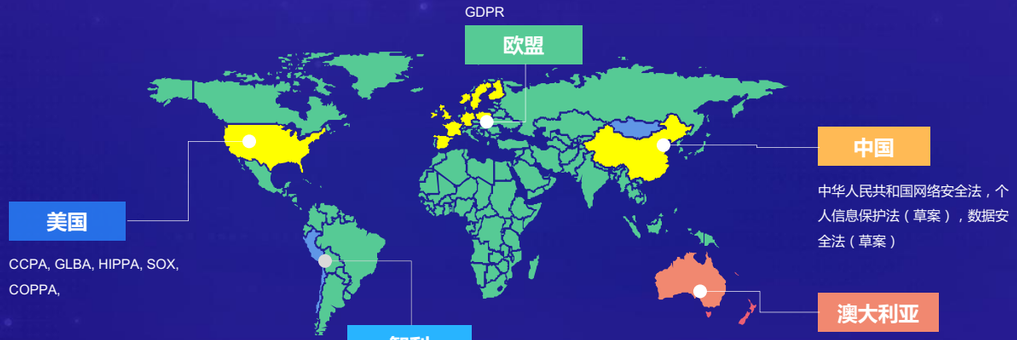
2019年9月 — 内部员工将用户身份信息、电话号码、余额、交易记录等信息售卖谋利，涉及公民5万多条，涉案金额2000多万元

2020年5月 — YouTube因违规处理儿童个人信息，被FTC开出1.7亿美元罚单

数据安全受到全球各个国家的高度重视

全球107个国家和地区已制定数据安全和隐私保护法律

欧盟：《通用数据保护条例（GDPR）》
美国：《加州消费者隐私法案（CCPA）》
中国：《网络安全法》
《数据安全法（草案）》
《个人信息保护法（草案）》



资料来源：中国电子标准研究院

十四五规划

全文“数据安全”出现了**5次**，“数据要素”出现了**4次**，数据安全已成为国家、社会发展面临的重要议题。

数据安全建设融入到各个篇章中，对建设数字化中国和打造网络安全强国做出了重要部署，政策导向明确，数据安全的监管力度清晰，数据安全在未来仍将是一个重大挑战。

赋能数据开放、激活数据价值

赋能数据开放、激活数据价值

数据安全保护能力建设路径

确立顶层设计

数据安全整体规划

业务场景 | 用户行为

数据安全 | 基础设施

安全管理 | 安全技术 | 安全运营

战略支撑基础：领导重视、资源保障、法规支持

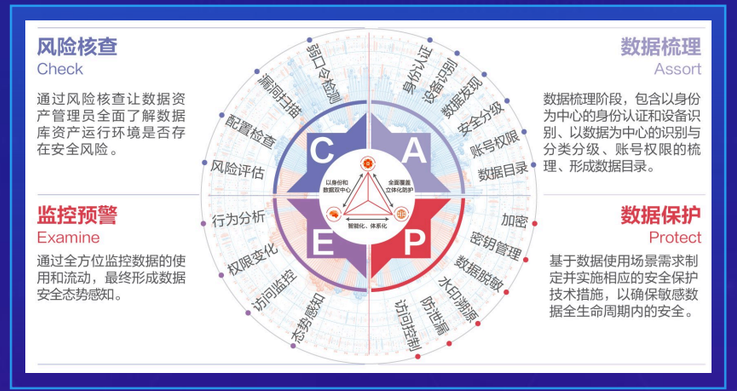
梳理数据安全风险，完善能力框架

组织有多少数据，数据如何分布？

什么是敏感数据，敏感数据在哪里？

谁有权访问组织的数据？

是否采取了防护、监控、告警等措施？



数据全生命周期监管

数据安全十大风险

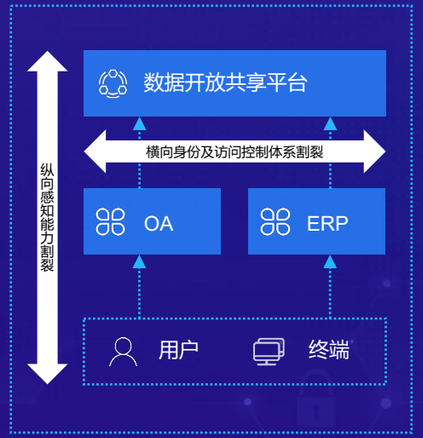
数据自身安全风险

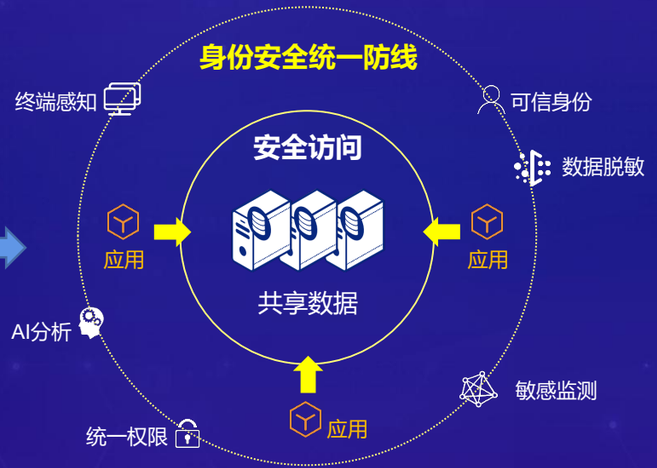
- 账号安全风险
- 权限失控风险
- 数据未分类分级
- 数据库漏洞利用风险
- 数据存储风险

数据使用安全风险

- 终端数据泄露风险
- 数据共享泄露风险（API安全）
- 数据泄露后，无法溯源和追责
- 审计线索不足，数据泄露无法追溯
- 开发测试环节数据泄漏风险

杜绝数据共享泄露风险：API安全隐患成为核心数据泄露的典型途径





身份安全统一防线

以零信任理念，从最终用户访问开始，构建全局可信数字身份体系

终端、应用、API、数据四重防线，最小化共享数据潜在暴露面

智能安全能力底座

终端安全环境联动感知，保护数据端到端的安全流转

基于大数据的智能分析能力，全流量接入，潜在异常与威胁无所遁形

数据安全能力加持

完整的监测+防护一体化解决方案，预警与处置闭环

水印、敏感数据监控、分级分类等能力加持API安全防护

保障数据安全的使用：数据脱敏

某医疗单位的肿瘤样本数据交付给第三方研究机构进行数据分析，如何保障分析结果的前提下，不泄露敏感数据？

运维侧人员访问敏感数据

能接触到开发敏感数据的人员多，组成复杂，管理难度大，存在敏感数据泄露风险。

生产数据跨区域流转至开发库

生产数据由高级别安全区域向低级别安全区域流转。

第三方公司人员访问敏感数据

第三方人员可接触到生产系统或生产数据，可能违规进行数据导出操作。

保障数据安全的使用：国内首家支持面向机器学习任务的脱敏算法（1/2)

某医疗单位的肿瘤样本数据交付给第三方研究机构进行数据分析，如何保障分析结果的前提下，不泄露敏感数据？

- ✓ 保障脱敏后在多个业务数据源中的患者信息仍保持可关联性；
- ✓ 脱敏后的数据可保留原数据分布不变，以及保留其各类常用统计特征，即保障分析结果不因脱敏而影响；
- ✓ 数据泄露之后，可以根据脱敏的信息进行溯源。

注：所有数据均为示例均为实验室仿照的数据，不涉及任何真实人员。如匹配到真实信息，纯属巧合。

一、样本编号：使用一致性关联算法，对样本编号（图中身份证号）进行脱敏，使得脱敏后的样本编号在多个业务数据源中保持一致，保留了多数据源中患者信息的可关联性，保障了后续分析任务中数据的丰富性

脱敏前数据

表1：患者基本信息表

身份证号	姓名	性别	籍贯
330521198702174626	张三丰	男	杭州
37078219650816137x	王小雨	女	北京
110363197804213422	张驰	男	上海

表2：患者诊断表

身份证号	肿瘤大小	肿瘤颜色	诊断结论
110363197804213422	5.596	暗红	良性
37078219650816137x	20.247	黑	恶性
330521198702174626	10.125	深红	良性

“身份证号”

一致性关联算法

脱敏后数据

表1：患者基本信息表

身份证号	姓名	性别	籍贯
340721192707124636	张大山	x	h
230522197603165768	王大雨	y	b
330463198411083422	张飞	x	s

表2：患者诊断表

身份证	肿瘤大小	肿瘤颜色	诊断结论
330463198411083422	3.229	a	A
230522197603165768	23.865	b	B
340721192707124636	9.544	c	A

20+数据库

支持MySQL、Oracle、MS SQLServer、ElasticSearch等关系型&非关系型20多种数据库

50+敏感数据识别算法

包括机器学习、NLP、文档指纹等先进AI技术创建识别规则、实体识别模型。包括复杂姓名、复姓、手机号、身份证、车牌号等

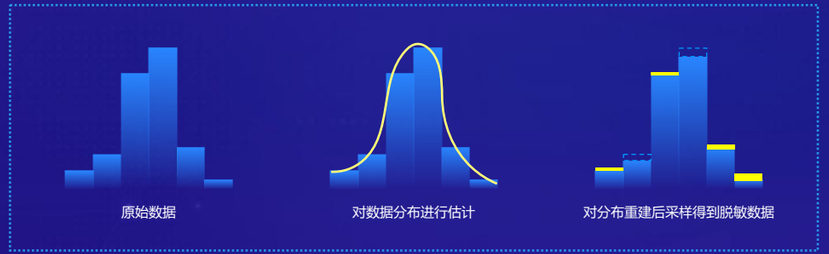
内置主流行业法规、保障数据合规

内置了包括网络安全法、金融、证券、电信、GDPR、CCPA、等保、数据安全法和个人信息安全规范等数十种法律法规

保障数据安全的使用：国内首家支持面向机器学习任务的脱敏算法（2/2)

二、样本特征：以肿瘤大小这一类数值类型特征为例，使用分布重建的脱敏算法，脱敏后的数据可保留原数据分布（图右），可保留各类常用统计特征（图左）。

肿瘤大小	脱敏前	脱敏后
最大值	30.753	30.774
最小值	5.596	4.729
均值	15.343	15.332
标准差	7.545	7.653
中位数	9.524	9.765



三、样本标签：主要针对枚举类型的样本标签（图左），可使用保留类别频次特征算法，无需先验映射字典，自动根据标签分布对其进行编码，去除字段内容含义，仅保留类别区分性。对样本类别属性特征（如肿瘤颜色）也可使用该算法进行脱敏。

诊断结论	脱敏前	脱敏后
	良性	A
	恶性	B
	良性	A

肿瘤颜色	脱敏前	脱敏后
	暗红	a
	黑	b
	深红	c

价值：面向机器学习的脱敏算法在剔除数据敏感性的同时最大程度保留了AI建模相关数据的可用性，即 样本编号+样本特征+样本标签=高质量的AI模型

注：所有数据均为示例均为实验室仿照的数据，不涉及任何真实人员。如匹配到真实信息，纯属巧合。

保障数据溯源的隐蔽性和可靠性：智能水印溯源

算法名称	应用场景	特点
伪行种子算法	对单条数据的查询	数据查询具有真实性，不适用数据统计
伪列种子算法	对数据的统计查询	数据统计不失真，不适用单条数据精确查询
仿真种子算法	数据的查询统计分析	对用户名、密码等数据高度仿真
文本属性算法	文本型数据查询	通过不可见字符嵌入水印
数值属性算法	包含数值数据的查询	应用范围较广，实用性强，但不适用于数据统计

原始数据

表：患者诊断表

身份证号	肿瘤大小	肿瘤颜色	诊断结论
110363197804213422	5.596	暗红	良性
37078219650816137x	20.247	黑	恶性
330521198702174626	10.125	深红	良性
632121197802097667	3.417	暗红	良性

水印嵌入后数据

表：患者诊断表

身份证号	体重	肿瘤大小	肿瘤颜色	诊断结论
110363197804213422	67.23	5.596	暗红	良性
37078219650816137x	50.94	20.245	黑	恶性
330521198702174626	73.56	10.127	深红	良性
530181201301260622	64.32	15.343	暗红	良性
632121197802097667	48.35	3.417	暗红	良性

1、基于DeepFM等推荐算法，根据业务场景智能推荐并生成仿真列（图中红框的体重；根据患者，肿瘤等医疗场景自动推荐生成符合业务场景且不影响后续分类任务的体重特征，数据使用者无法察觉该列为水印仿真列）

2、基于核密度估计（Kernel Density Estimation），使得智能插入的高仿真行对数据分布无影响，不影响AI模型的训练结果（图中红底所在行，不影响业务可用性）

3、通过智能算法计算对某些位置的数值型数据进行最低有效位（Least Significant Bits）修改（图中蓝框红字，基本不影响数值计算及判断的使用场景），对文本型数据嵌入空格、回车符（图中黑块）等不可见字

注：所有数据均为示例均为实验室仿照的数据，不涉及任何真实人员。如匹配到真实信息，纯属巧合。

赋能数据开放、激活数据价值

数据要素的背景（政策）

中国数据要素市场激活

政策面

2017年12月

央政治局第二次关于“实施国家大数据战略“提出“数据是新的**生产要素**，是**基础性资源**和战略性资源，也是重要生产力”

2019年10月

十九届四中全会提出健全劳动、资本、土地、知识、技术、管理、**数据**等生产要素由市场评价贡献、按贡献决定报酬的机制

2020年4月

《中共中央 国务院关于构建更加完善的要素市场化配置体制机制的意见》提出“加快培育**数据要素市场**”

2020年7月

2020年7月《数据安全法（草案）》、《个人信息保护法（草案）》指出对政府数据开放和个人隐私保护的重要性

地方政府陆续出台

《浙江省公共数据开放与安全管理暂行办法》、《上海市公共数据开放管理办法》、《福州市公共数据开放管理暂行办法》、《广西公共数据开放管理办法》（征求意见稿）等

实践面

北京大数据交易中心

交易中心出台关于数据要素市场建设、定价和交易相应的规范和条例，探索政企数据必须要经过大数据交易中心模式，保障政企数据安全

上海临港推动数据跨境流通试点

上海临港探索汇集跨境贸易、金融、交通领域的数据汇聚和定价策略，促进跨境数据、国内政务数据、国内企业数据的流通和交易

数据共享的难点和技术风险

数据安全岛：实现数据开放可用不可见

技术优势

✓ BDTEE可信执行环境

- 基于大数据环境下数据要素特点，构建大数据可信执行环境(BDTEE)技术
- 基于单向网闸、防火墙和权限管控，将数据域、计算域和操作域分离

✓ 隐私计算

- 综合运用MPC多方安全计算、差分隐私、PSI隐私求交集、半同态加密和联邦学习解决个人信息隐私保护

✓ 数据全生命周期安全

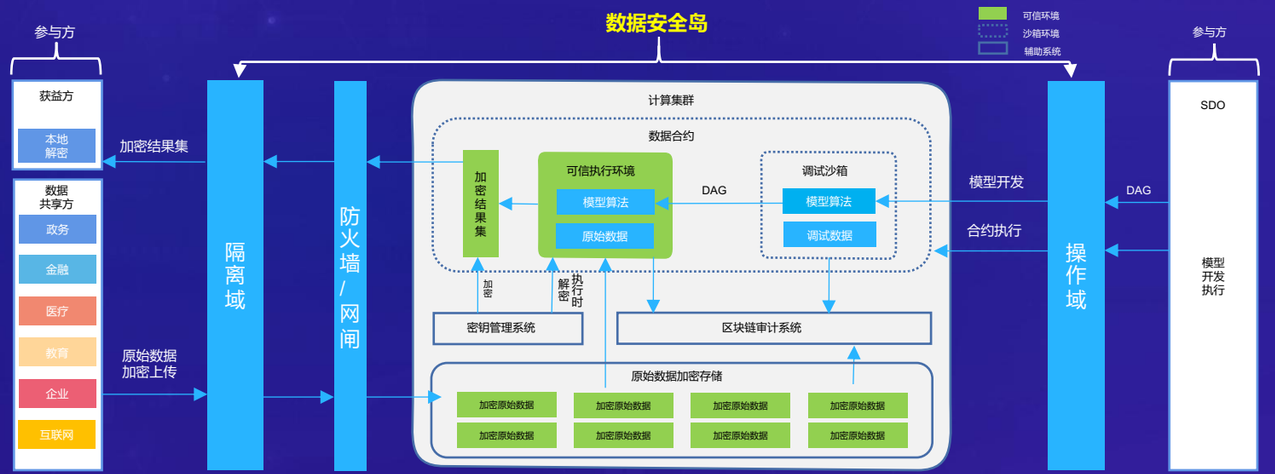
- 数据采集所有权管理、数据全生命周期加密（数据集加密上传、数据密文存储和数据集密文下载）、数据销毁和区块链存储审计信息等技术保障数据安全
- 基于安全运营中心的数据安全态势感知、安全合规审批和云安全防护维度，保障安全岛平台的网络安全、云平台安全和数据风险持续监测

BDTEE可信大数据执行环境

国内首创大数据环境下的可信执行环境

安全功能

- **Isolation-执行环境隔离**
  调试和执行环境分离，数据可用不可见
- **Attestation-身份验证**
  访问用户身份需私钥签名认证
- **Sealing-数据加密**
  数据全生命周期加密，支持国密算法
- **Tracing-溯源**
  用户操作、审批等行为通过区块链固化，可追踪溯源
- **Verifiable-可验证**
  执行过程非黑盒，安全可验证



数据安全岛主要的应用场景

大数据局：

政务数据开放应用、赋能智慧城市和数字经济

应用场景：政务数据开放，营造良好的营商和生活环境。通过归集住房数据、流动人口、常住人口数据综合分析，赋能教育、便民、医疗、交通和金融等数据要素场景

核心需求：释放数字政府和智慧城市的数据要素价值，促进数据安全流通

- ✓ 教育服务：精准预测学校未来的学生入学，提前规划新校区和师资
- ✓ 医疗服务：保护个人隐私下，医生使用患者的多维数据进行疾病诊断
- ✓ 交通服务：保护个人隐私和驾驶动态行为，精准预测个人保险投放
- ✓ 金融服务：精准营销场景下，建立个人隐私精准画像，防范隐私泄露

公安：

公安大数据开放创新、警务创新，协同办案

应用场景：流动人口信息结合就业登记情况分析流动人口就业状况，对就业困难的流动人口精准帮扶和管控，促进社会和谐

核心需求：释放公安人口等数据全要素，赋能公安内部使用和外部开放

- ✓ 警务创新：支持各种警种在追查办案、交通指挥、网络犯罪等领域，保护个人隐私情况下，融合更多的维度数据，增加服务价值
- ✓ 协同办案：支持跨部门，跨单位的数据协同和开放场景，精准的预测案件趋势和研判分析，精确打击犯罪行为
- ✓ 政务赋能：延展公安数据边界，保障原始业务数据存放安全岛，防范数据多次拷贝泄露风险

大数据交易中心：

探索数据要素市场化、数据变现等

应用场景：促进政企、企业与企业之间的数据交易，提升数据价值

核心需求：

- ✓ 数据定价：灵活的通过预定价、固定定价、实时定价、协议定价和拍卖定价，解决不同场景下的数据要素定价
- ✓ 数据交易：对数据加工、评估、分析、建模等各类增值服务的供需双方进行线下撮合、线上交易
- ✓ 定制交易：需要融合多维度数据或者个性化数据需求时，可由多家提供商线下组合协商，定制完成后各提供方按约定分成

数据服务公司：

数据流通主体单位间，提升数据服务

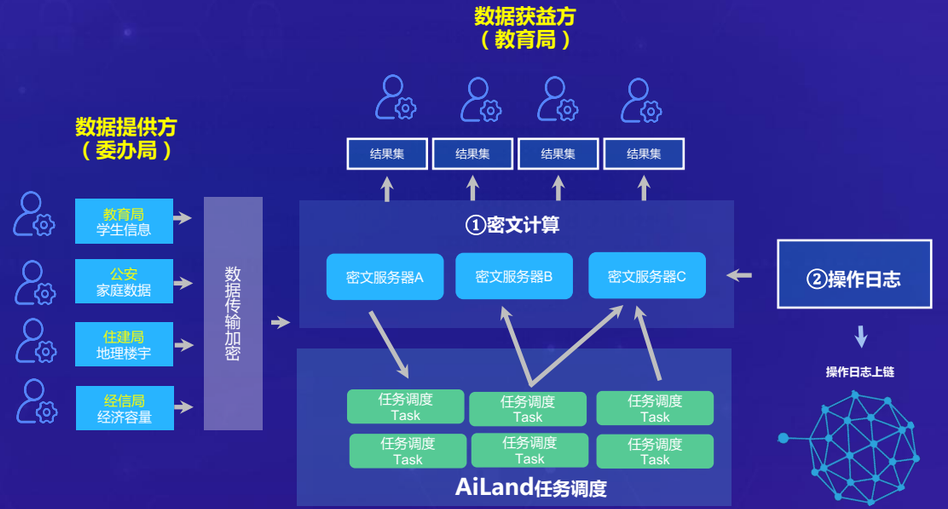
应用场景：金融服务、保险服务、广告推荐服务等公司

核心需求：

- ✓ 金融服务：通过多维度的个人、征信、消费等数据，跨越政府、企业间的数据融合后，为消费者提供深度金融服务
- ✓ 保险服务：通过融合政府信用、车联网驾驶数据，企业多主体间车辆维保数据碰撞，为个人提供精准的保险定价
- ✓ 广告服务： 通过取得大数据交易中心广告主体商、融合本企业个人数据，为消费者提供精确和个性化的服务

政务场景：在保护学生隐私数据前提下助力规划局学校选址

- ✓ **业务场景**：教育局需要对整个城市的学校进行精确选址规划，需要对近10年内辖区内学生数据、对应家庭户数，周边楼宇数量和分布，该区域经济体量联合建模得出学校建设规模和占地数量，防范涉及儿童数据隐私泄露。
- ✓ **安全风险**：教育局因为缺乏安全可信环境，导致数据提供方不愿意提供更多的数据给教育局，从而选址规划时常不够精准，造成社会经济损失。另外教育局本身无法查看各个委办局的明文业务数据。
- ✓ **解决方案**：通过安全岛自动化注册密文计算服务，将各个委办局提供的数据随机生成中间数，通过密文算子将各个委办局本地数据融合计算，最后将汇总计算数据给到数据获益方教育局，从而避免各方业务数据泄露。



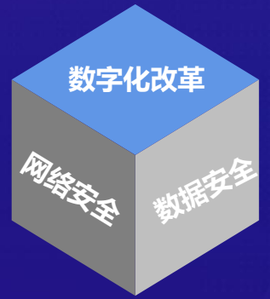
1.学生隐私数据密文计算，杜绝隐私数据泄露；
2.教育数据、公安数据、住建数据不出本地，完成融合学习，避免各方业务数据泄露。

赋能数据开放、激活数据价值

数据安全监管

数据安全监管

- ✓ 形成大数据+AI为驱动的技术保障能力
- ✓ 形成全天候全方位的监管监测能力
- ✓ 构建以数据为核心的安全防护能力
- ✓ 建设一体化的安全运营支撑能力



安全：赋能数据开放、激活数据价值

2021 WEST LAKE CYBERSECURITY CONFERENCE
西湖论剑·网络安全大会

2021

CYBERSECURITY :
THE FOUNDATION OF DIGITAL REFORM

谢谢您的观看指导!

西湖论剑·网络安全大会

扫码了解更多西湖论剑资讯