

人工智能数据安全风险与治理



胡绍勇 上海观安信息技术股份有限公司

2019年8月30日



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security



86 (21) 62090100



演讲人：胡绍勇



<https://www.idss-cn.com/>



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security





01

人工智能技术与数据安全风险

ARTIFICIAL INTELLIGENCE TECHNOLOGY DEVELOPMENT AND DATA SECURITY RISKS

02

国内外应对与举措

DOMESTIC AND FOREIGN RESPONSES AND INITIATIVES

03

人工智能数据安全治理

ARTIFICIAL INTELLIGENCE DATA SECURITY GOVERNANCE

04

国内外优秀实践案例

DOMESTIC AND FOREIGN EXCELLENT PRACTICE CASES

人工智能技术发展



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security



第一阶段：人工智能起步期

1956年达特茅斯会议标志着人工智能的诞生；1957年弗兰克·罗森布拉特提出了感知器神经网络模型；1970年受限于计算力，进入寒冬。

第二阶段：专家系统推广

1980年XCON专家系统出现，每年节约4000万美金；1990~1991年人工智能计算机DARPA没有实现，政府缩减投入，进入二次低谷。1997年IBM的深蓝战胜国际象棋冠军。

第三阶段：深度学习

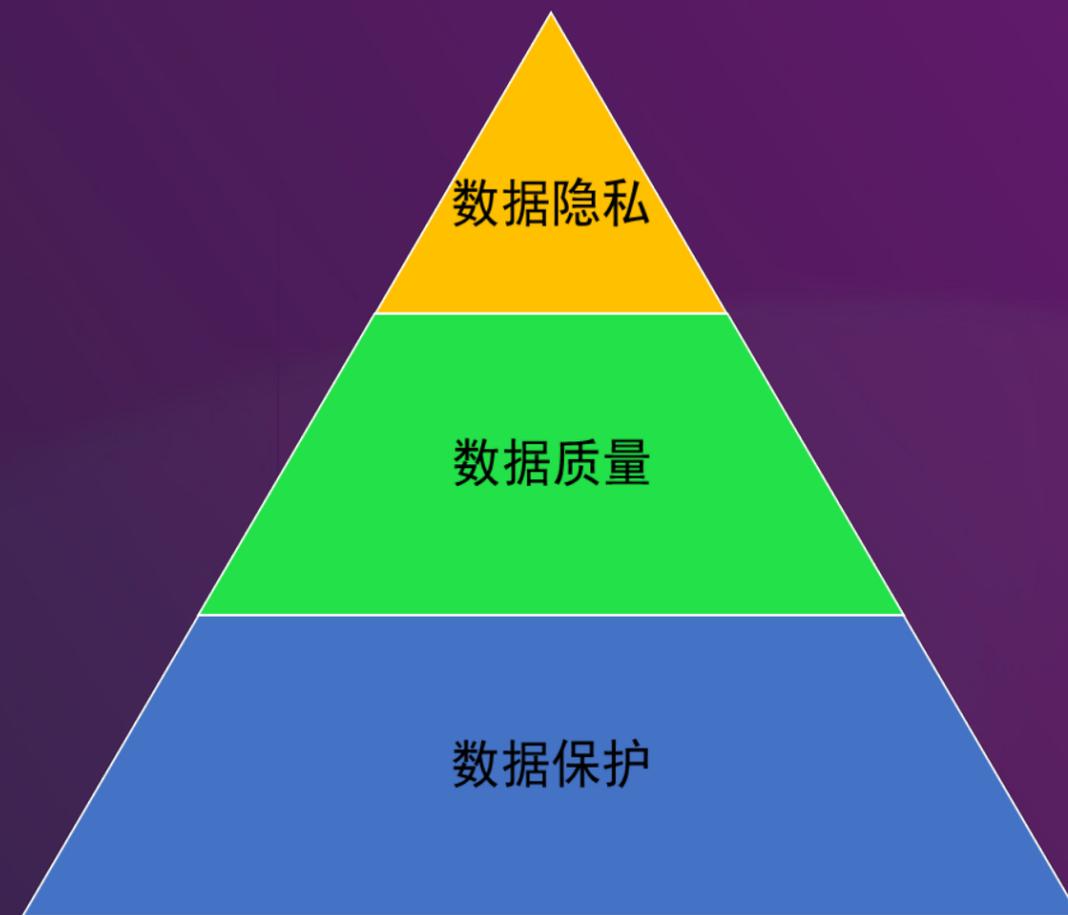
2006年Hinton提出“深度学习”的神经网络；2013年深度学习在语音和视觉识别上有重大突破，识别率超过99%和95%；2016年AlphaGo运用深度学习算法战胜围棋冠军



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security



人工智能数据安全风险



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

人工智能与数据隐私风险



训练、测试数据采集与隐私



现场数据采集与隐私



现场数据用于产品改进



数据分析挖掘与隐私



逆向攻击与隐私



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

人工智能与数据质量



数据集的规模不足



数据集的多样性和
均衡性不足



数据集的标注质量低



数据集遭到投毒攻击

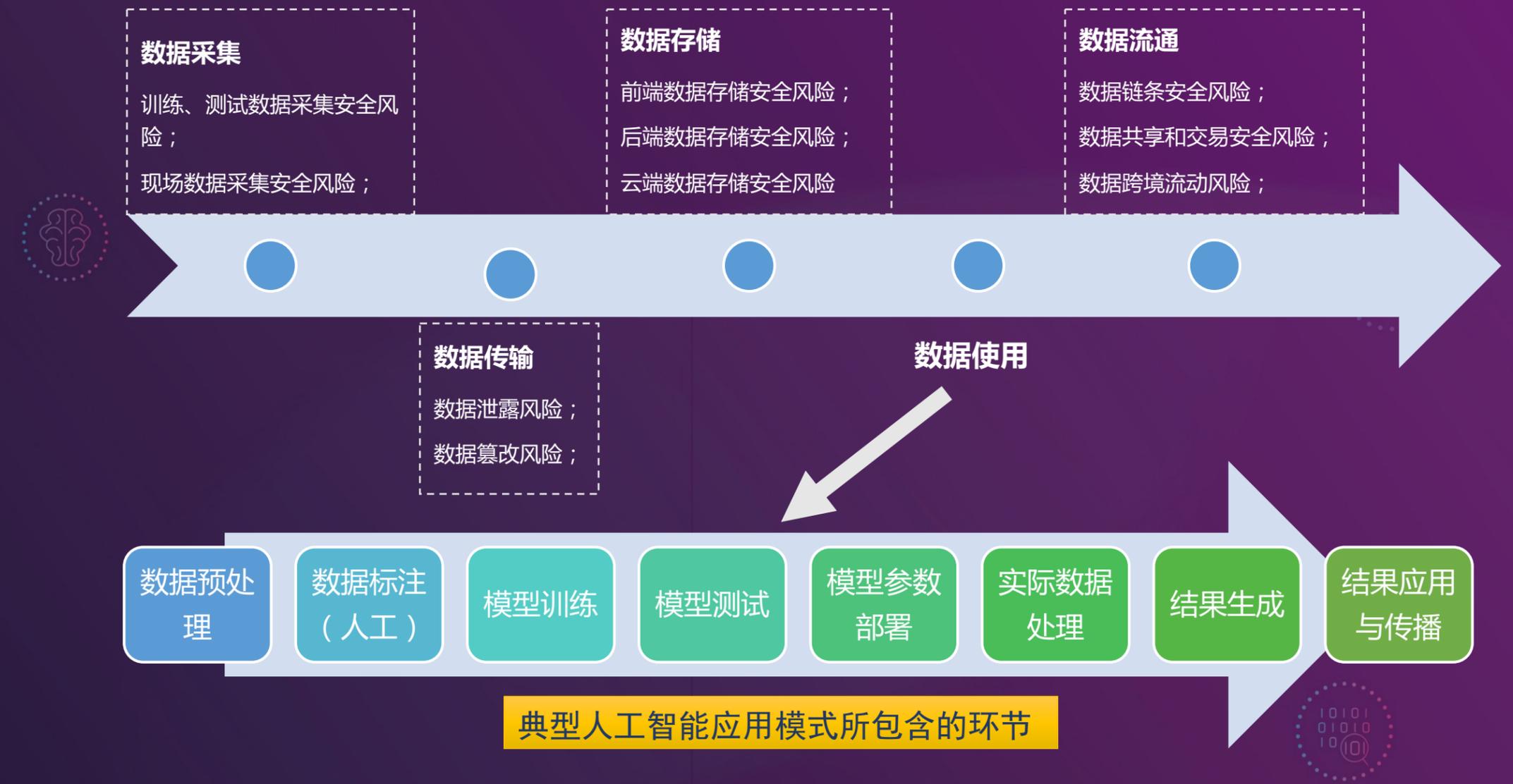


2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security



人工智能与数据保护



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security



01

人工智能技术与数据安全风险

ARTIFICIAL INTELLIGENCE TECHNOLOGY DEVELOPMENT AND DATA SECURITY RISKS

02

国内外应对与举措

DOMESTIC AND FOREIGN RESPONSES AND INITIATIVES

03

人工智能数据安全治理

ARTIFICIAL INTELLIGENCE DATA SECURITY GOVERNANCE

04

国内外优秀实践案例

DOMESTIC AND FOREIGN EXCELLENT PRACTICE CASES

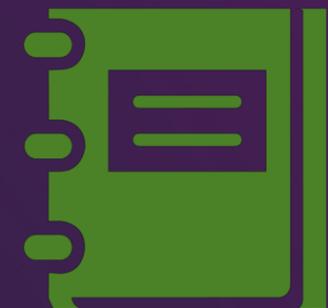


2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security



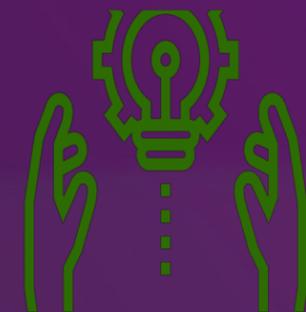
2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

国内外应对举措



政策法规

- 倡议层面
- 法规层面
- 标准指南



技术发展

- 保护隐私的机器学习
- 数据偏见检测技术
- 数据生成技术
- 减少数据需求技术
- 针对数据投毒的防御



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

国内外应对举措存在的差距



法规层面

当前法律存在盲点；
当前法律存在不适用性；

企业意识

数据安全意识薄弱；
缺乏合规驱动；

标准层面

缺乏人工智能数据安全通用标准；
缺乏人工智能细分应用领域的
数据安全标准；

技术层面

各种应对技术仍处于不成熟阶段；
数据安全风险评估能力弱；
防御技术不能抵御所有攻击类型；



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security



01

人工智能技术与数据安全风险

ARTIFICIAL INTELLIGENCE TECHNOLOGY DEVELOPMENT AND DATA SECURITY RISKS

02

国内外应对与举措

DOMESTIC AND FOREIGN RESPONSES AND INITIATIVES

03

人工智能数据安全治理

ARTIFICIAL INTELLIGENCE DATA SECURITY GOVERNANCE

04

国内外优秀实践案例

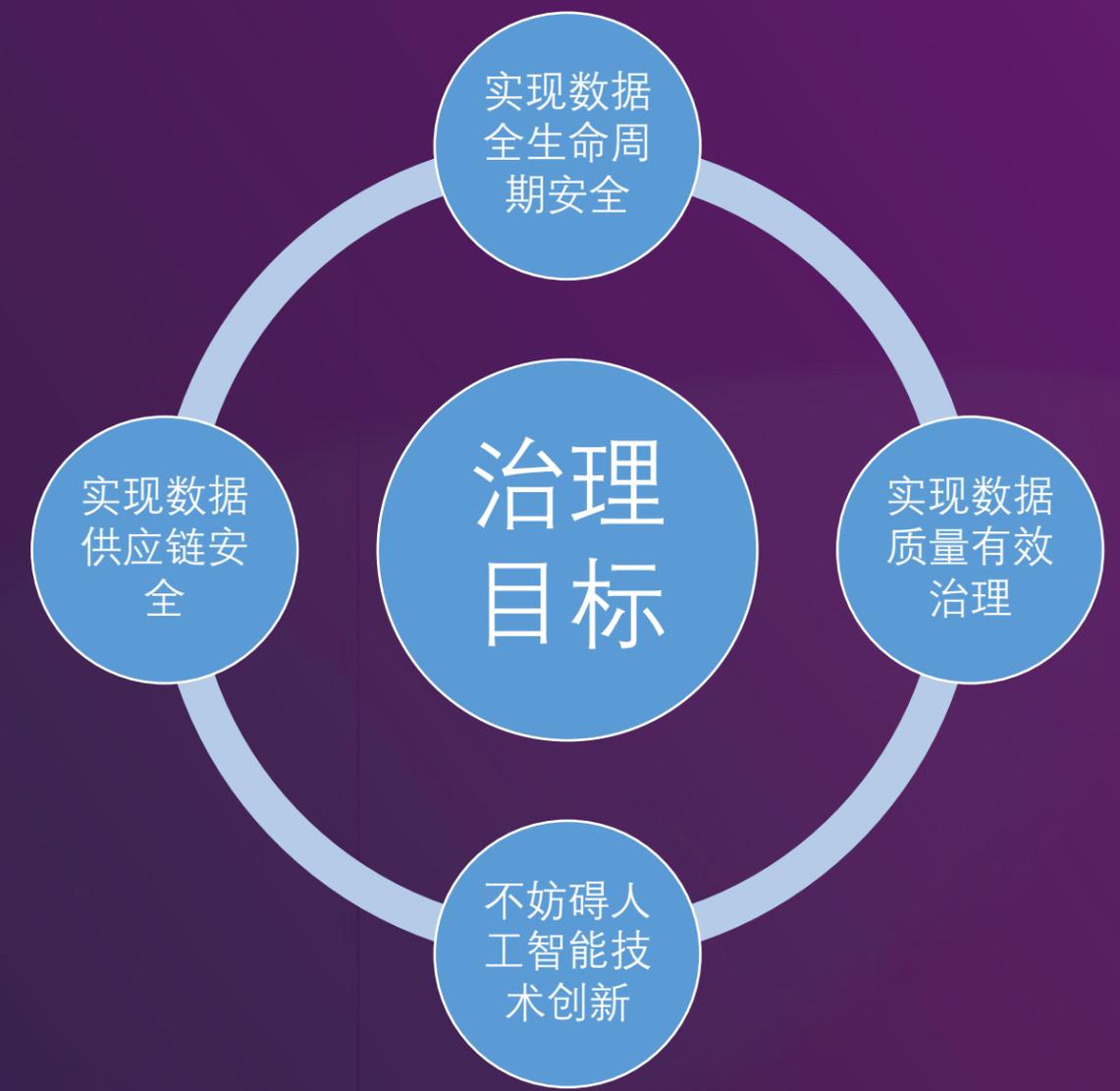
DOMESTIC AND FOREIGN EXCELLENT PRACTICE CASES

目录

CONTENT



治理目标



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

治理框架及措施

政策制定者

- 更新和完善现有法规，加快立法进程
- 促进最佳实践、指南、标准的形成
- 促进数据开放和建立公共数据集
- 加大对相关技术研发的支持
- 促进测评能力建设

监管

AI开发企业 & AI应用企业



关键技术支撑

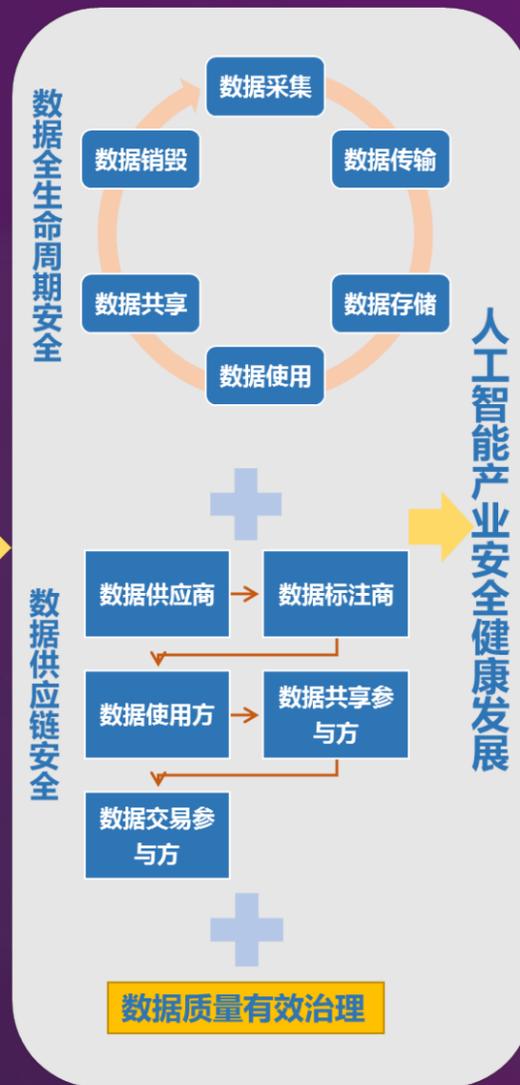
AI技术

- 基于隐私的机器学习技术
- 减少数据需求的技术
- 数据偏见检测技术
- 针对AI数据的攻击防御技术

数据安全技术

- 加密技术、访问控制
- 个人信息去标识化、数据脱敏
- 数据标签、数据交换共享管控
- 数据安全风险评估技术和能力

治理目标



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

数据全生命周期安全建设



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

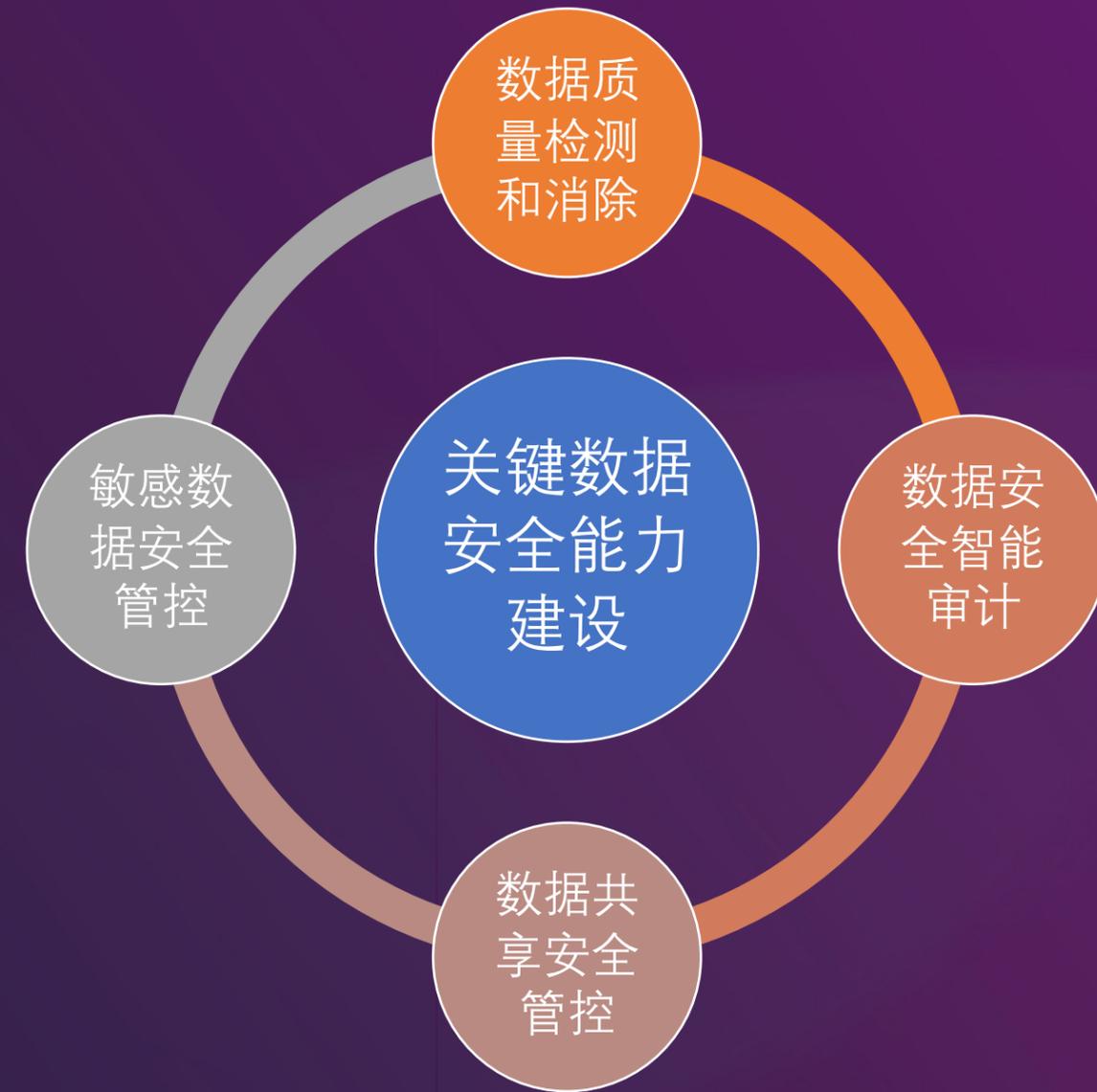


2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security





关键数据安全能力建设



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security





01

人工智能技术与数据安全风险
ARTIFICIAL INTELLIGENCE TECHNOLOGY DEVELOPMENT AND DATA SECURITY RISKS

02

国内外应对与举措
DOMESTIC AND FOREIGN RESPONSES AND INITIATIVES

03

人工智能数据安全治理
ARTIFICIAL INTELLIGENCE DATA SECURITY GOVERNANCE

04

国内外优秀实践案例
DOMESTIC AND FOREIGN EXCELLENT PRACTICE CASES

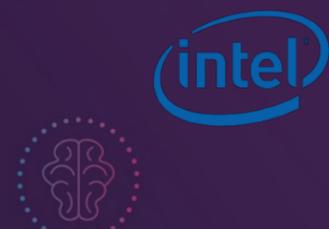


2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

国内外优秀案例



英特尔发布开源版HE-Transformer



TensorFlow Federated learning



TensorFlow的新模块 TensorFlow Privacy



差分隐私技术，保护用户共享给Apple的信息



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security



开源工具包AI Fairness 360



结构化域随机化系统



机器流量防控系统

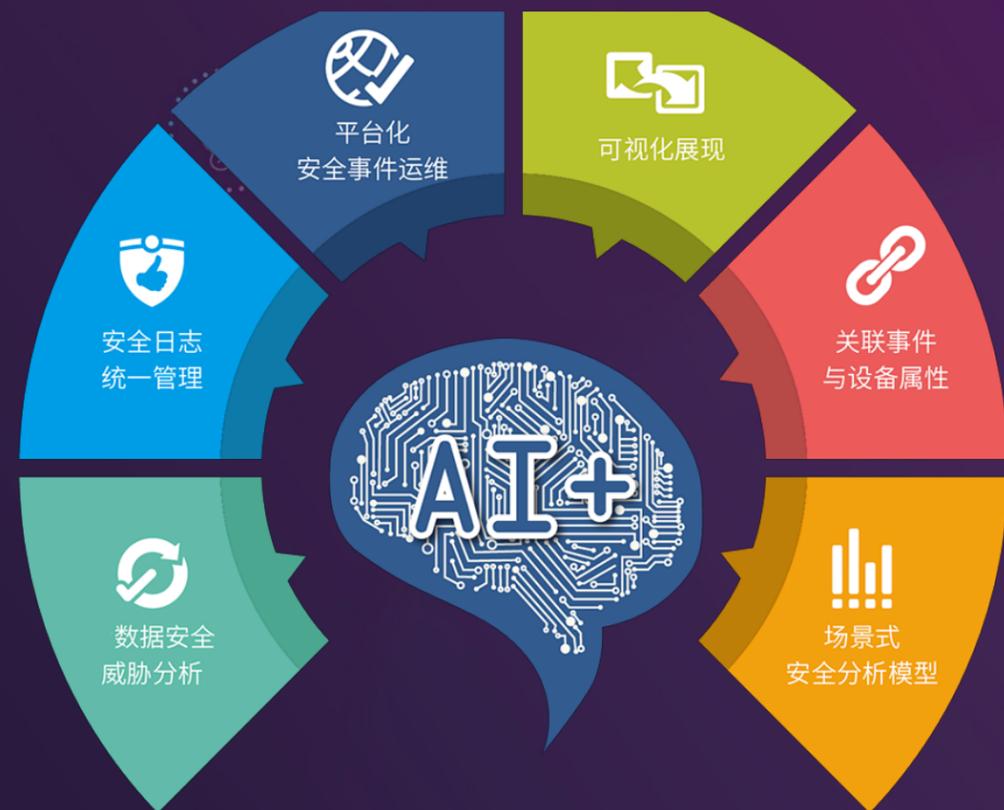


联盟AI系统并开源联盟AI解决方案FATE

2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

国内外优秀案例

观安信息数据安全解决方案助力智能客服系统敏感数据防护。



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security



谢谢聆听!

Thank you for listening



86 (21) 62090100



汇报人: 胡绍勇



<http://www.idss-cn.com>



2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

2019世界人工智能安全高端对话
2019 High-level Dialogue on World AI Security

