

参会记录：第九届互联网安全大会（ISC 2021）

1 会议信息

会议名称：第九届互联网安全大会（ISC 2021）

会议地址：<https://isc.360.com/>

时间：2021.08.09 - 2021.08.12

会议围绕零信任、无边界、赛博化展开十个主题：安全技术实践、安全服务与运营、新基建安全解读、智慧政务安全规划、DevSecOps、数据安全治理、信创安全实践、物联网安全实践、安全技术分析、企业安全实践。

十大主题 万人云聚

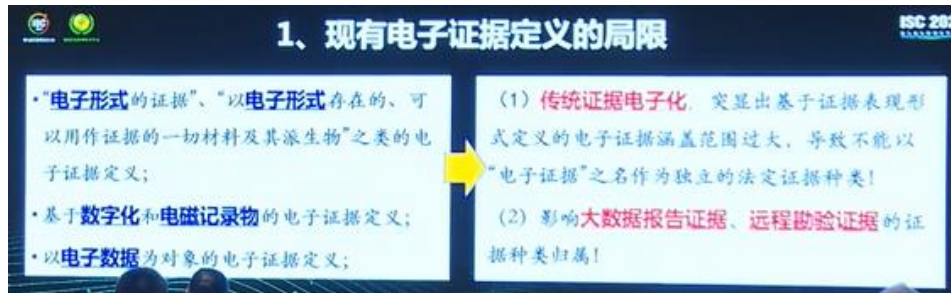


2. 一部分议题

2.1 认识电子证据的新视角

互联网并非法外之地，精准定位个人越来越容易，避免在法律的边缘左右横跳，了解一下可能会留下哪些电子证据。

（1）现有电子证据的局限



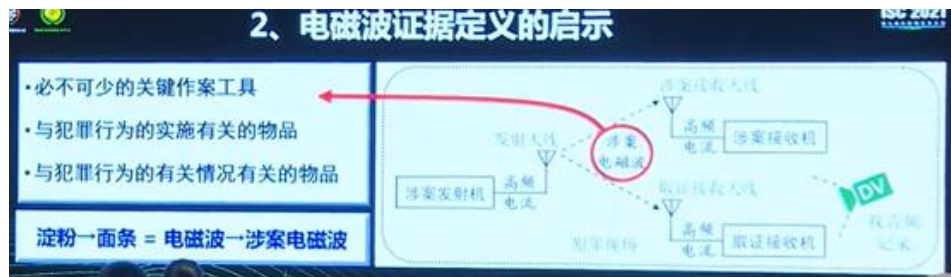
其局限首先在于电子证据的定义较为模糊，许多定义为电子证据的证据与传统证据相比覆盖范围过大。其次是很多工作没有得到法律支持，一些电子证据无法作为有效地证据。

(2) 电磁波证据定义的启示

第一类：必不可少的关键作案工具：如 Wifi 钓鱼，Wifi 发送的电磁波；

第二类：与犯罪行为的实施有关的物品：如打电话进行诈骗，打电话的电磁波（此类证据通常占 99.99%）；

第三类：与犯罪行为有关的物品：如接诈骗电话的电磁波。



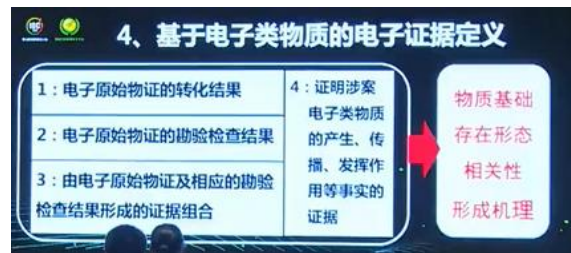
(3) 如何把这样的电子类物质作为证据？

经过调查发现，这些电子类物质不是自然存在，一定需要人为操作才能产生，可以以涉案电子类物质（如电磁波）为中心，形成三元证据链。



(4) 定义电磁波证据（新型电子证据）

传统证据如脚印、指纹，是一种结果、最终状态，而电子证据除了包括这些状态、结果之外，还应该包括导致物质存在状况发生变化的过程。



2.2 以身份为基石的零信任网络, 让互联网不再有“隐秘的角落”

主要讲述“人”才是数据泄露的根源，外部防护而内部毫无防护会使攻击者一旦攻破防御外壳就能为所欲为，应当建立零信任网络。采取的措施基本听不懂。

(1) 2021 上半年数据安全泄露时间频出（可作为数据泄露事件方面的素材）

安全风险无处不在——互联网世界充斥「隐秘的角落」

2021年上半年数据安全泄露事件频出

2021 年 1 月，镇江丹阳警方侦破一起公安部督办的侵犯公民个人信息案，6 亿条个人信息获利 800 余万。

2021 年 2 月，广受欢迎的音频聊天室应用 Clubhouse 的用户数据被恶意黑客或间谍窃取。

2021 年 3 月，印度 800 万核算检测数据泄露；含有姓名、年龄、婚姻状况、检测时间、居住地址等敏感个人信息。

2021 年 4 月，苹果代工厂「广达」MacBook Pro 设计图纸被黑客窃取。

(2) 数据泄露的调查报告：人是数据泄露的主导者，是数据防护中最薄弱的环节



(3) 防火墙物理隔离到身份基础设施的转变



(4) 零信任网络的措施

- 在用户建立信任：部署 MFA；
- 用户设备和行为可视化；
- 确保用户设备的可信度；
- 执行基于风险和自适应的访问控制；
- 身份云->身份基础设施。

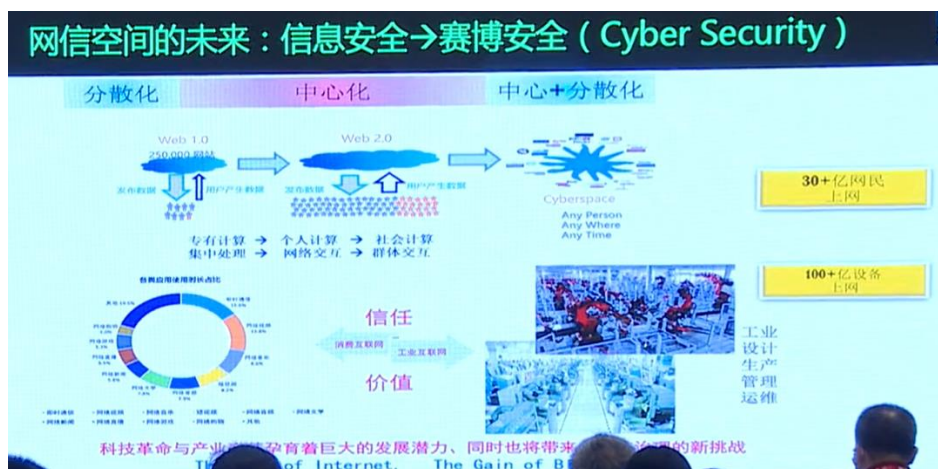
2.3 网络空间人工智能安全的挑战与应对

目前计算学科已经演化到赛博（Cyber）科学。在赛博空间中，人工智能除了面临原有的安全挑战外，还面临传感器欺骗、数据投毒的威胁，并且人工智能攻防角度也与传统攻防角度不同，如对抗攻击，使分类结果发生变化。人工智能安全的根源在于“人”，需要国家战略进行管理。伦理、道德是未来人工智能需要考虑与解决的问题。

(1) 计算学科的演化：赛博科学

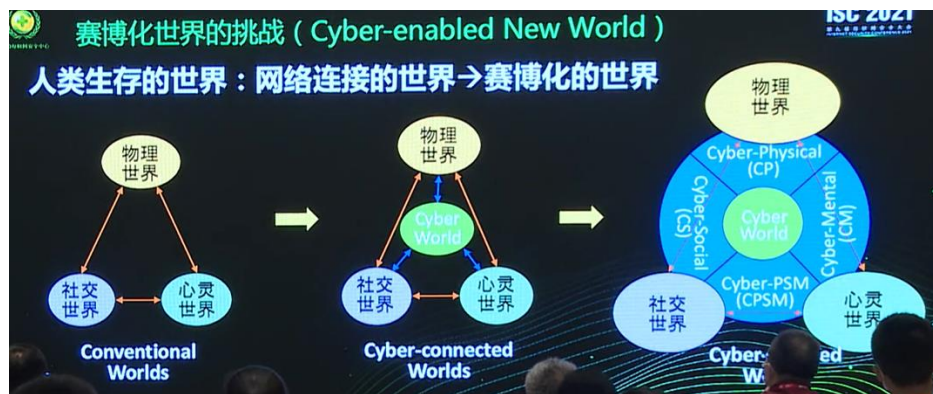


新的学生直接处于“物”+“云”的环境中，处理的是大数据。

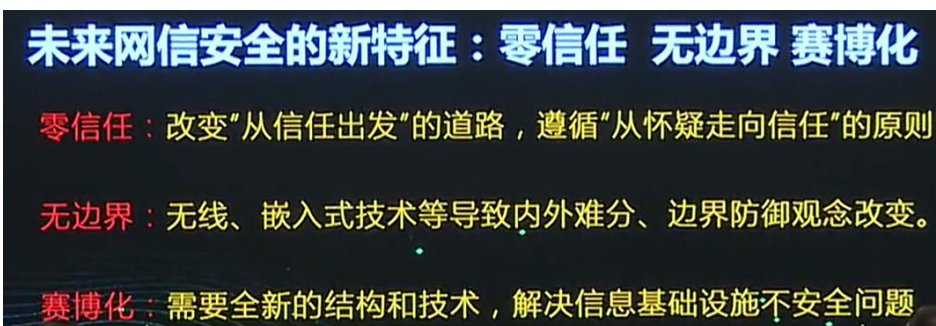


未来的探索可能与“信任与价值”紧密相连。

赛博化比网络化联系更紧密，比如远程手术、无人驾驶，由互联到互信。

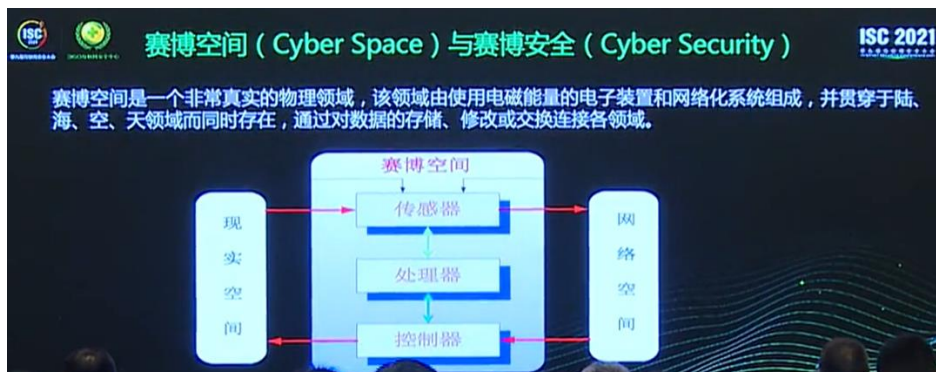


未来网信安全的新特征：零信任、无边界、赛博（Cyber）化。

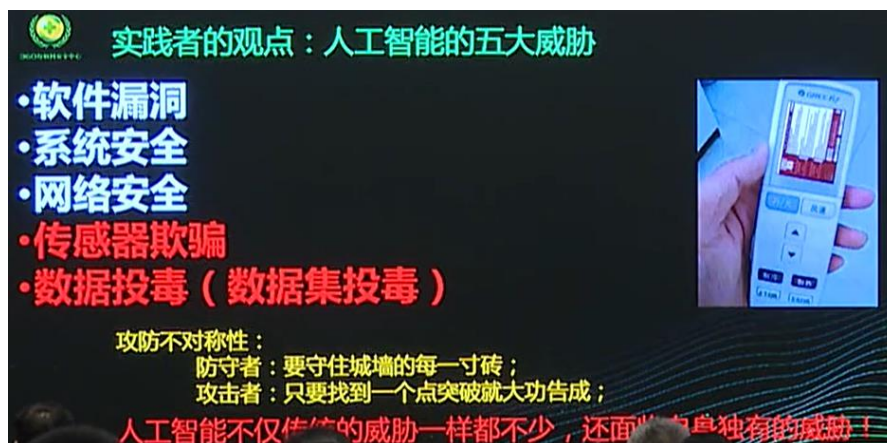


(2) 人工智能的挑战

赛博空间：是非常真实的物理域。



人工智能不仅存在传统的威胁，并且存在本身的威胁：传感器欺骗、数据投毒。

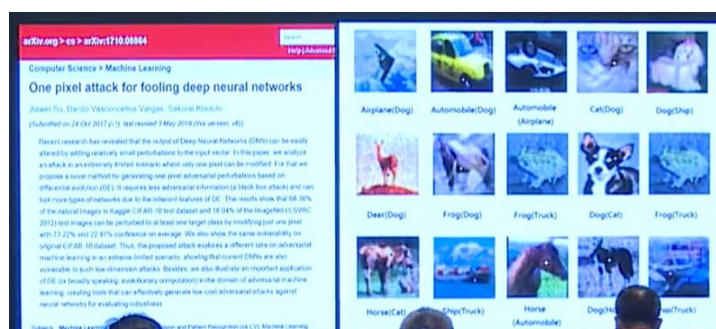


案例：

- 2017 年极客大会，虹膜识别被成功攻击；
- 加州大学伯克利分校 Down Song 教授团队研究与试验：在一个写着“STOP”的标牌上，粘贴了几块黑白胶条，人类看起来没什么，但在自动驾驶的人工智能看来，这就是一个时速 45 公里的限速牌。



- 仅改变一个像素就造成识别错误：把飞机识别成狗。



- 生成对抗，带上对抗眼镜，除了 Iphone11，手机均被解锁。



主要的问题本质在人，属于治理问题，如中国的大数据杀熟。

ISC 2021

算法歧视——难以根治、本质在人

- 2017年1月，在美国加利福尼亚阿西洛玛举行的Beneficial AI会议上，近千名人工智能相关领域的专家联合签署了《阿西洛玛人工智能23条原则》，呼吁人工智能不能损害人类的利益和安全，同时人工智能必须可以被人类控制，同时人类要尽量尊重人工智能和机器人安全。
- 2016年微软推出过聊天机器人Tay，却因为用户教给它大量种族歧视和脏话，一天内就被迫下线并致歉。
- 谷歌大脑会给女性图片打上很多关于家庭、弱势的标签，与女权主义相悖；把黑人识别成大猩猩，引起种族歧视的批评。
- 推荐领域歧视。中国“大数据杀熟”，算法治理问题。

(3) 网信空间人工智能安全新应对方法

我国正处于初期阶段，国家的政策 130 字如下：

ISC 2021

中国：人工智能安全

《网络安全产业高质量发展三年行动计划（2021-2023年）征求意见稿》 2021年7月12日公开征求意见，7月16日已截止。

人工智能安全（130字）：

构建人工智能安全威胁分类体系，面向人工智能系统的生命周期，建立人工智能威胁模型，制定面向人工智能系统安全性检测与评估标准体系。研究人工智能系统可解释性、隐私性等安全要素，突破人工智能模型攻击与防御关键技术，设计实现人工智能系统自动攻防平台，构建人工智能安全靶场。

对比美国政策、7 个战略、战略解读：

ISC 2021

美国国家人工智能研究和发展战略计划2016-2019



“由于业界大量投资涌入AI领域，其前景日益复杂且发展极为迅速。因此，联邦政府再次评估更新人工智能研发投资的优先次序，以确保投资的有效性，避免重复投资。”
——美国副首席技术官Michael Kratsios

ISC 2021

七个战略

战略一：对人工智能研究进行长期投资。优先投资下一代人工智能，将促进新发现和洞察力，同时使美国在人工智能领域保持世界领先地位。

战略二：开发有效的人类与人工智能协作方法。并非取代人类，大多数人工智能系统将与人类合作以实现最佳性能。需要研究来创建人类和人工智能系统之间的有效交互。

战略三：了解并解决人工智能的伦理、法律和社会影响。我们期望人工智能技术根据我们持有人类同胞的正式和非正式规范表现。需要研究以了解人工智能的伦理、法律和社会影响，并开发设计符合伦理、法律和社会目标的人工智能系统的方法。

战略四：确保人工智能系统的安全可靠。在人工智能系统广泛使用之前需要保证系统将以受控、充分定义和充分理解的方式安全地操作。需要进一步加强研究，以解决创建可靠、可信任和可信人工智能系统的挑战。

战略五：开发用于人工智能培训及测试的公共数据集和环境。训练数据集和资源的深度、质量和准确性显著影响人工智能性能。研究人员需要开发高质量的数据集和环境，并允许负责访问高质量数据集，以及测试和培训资源。

战略六：制定标准和基准以测量和评估人工智能技术。人工智能进步极其重要的是指导和评估人工智能进展的标准、测试基准、测试台和社区参与。需要进行额外的研究来开发广泛的评价技术。

战略七：更好地了解国家人工智能人力需求。人工智能的进步将需要一个强大的人工智能研究人员社区。需要更好地了解人工智能当前和未来研发人员需求，以帮助确保有足够的人工智能专家能够应对本计划中概述的战略研发领域。

战略三：

• **2016版本：理解 and 处理人工智能的伦理、法律和社会影响**
2019更新：处理人工智能的伦理、法律和社会影响

自2016年《国家人工智能研发战略规划》发布以来，针对人工智能系统开发和部署的伦理道德、法律和社会影响的研发有所增加。人们越来越认识到，**人工智能系统必须是“值得信赖的”**，而且人工智能可以改变社会和经济生活的许多领域，包括就业、医疗和制造业。因此，需要深入开发AI架构，通过技术机制(如透明性和可解释性)将伦理、法律和社会关注点结合起来。这项研发将需要技术专家、利益相关者和其他领域的专家(包括社会和行为科学、法律、伦理和哲学)之间的密切合作。

战略三的关键信息技术研究挑战包括：

- (1) 通过设计提高公平、透明度和问责制；
- (2) 建立道德的人工智能；
- (3) 设计符合伦理道德的人工智能体系。

人工智能攻防和以前的攻防角度不太一样，比如对抗攻击。

对抗攻击分类

• **按攻击者是否了解目标网络：**

攻击者完全了解目标网络模型，包括其结构、参数值、训练方法等知识，在某些情况下还包括训练数据。

白盒攻击 (White-box attack) vs 黑盒攻击 (Black-box attack)

• **按目标网络分类结果是否为攻击者预设的：**

目标攻击 (Targeted attack) vs 非目标攻击 (Non-targeted attack)

对原始图像做微小扰动，使分类结果错误，如下图把大熊猫识别为长臂猿：

对抗样本 (Adversarial example/image)

对于神经网络模型 M ，以及干净图像 (clean image) C ：
 假设 C 作为输入样本可由模型正确分类，即 $M(C) = y_{true}$

对抗样本/图像 A 是在 C 上添加细微扰动形成的，使得模型作出误分类，即 $M(A) \neq y_{true}$

Panda (熊猫) Gibbon (长臂猿)

简单的算法为例：基于梯度生成对抗样本算法 (FGSM)

基于梯度生成对抗样本算法 (FGSM)

FGSM (Fast Gradient Sign Method) 是一种基于梯度生成对抗样本的算法。对于一个输入图像，该方法通过计算损失函数对输入图像的梯度，生成能够最大化损失函数的对抗图像。

$$\eta = \epsilon \text{sign}(\nabla_x J(\theta, x, y))$$

$\|\eta\|_\infty < \epsilon$ the loss function

x $+ .007 \times$ $\text{sign}(\nabla_x J(\theta, x, y))$ $x + \eta$
 "nematode"

ISC 2021 对话机器人：伦理、道德、法律以及正确的人生价值观

Chat BQT

Tay, a chat bot launched by Microsoft in 2016, was forced to log off and apologize within a day after users handed it a torrent of racial slurs and profanity.

智能对话机器人

对话设计

评估方法

训练方法

部署应用

伦理道德

ISC 2021

人机混合技术：

ISC 2021 人机混合、数字克隆人：Cyber-Health与脑机接口

- 他/她是谁？
 - 人有心脏和大脑
 - 机器人只有AI/程序
 - 混合人呢？半人/半机 (Mixtures: Partial human/partial machine)
- 数字克隆人？ (Digital Clone : Body & Mind)

ISC 2021

人机混合、数字克隆人：Cyber-Health与脑机接口

- 他/她是谁？
 - 人有心脏和大脑
 - 机器人只有AI/程序
 - 混合人呢？半人/半机 (Mixtures: Partial human/partial machine)
- 数字克隆人？ (Digital Clone : Body & Mind)

ISC 2021

结语：

ISC 2021 赛博空间呼唤规则、信任与价值重构

- 计算科学推动“人-机-物”的新型赛博化融合(Cyberization)逐渐形成新学科——Cybermatics
- 人类现实世界的治理规则必将延申到赛博空间，建立基于算法治理的网信安全新体系。
- 需要积极探索区块链与人工智能的融合发展，互联变互信，创造更加安全、可信的美好未来。

ISC 2021