

信息安全环境与工具理论作业



2020-9-7

161820127

廖俊轩

# 大数据环境下的信息安全问题

## 发展现状：

一是已有众多成功的大数据应用，但就其效果和深度而言，当前大数据应用尚处于初级阶段，根据大数据分析预测未来、指导实践的深层次应用将成为发展重点。

二是大数据治理体系远未形成，特别是隐私保护、数据安全与数据共享利用效率之间尚存在明显矛盾，成为制约大数据发展的重要短板，各界已经意识到构建大数据治理体系的重要意义，相关的研究与实践将持续加强。

三是数据规模高速增长，现有技术体系难以满足大数据应用的需求，大数据理论与技术远未成熟，未来信息技术体系将需要颠覆式创新和变革。

大数据是信息技术发展的必然产物，更是信息化进程的新阶段，其发展推动了数字经济的形成与繁荣。信息化已经历了两次高速发展的浪潮，始于上世纪80年代，随个人计算机大规模普及应用所带来的以单机应用为主要特征的数字化（信息化1.0），及始于上世纪90年代中期，随互联网大规模商用进程所推动的以联网应用为主要特征的网络化（信息化2.0）。当前，我们正在进入以数据的深度挖掘和融合应用为主要特征的智能化阶段（信息化3.0）。在“人机物”三元融合的大背景下，以“万物均需互联、一切皆可编程”为目标，数字化、网络化和智能化呈融合发展新态势。

作为人口大国和制造大国，我国数据产生能力巨大，大数据资源极为丰富。随着数字中国建设的推进，各行业的数据资源采集、应用能力不断提升，将会导致更快更多的数据积累。到2020年，我国数据总量有望达到8000EB（1018），占全球数据总量的21%，将成为名列前茅的数据资源大国和全球数据中心。

## 大数据环境下存在的信息泄露问题

一是大数据治理体系尚待构建。首先，法律法规滞后。目前，我国尚无真正意义上的数据管理法规，只在少数相关法律条文中有涉及到数据管理、数据安全等规范的内容，难以满足快速增长的数据管理需求。其次，安全隐患增多。近年来，数据安全和隐私数据泄露事件频发，凸显大数据发展面临的严峻挑战。在大数据环境下，数据在采集、存储、跨境跨系统流转、利用、交易和销毁等环节的全生命周期过程中，所有权与管理权分离，真假难辨，多系统、多环节的信息隐性留存，导致数据跨境跨系统流转追踪难、控制难，数据确权和可信销毁也更加困难。同时也增大了信息泄露的风险。

二是核心技术薄弱。基础理论与核心技术的落后导致我国信息技术长期存在“空心化”和“低端化”问题，大数据时代需避免此问题在新一轮发展中再次出现。近年来，我国在大数据应用领域取得较大进展，但是基础理论、核心器件和算法、软件等层面依旧比较落后，可能因为技术落后而产生数据泄露的风险。

## 阻止手段

### 建立系统全面的大数据治理体系

大数据治理须从营造大数据产业发展环境的视角予以全面、系统化考虑。大数据治理体系建设涉及国家、行业和组织三个层次。在国家层次，重点是出台数据安全与隐私保护的法律法规，保障国家、组织和个人的数据安全。在行业层次，重点是要在国家相关法律法规框架下，充分考虑本行业中企业的共同利益与长效发展，建立规范行业数据管理的组织机构和数据管控制度，制定行业内数据共享与开放的规则和技术规范，促进行业内更加安全规范地使用数据。在组织层次，重点是要提升企业对数据全生命期的管理能力，促进企业内部和企业间的数据流通，提升数据变现能力，保障企业自身的数据安全及客户的数据安全和隐私信息。

但同时我们也不能因噎废食。要充分平衡数据共享开放和隐私保护、数据安全的关系。如，数据集中管理可能带来保管上的安全问题，然而数据融合才能产生价值，一定程度的集中是趋势所在，也更利于建立更强大可靠的保护机制；多源数据的融合可能导致信息泄露，然而在确知风险前，是否需要因其“可能性”而拒绝技术的应用？数据脱敏仍然可能存在隐私泄露的风险，是否允许个体在知情前提下“用隐私换方便”、“用隐私换治疗换健康”？是否允许使用符合当前“标准”、但无法确保未来一定不出现信息泄露的脱敏方法，并对相关应用予以免责？当然，加强兼顾隐私保护、数据安全和数据流动利用的新技术研发，也非常必要。当前，如安全多方计算、同态加密、联邦学习等技术研发，希望允许拥有数据的各方在不向其他组织或个人公开数据中所含敏感信息的情况下，实现数据的融合利用。虽然这些技术尚处于发展的初级阶段，但其应用前景广阔。

### 未雨绸缪，防范大数据发展可能带来的新风险

大数据发展可能导致一系列新的风险。例如，数据垄断可能导致数据“黑洞”现象。一些企业凭借先发展起来的行业优势，不断获取行业数据，但却“有收无放”，呈现出数据垄断的趋势。这种数据垄断不仅不利于个人数据的安全，而且有可能对国家安全带来冲击和影响。

### 大力发展大数据行业

面对数据规模高速增长，现有技术体系难以满足大数据应用的需求，大数据理论与技术远未成熟。未成熟同时也就意味着有很多尚不明确的漏洞。故要发展大数据，让这个行业变得更加成熟，进而减少因为漏洞带来的数据不安全。

# 人工智能环境下的信息安全问题

## 发展现状：

人工智能可划分为弱人工智能、强人工 智能和超人工智能三个阶段。弱人工智能擅长于在特定领域、有限规 则内模拟和延伸人的智能；强人工智能具有意识、自我和创新思维， 能够进行思考、计划、解决问题、抽象思维、理解复杂理念、快速学 习和从经验中学习等人类级别智能的工作；超人工智能是在所有领域 都大幅超越人类智能的机器智能。

通用人工智能尚处于起步阶段。目前，虽然专用人工智能领域已取得突破性进展，但是通用人工智能领域的研究与应用仍然任重而道远，人工智能总体发展水平仍处于起步阶段。当前的人工智能系统在信息感知、机器学习等“浅层智能”方面进步显著，但是在概念抽象和推理决策等“深层智能”方面的能力还很薄弱。总体上看，目前的人工智能系统可谓有智能没智慧、有智商没情商、会计算不会“算计”、有专才而无通才。

## 人工智能环境下存在的信息泄露问题

### 人工智能自身面临的数据安全风险

训练数据污染可导致人工智能决策错误。数据投毒通过在训练数据里加入伪装数据、恶意样本等破坏数据的完整性，进而导致训练的算法模型决策出现偏差。数据投毒主要有两种攻击方式：一种是采用模型偏斜方式，主要攻击目标是训练数据样本，通过污染训练数据达到改变分类器分类边界的目的。例如，模型偏斜污染训练数据可欺骗分类器将特定的恶意二进制文件标记为良性。另外一种是采用反馈误导方式，主要攻击目标是人工智能的学习模型本身，利用模型的用户反馈机制发起攻击，直接向模型“注入”伪装的数据或信息，误导人工智能做出错误判断。随着人工智能与实体经济深度融合，医疗、交通、金融等行业训练数据集建设需求迫切，这就为恶意、伪造数据的注入提供了机会，使得从训练样本环节发动网络攻击成为最直接有效的方法，潜在危害巨大。在自动驾驶领域，数据投毒可导致车辆违反交通规则甚至造成交通事故；在军事领域，通过信息伪装的方式可诱导自主性武器启动或攻击，从而带来毁灭性风险。运行阶段的数据异常可导致智能系统运行错误。一是人为构造对抗样本攻击，导致智能系统产生错误的决策结果。人工智能算法模型主要反映了数据关联性和特征统计，而没有真正获取数据因果关系。针对算法模型这一缺陷，对抗样本通过对数据输入样例添加难以察觉的扰动，使算法模型以高置信度给出一个错误的输出。对抗样本攻击可实现逃避检测，例如在生物特征识别应用场景中，对抗样本攻击可欺骗基于人工智能技术的身份鉴别、活体检测系统。

模型窃取攻击可对算法模型的数据进行逆向还原。人工智能算法模型的训练过程依托训练数据，并且在运行过程中会进一步采集数据进行模型优化，相关数据可能涉及到隐私或敏感信息，所以算法模型的机密性非常重要。但是，算法模型在部署应用中需要将公共访问接口发布给用户使用，攻击者可通过公共访问接口对算法模型进行黑盒访问，依据输入信息和输出信息映射关系，在没有算法模型任何先验知识（训练数据、模型参数等）情况下，构造出与目标模型相似度非常高的模型，实现对算法模型的窃取，进而还原出模型训练和运行过程中的数据以及相关隐私信息。

开源学习框架存在安全风险，可导致人工智能系统数据泄露。人工智能开源学习框架实现了基础算法的模块化封装，可以让应用开发人员无需关注底层实现细节，大大提高了人工智能应用的开发效率。谷歌、微软、亚马逊、脸书等企业都发布了自己的人工智能学习框架，在全球得到广泛应用。但是，人工智能开源学习框架集成了大量的第三方软件包和依赖库资源，相关组件缺乏严格的测试管理和安全认证，存在未知安全漏洞。近年来，360、腾讯等企业安全团队曾多次发现TensorFlow、Caffe、Torch 等深度学习框架及其依赖库的安全漏洞，攻击者可利用相关漏洞篡改或窃取人工智能系统数据。

### 人工智能应用导致的数据安全风险

人工智能应用可导致个人数据过度采集，加剧隐私泄露风险。随着各类智能设备（如智能手环、智能音箱）和智能系统（如生物特征识别系统、智能医疗系统）的应用普及，人工智能设备和系统对个人信息采集更加直接与全面。相较于互联网对用户上网习惯、消费记录等信息采集，人工智能应用可采集用户人脸、指纹、声纹、虹膜、心跳、基因等具有强个人属性的生物特征信息。这些信息具有唯一性和不变性，一旦被泄露或者滥用会对公民权益将造成严重影响。2018年 8 月，腾讯安全团队发现亚马逊智能音箱后门，可实现远程窃听并录音。2019 年 2 月，我国人脸识别公司深网视界曝出数据泄露事件，超过 250 万人数据、680 万条记录被泄露，其中包括身份证信息、人脸识别图像及 GPS 位置记录等。

## 阻止手段

### 明晰发展与安全并举的治理思路

一是推进人工智能数据资源建设，在发展中解决安全问题。建立健全适合我国国情的数据流通共享机制，推动政府和行业数据开放，培育规范数据交易市场，鼓励不同市场主体安全的进行数据交换，构建支撑我国人工智能产业发展的优质数据资源，在发展中规避数据偏见、数据权属等人工智能数据安全问题。二是加强人工智能数据安全治理能力，以安全促进发展。基于人工智能数据安全风险研究，依托现有数据安全管理机制和技术手段，加大人工智能应用场景下数据安全防护技术研究，同时，促进人工智能技术在数据安全治理与网络攻防对抗等领域中的应用，实现人工智能数据安全风险的提前感知和预防，规避训练数据污染、数据智能窃取等数据安全风险，促进人工智能安全发展。

### 建立人工智能数据安全法律法规

一是推进人工智能和数据安全相关立法工作。在国家层面，推进《数据安全法》、《个人信息保护法》以及人工智能相关法律出台，明确人工智能数据安全法律原则，确立不同参与主体在人工智能生命周期各阶段所享有的数据权利与承担的安全责任，设立人工智能数据安全问责制和救济制度，并对人工智能相关数据过度采集、偏见歧视、资源滥用、深度伪造等突出问题进行规制，为人工智能数据安全管理提供基本法律依据。二是完善人工智能数据安全相关部门规章。依据国家相关法律，结合人工智能在不同领域应用中的特点，针对各领域关键突出人工智能数据安全风险，制定和细化相关部门规章，提出对所属领域的人工智能算法设计、产品开发和成果应用等过程中数据安全要求。三是开展人工智能数据安全执法。加强对人工智能数据收集、使用、共享等高风险环节安全执法，特别是对数据过度采集、数据资源滥用、侵犯个人隐私、违背道德伦理等行为加大执法惩戒力度，创新和规范人工智能数据安全事件调查取证方法和程序，促进人工智能数据安全法律和规章有效落地执行。积累执法经验并总结不足，形成反馈机制持续完善相关法律和部门规章。

### 完善人工智能数据安全监管措施

一是开展人工智能数据安全监督惩戒。依照国家法律法规，政府部门针对数据过度采集、数据偏见歧视、数据资源滥用等人工智能数据安全风险，通过线上线下多种方式实施监督检查，及时发现和防范安全隐患。针对基于人工智能的网络攻击、深度伪造等严重不良行为，利用技术手段监测和社会公众监督等方式，及早发现，降低危害，加强惩戒。二是开展人工智能数据安全检测评估。依托行业组织或者第三方机构，构建人工智能数据安全检测评估平台，制定人工智能产品、应用和服务的数据安全检测评估方法和指标体系，研发安全检测评估工具集，通过测试验证提升人工智能产品安全性和成熟度，降低人工智能数据安全风险。通过检测评估强化企业的数据安全与隐私保护，为人工智能研发和广泛应用提供海量数据支撑。

### 健全人工智能数据安全标准体系

一是完善我国人工智能数据安全标准体系，加快急需重点标准研制。在我国人工智能安全标准框架下，加快研制人工智能数据安全标准体系，制定人工智能数据安全标准推进计划。重点加快推进人工智能数据安全评估、人工智能平台数据安全保护、自动驾驶用户隐私保护等行业急需重点标准研制工作。二是优化我国人工智能数据安全标准化组织建设。推动国家信息安全标准化技术委员会、中国通信标准化协会等国家及行业标准化组织成立人工智能安全研究组，促进国家、行业和团体标准化组织联合有序推进人工智能数据安全标准出台。三是加强国际人工智能数据安全标准化工作。组织国内企业、科研院所等多方力量加强研究储备，在IEEE、ISO/IEC、ITU等国际标准化组织中联合发声,提出更多人工智能数据安全相关提案，贡献更多中国力量和方案，实质性参与和主导人工智能数据安全相关国际标准工作。

### 创新人工智能数据安全技术手段

一是加强人工智能数据安全保护基础理论研究和技术研发。利用国家专项和社会基金引导产学研各界联合开展人工智能数据安全风险产生机理和防御理论的研究，并突破小样本学习、联邦学习、差分隐私等人工智能数据安全保护核心关键技术。二是建设完善我国人工智能开源学习框架，提供保障数据安全的人工智能基础研发平台。鼓励企业建设完善人工智能开源学习框架，增强框架内置数据安全设计和技术措施。并且通过我国市场优势，加快培育自有人工智能开源平台共享应用生态圈和产业链。三是促进人工智能在数据安全领域中的应用。鼓励人工智能企业和数据安全企业充分发挥各自优势，通过成立联合实验室、共同投资等多种方式，开展人工智能技术在数据安全治理领域的应用研究和产品技术研发。

### 培养复合人工智能数据安全人才

一是完善学校人工智能数据安全教育。鼓励高校尽快形成人工智能与网络信息安全交叉学科的人才培养模式，组建和壮大人工智能安全师资队伍，促进国内外人工智能安全学生和教师共同开展研究，扩大人工智能数据安全人才培养规模、提高人工智能数据安全人才培养质量。二是加大企业人工智能数据安全人才培养。鼓励企业内部创办培训机构，或与科研机构、高校等建立联合人工智能数据安全培训基地，加强企业人员人工智能数据安全管理和技术能力培训。三是加强国外人工智能数据安全人才引进。制定人才政策引进专项人才，支持高校或企业引进世界一流人工智能数据安全领军人才；鼓励企业通过资本运作等方式吸纳掌握核心技术的人工智能数据安全团队。

# 云计算环境下的信息安全问题

## 发展现状：

2017年我国云计算整体市场规模达691.6亿元，同比增长34.32%。其中，公有云市场规模达到264.8亿元，同比增长55.7%，预计2018—2021年仍将保持快速增长态势，到2021年市场规模将达到902.6亿元；私有云市场规模达到426.8亿元，同比增长23.8%，预计未来几年将保持稳定增长，到2021年市场规模将达到955.7亿元。

从市场结构来看，我国云计算市场以私有云为主，2017年其占比达到61.71%，而在全球云计算市场中，公有云市场份额达到88.35%。国内外云计算市场的结构差异主要来源于国内客户对云计算的了解不足、云计算标准缺失、与原信息系统的兼容性问题等因素，目前大中型企业是我国云计算服务的主要用户，而出于对安全性和可控性的追求，该部分客户通常选择私有云作为其IT部署架构，造成了我国云计算市场中公有云的市场份额较少。

同时，从服务模式来看，我国公有云市场内部也存在着企业级SaaS服务市场占比偏低的问题：2017年，国内IaaS市场成为游戏、视频、移动互联网等领域中小企业IT资源建设的首选，占据了云计算市场56.16%的份额；而由于我国SaaS服务的开放程度不够、易用性不足以及市场环境欠佳，尽管SaaS服务亦在公有云市场规模中占据了一定比例，但是其市场份额主要来自于以云存储等基础服务为主的消费级SaaS市场，企业级SaaS市场仅占SaaS服务市场规模的10%。我国企业级SaaS市场的发展缓慢也是导致了SaaS市场与其上游PaaS市场的规模仅占我国公有云市场的4成，与它们在全球公有云市场中约6成的市场规模形成了对比。

总的来说，我国云计算市场规模增长迅速，但是其体量仍与我国的经济总量并不相称：2017年，我国云计算市场的全球份额占比仅为约9%，而美国云计算四巨头亚马逊、微软、谷歌以及IBM的云计算营收达到全球市场份额的约40%，其中，仅亚马逊一家就占据全球云计算市场份额的31.50%。

这主要是由于，目前我国云计算市场用户仍以互联网原生行业，如游戏、电商、视频为主。该领域用户对云计算的接受与熟悉程度较高，且其自身特点亦适合云计算的部署方式，因此最早完成云计算架构的部署，而金融、政府、工业等对私密性、稳定性、实时性要求较高、系统迁移难度较大的行业其整体迁移时间较晚。未来，随着国家对于云计算发展的大力支持、对物联网所产生的海量大数据的存储与分析需求不断增长，以及相关云计算技术的继续更新与优化，我国云计算产业链的下游应用市场将得到持续拓展，云计算市场亦将随之不断壮大，为相关公司带来显著成长红利。

## 云计算环境下存在的信息泄露问题

### “云计算”数据平台安全问题

（1）当前云计算、数据存储、数据分析等一系列关键技术的快速发展直接带动了信息管理系统软硬件网络架构的深刻改变，由此直接带来了软件、硬件、协议等多方面的信息系统漏洞和潜在的安全隐患，而实际上现有的安全管理系统防护措施不足以防御这些未知系统漏洞可能带来的安全威胁和风险。

（2）当前大数据平台一般基于 Hadoop、Spark 等数据框架实现，缺乏健全的安全机制，尤其是在身份认证、权限控制、安全审计等方面不够完善，安全保障能力比较薄弱。同时大数据技术迅猛发展不断

催生出新的网络攻击手段，导致传统的入侵检测等防护技术无法有效抵御外界的强大攻击。

（3）在传统的网络防护模式中，不同网络域的安全防护边界一般比较清晰，而由于大数据安全防护技术主要采用分布式的数据存储和底层计算架构，其系统底层复杂、开放的计算架构特点使得大数据网络环境下的安全边界越来越模糊，之前基于传统边界安全的系统防御理念和技术基本无法得到广泛适用。

### “云计算”数据安全问题

（1）“云计算”数据处理平台汇聚了行业用户海量多源的数据，这些数据分别存储在数据池中，部分数据可能包含国家安全、社会治理、经济运行等敏感领域的信息，利用先进的大数据综合分析技术进行数据挖掘和分析时，需要进一步强化对数据的隔离和访问控制，否则将引发行业用户数据泄露的风险。

（2）“云计算”数据来源广泛、种类多样，并且被众多基于互联网的“云计算”设备和应用软件所采集，对“云计算”中输入的数据进行完整性校验，是采集“云计算”数据真实可信的重要保证。如果技术人员分析使用的数据是虚假有误的，势必会影响到“云计算”结果的客观性，甚至导致重大决策的失误。

### “云用户”个人隐私安全问题

（1）用户个人信息被滥用。现代社会信息资源开放、信息共享的基本诉求与对个人隐私信息安全保护的诉求存在天然矛盾。随着数据处理技术的不断创新，用户个人信息如社交网络账号、银行卡信息、地理位置信息等容易被不良企业非法收集、使用或清除，更有甚者将个人信息进行非法交易，形成数据流转的交易链条。

（2）数据深度挖掘增加保护难度。大数据时代，私人领域和公共领域的界限越来越模糊，人们利用大数据技术进行关联分析和深度挖掘，借助各种社交、购物平台从看似不相关的数据中获得个人身份信息、搜集个人行为习惯，甚至包含一些敏感信息。因此，个人信息的保护难度进一步增大。

## 阻止手段

使用虚拟化技术，实现有效隔离在云计算应用中要进一步实现数据的安全存储与虚拟化传输，需要提高虚拟化安全系数，实现不同数据的物理隔离和虚拟隔离。第一，要提高虚拟化软件的可靠性。虚拟化软件的作用是对多个用户所产生数据进行隔离，并且对用户访问权限进行限制。只有将众多数据在物理上进行隔离，才能为不同的用户数据提供虚拟化和数据隔离服务。第二，安装虚拟服务器时要从逻辑上对虚拟服务器进行独立的区域划分，并考虑使用多核心微处理器，实现对 CPU 的隔离与数据缓存。

设置身份认证和访问权限，确保用户安全“云计算”数据服务的对象是各大用户，通过“云计算”服务的用户越多，数据也会随之增加，这就需要服务商加强对“云计算”用户真实身份的管理和保护。“云计算”服务商除了提供更加人性化和理性化的身份信息监管之外，还应通过严格、统一的用户身份信息认证程序提升用户信息访问的安全，避免用户信息被恶意泄露。一方面要保证用户对云计算技术的利用率，另一方面也要对用户访问数据库的权限进行适度限制，可以从比较单一的身份认证模式转变为多种安全凭证并存的联合身份认证方式。

依靠法律手段，保障数据安全大数据背景下“云计算”数据管理解决方案一般采用全自动智能化的模式，即将所有用户使用“云计算”所需要的数据和信息进行获取、存储、管理和分析，并自动交付给应用程序运行。这一模式的最大特点是科技含量和自由度都很高，这代表着云计算技术的高水平发展，却在某种程度上意味着用户在安全性方面丢失必要的主动权。近年来随着国家对网络安全的不断重视，《网络安全法》与网络等级安全保护 2.0 等纷纷出台，分别对“云安全”平台、云上安全用户和云上安全产品提出了一套相应的网络安全管理要求，共同构建稳定安全的网络环境。