

数据，这一数字经济时代形成的资产，所有权是你的，还是我的？

近日，十三届全国人大常委会第二十九次会议审议了“数据安全法草案”等。数据安全法问题，即将走向法治新阶段。

近年来，“特斯拉维权”“护险行动”等事件，引发社会对数据归属权、所有权等问题的广泛讨论。而由数据侵权带来的损失和危害，也日益凸显。

2020年7月，《中华人民共和国数据安全法(草案)》向社会公开征求意见；今年5月，国家互联网信息办公室发布关于《汽车数据安全管理办法(征求意见稿)》，意味着这些问题即将获得确定答案。

### 数据产权模糊不清

当今社会，正在从信息时代跨入数字经济时代。发展数字经济，成为我国实现新旧动能转换、培育新业态发展的重要路径。

处于数字经济时代初期，无序发展带来了诸多隐患。今年5月13日，工业和信息化部通报下架90款侵害用户权益的App。这些被下架和通报的App的侵权行为主要包括“违规收集个人信息”“超范围收集个人信息”“收集与其提供的服务无关的个人信息”“强制、频繁、过度索取权限”等。

“现在面临最大的挑战之一，是数据产权模糊。”在日前举行的2021北京智源大会上，清华大学国家金融研究院院长朱民在主旨报告中表示，“数据谁采集谁拥有，随意复制，触犯隐私权，造成数据泄露等。”

朱民认为，目前数据最关键的问题是产权不清楚。例如，消费者在网上购物，消费支付是银行、网络是平台、购物是商家，数据在那么多地方流通，“谁是数据的最后拥有者，这是很值得讨论的法律问题”。

美国加州大学伯克利分校计算机系教授宋晓冬介绍，美国加州政府的机动车辆管理局也在销售驾驶员个人的信息，且每年可以获利达到千万美元的级别。“这些数据在被买卖的时候，都声称是被匿名处理之后才进行的，但从实际研究中发现，这些匿名的数据其实并不能充分保护用户的隐私。”

国际咨询机构波耐蒙研究所受IBM委托，针对遭遇数据泄露的524家公司企业做了调查研究，形成了《2020数据泄露事件损失报告》。该报告显示，2020年数据泄露事件平均损失达386万美元；行业之间数据安全事件损失的差异巨大，医疗保健行业数据泄露事件的平均损失为713万美元，而遗失或被盗数据的平均损失为146美元。

“隐私和安全问题突出，带来的损失日益凸显。”同时朱民表示，数据不流通就会形成数据孤岛，孤岛问题造成数据的集聚和密度下降，数据使用效率就大打折扣，“这是我们面临的最根本的也是最基本的挑战”。

### 数据资产化

在现有的数据买卖过程中，虽然交易量很大，但是用户其实并没有从中得到足够的收益。宋晓冬表示，对数据的确权很重要。

“数据本身没有价值，但数据在使用的时候，体现出了巨大的价值，这为我们考虑数据资产化提供了很重要的前提。”朱民在多个场

# 数据是你的还是我的？

■本报记者郑金武



图片来源:unsplash

合呼吁将“数据资产化”。

“随着大数据技术的兴起，数字经济的发展越来越让我们意识到，数据不是一种资源，更是一种资产，而且越来越明确的是，只有把数据从资源变成资产，智能社会才能得以运转。”朱民强调，“智能社会的出发点和起点是数据，而数据如果不是资产的话，没办法承担起重大的责任和经济功能。”

事实上，“数据是一种资产”的观点，早在十年前就有专家提出。数据资产有一系列的特征，包括资源丰富、更新频繁、品种多样；有准公共物品的特征；可以不断产生原始和衍生的数据，因此可以不断产生新的价值；数据可以被交易，能够多方面满足人类生存和发展的需求；数据集合使用的话，其价值更高。

“这些特征使得数据这种资产的性质非常好。”朱民表示，数据要有足够的规模、足够的维度、足够的密度，要支撑这3个基本概念，数据必须资产化。

对此，翼方健数公司联合创始人陈恂担忧的是，“我把我的数据给了你，我同时还可以复制一份数据给别人。这种非经济类的特征或决定了数据所有权并不适合交易”。因为“一是数据是虚拟的、可重复使用的，二是数据有相对高昂的固定成本，但却可以几乎零成本复制，三是数据具有非竞争性，不是排他的，一份数据可以使用在不同的场合”。

“如果把数据交易放在所有权交易的层面的话，会遇到很大的困难。”陈恂认为，在现阶段，数据的所有权模糊不清，让数据交易处于一个尴尬的境地。

朱民对此也表示认同。“数据有一系列非金融学的维度，这是我们面临的很大挑战。同时，隐私、合规、保密等，这些非经济学的维度，解决起来还是挺难的，需要我们整个社会提升治理能力。”

### 释放数据真正的能量

数字经济时代，数据成了重要的生产要素。“这个时代是数据先行的时代，只有有了数据，才有可能从数据当中去攫取知识，最后被用到社会中去。”陈恂表示。

全球数据经济在飞速增长。宋晓冬介绍，据欧盟估测，2020年个人数据产生的价值占了欧洲GDP总量约8%，全球经济产值估计达到万亿美元级别。

在此大背景下，数据安全作为网络安全产业重要组成日益增强。调研数据显示，我国2020年数据安全产业规模约356亿元，预计未来五年行业复合增长率在15%左右。随着数据安全法落地，数据安全上升到国家安全层面，行业将进入高速发展期。

陈恂介绍，随着人工智能的发展，数据的

# 概念莫混淆：智能计算机不是超级计算机

■本报记者 赵广立

日前，武汉城市人工智能(AI)计算中心建成投用，媒体在报道该中心算力时称：“中心一期建设规模100P FLOPS人工智能算力……其算力相当于5万台高性能计算机。”该报道一出，引发高性能计算领域专家关注。相关专家对《中国科学报》表示，该报道内容对“超级计算机”与“智能计算机”认识不清、概念混淆，极易引发误导。

中国科学院计算技术研究所研究员、中国计算机学会高性能计算专业委员会秘书长张云泉告诉《中国科学报》，“智能计算机不是超级计算机，两者的概念应该区分清楚，否则可能会引起行业混乱。”

### 超级计算机还是智能计算机

目前业界用于衡量超级计算机的Linpack测试，测试的是超级计算机的“双精度浮点运算能力”，即64位浮点数字的计算(FP64)。在以二进制所表示的数字精度中，还有单精度(32位，FP32)、半精度(16位，FP16)以及整数类型(如INT8、INT4)等。数字位数越高，意味着人们可以在更大范围的数值内体现两个数值的变化，从而实现更精确计算。

与许多科学计算等不同，AI所需的算力不需要太高精度。“比如一些AI应用需要处理的对象是语音、图片或视频，运行低精度计算甚至整型计算即可完成推理或训练。”张云泉说，这种专用计算机处理AI算法速度快、能耗低，这是由其特点决定的。

智能计算机是一种专用算力，在推理或训练等智能计算方面的表现出色，但由于AI推理或训练一般仅用到单精度甚至半精度计算、整型计算，多数智能计算机并不具备高精度数值计算能力，这也限制其在AI计算之外的应用场景使用。

而超级计算机是一种通用算力，其设计目标是提供完备、复杂的计算能力，在高精度计算方面能力更强、应用范围更广，比如科学家常使用超级计算机进行行星模拟、新材料开发、分子药物设计、基因分析等科学计算和大数据分析。

中国工程院院士陈左宁曾形象地将使用

超级计算机做AI计算比喻成“大马拉小车”，来说明超级计算虽然“十项全能”，但毕竟不是为AI量身打造。智能计算机由此兴起。

此前曾一度热炒的AI与超算融合，其实正是将改良的计算机“AI专用化”，严格意义上来说，它们已经不再属于传统超算范畴。

“现在不论超级计算中心落成还是智能计算中心，都宣称算力是多少‘FLOPS’，其实这个单位是‘每秒浮点运算能力’，而一些智能计算机的单位其实是‘OPS’——每秒操作次数。如果不加区别地报道，大家很容易误认为是同一种计算精度、同一种计算能力。”张云泉说，这也导致一些地方以为花了小钱建了世界顶级的“大超级计算”，好像占了便宜；等项目上了马，将项目介绍给超算业内人士时，才恍然大悟。

### 用新指标引导行业健康发展

张云泉告诉记者，除了混淆智能计算机和超级计算机之外，业内还存在另外一种误导。

“有些厂商还会模糊智能计算机的推理性能和训练性能。”张云泉告诉《中国科学报》，“与推理相比，训练性能往往需要计算精度高一些，比如32位甚至64位；而大部分性能‘耀眼’的AI芯片，往往指的是其推理性能，而且可能只是理论值。”

对于AI计算而言，训练性能往往更重要——许多智能模型正是依赖于此。张云泉说，训练模型的计算量与参数量成正比，而且需要反复迭代，直到达到理想的效果。

如果要画一张AI所需算力的示意图，“推理”位于算力矩阵的最下层，因为半精度算力(FP16)或整型算力(如INT8)即可满足推理需要；排在其上的是“训练”，一般需要使用单精度算力(FP32)或半精度算力(FP16)；对算力需求最高的是类脑“模拟”，它需要双精度算力(FP64)和低精度算力同时支持。

“混淆传统超级计算与智能计算、混淆智能计算的训练性能和推理性能，这两种情况都可能会导致用户或地方政府错误决策——他们本意是建设强大的计算集群，最后却只建成

了一台只有推理性能的机器。”张云泉说。

对此，张云泉认为，需要一个简单、有效的指标来帮助判断系统的AI算力和整个高性能AI领域的发展状况，从而防止行业乱象。

2020年11月，张云泉联合清华大学教授陈文光、美国阿贡国家实验室研究员 Pavan Balaji 和瑞士苏黎世联邦理工学院教授 Torsten Hoeferl，与 ACM SIGHPCC 委员会共同发起了基于 AIPerf 大规模 AI 算力基准评测程序的“国际人工智能性能算力 500 排行榜”。记者注意到，这一榜单的算力单位是 OPS。

“超级计算与AI计算，一码归一码，需要一个新的标尺来引导AI计算行业走上健康发展的道路。”张云泉说。

### 国产AI芯片正待奋起直追

算力始于芯片。在AI芯片赛道上，我国拥有华为(昇腾)、百度昆仑、燧原等芯片设计企业，但即便如此，国内的智能计算机仍很少能绕过美国GPU巨头英伟达。

这是一个无奈的现实：国内上马诸多智能计算中心，英伟达是切切实实的受益者。

“专门做智能计算的AI芯片，只要核数足够多、主频足够高，就可以实现速度更快、在低精度计算中高出几个量级的性能。但如果某个计算集群既需要高精度计算又需要低精度计算，这对AI芯片的要求就高了。”张云泉说，英伟达的GPU各种精度的计算能力都很突出、比较均衡。这也是大部分国产AI芯片难以与英伟达GPU硬碰硬的原因之一。

不过，国产AI芯片并非完全没有机会。

首先，当下我国的算力基础设施，都有强烈的国产化意愿。即便英伟达、英特尔等巨头虎踞龙盘，但基于综合成本、生态等各类因素，国产化的大潮仍不可阻挡。

其次，就AI当前发展而言，场景、数据、模型、算力缺一不可，这也就意味着，中国将是未来全球AI算力富集地。AI芯片作为核心需要，不可能被一种形态、一种生态所垄断，寒武纪、昇腾等国产AI芯片的佼佼者仍坐拥巨大发展空间。

专家认为，虽然芯片是算力的主要来源和

驱动作用日益显现。事实上，数据已成为人工智能发展的重要条件之一。“要把数据作为生产要素，让它释放出真正的能量。”

而人工智能的发展，也促进了数据价值的挖掘。“人工智能的能力能够从这些数据当中去提炼洞察，提炼知识，同时把这些知识和洞察放在产品之中，最后服务社会。”朱民说。

那么，如何兼顾数据的使用与安全呢？“我们需要构建一个负责任的数据经济框架。”宋晓冬表示，“负责任，是指数据经济发展需要目标和原则”。

宋晓冬介绍，首先需要保护数据权益，数据权益是数据经济的基石，只有在保护数据权益的情况下，才能防止数据的滥用；其次是需要公平分配数据产生的价值，让用户能够从自身的数据受益；三是能够最有效地使用数据，来实现社会的福利和经济效率的最大化。

要达到这些目的，以前在物理世界中的经验要有所变革。“数据使用和隐私之间有天生的对立性，一方面我们需要更好使用数据，从数据中得到它的价值，另一方面我们又需要保护数据的隐私。”宋晓冬说。

朱民也表示，中国数据产生越来越多，在这个过程中，政府起着特别重要的作用。“政府在个人信息使用边界、数据开放、支持数据从资源走向资产等方面，应该发挥作用，打通数据流通通道，让数据动起来。”

### 守护安全从法律做起

朱民表示，我国在数据发展方面已走在世界的前沿，必须要抓紧构建生态，提升数据内生安全性和科技向善性、可信性。

宋晓冬介绍，近年来欧洲、美国都颁布了相关隐私法。“这虽然也会加重企业的负担，但更重要的是，相关法规的完善，一方面有利于保护数据的隐私性，另一方面也有利于打破数据孤岛，发挥数据的价值。”

我国从2018年开始，启动修订数据安全方面的法规。2020年7月公布的《中华人民共和国数据安全法(草案)》显示，国家坚持维护数据安全和促进数据开发利用并重，以数据开发利用和产业发展促进数据安全，以数据安全保障数据开发利用和产业发展。

同时该草案强调，国家实施大数据战略，推进数据基础设施建设，鼓励和支持数据在各行业、各领域的创新应用，促进数字经济发展；省级以上人民政府应当制定数字经济发展规划，并纳入本级国民经济和社会发展规划。

而国家网信办于今年5月发布的《汽车数据安全管理办法(征求意见稿)》指出，运营者收集个人信息应当取得被收集人同意，法律法规规定不需取得个人同意的除外；实践上难以实现的(如通过摄像头收集车外音视频信息)，且确需提供的，应当进行匿名化或脱敏处理；只有为了方便用户使用、增加车辆电子和信息系统安全性等目的，方可收集驾驶员指纹、声纹、人脸、心律等生物特征数据，同时应当提供生物特征的替代方式。

“我国的数据安全法律草案，规定了政府数据安全开放，建立政府数据开放平台等，总体的思路是鼓励公共数据共享和开放。”朱民表示，“世界正在离开信息时代进入智能时代，数据、算力和算法都在加速发展，我们要让数据成为创造力、生产力、竞争力和财富。”

最根本的物质基础，但是算力的生产、聚合、调度和释放是一个完整过程，需要复杂系统的软硬件生态共同配合，才能实现“有效算力”。因此，不能只关注芯片的单一性能指标，更要注重上层软件应用生态。

### 没有巨大算力无法发展AI?

在算力概念被混淆的背后，是AI计算有如脱缰野马一般疯长的算力需求。

由多位硅谷“大亨”联合建立的人工智能非营利组织OpenAI，在2020年5月推出了其新一代无监督的转化语言模型GPT-3，目前已有1750亿参数，训练数据量达到45TB(约1万亿单词量)。

GPT-3模型目前已经在语义搜索、文本生成、内容理解、机器翻译等方面取得重大突破。其最大价值是证实了机器在无监督下的自我学习能力，验证了纯粹通过扩大规模即可实现性能提升。

更壮观的是，万亿参数模型已经在路上。6月初，北京智源人工智能研究院发布了“悟道2.0”，宣称达到1.75万亿参数，超过之前由谷歌发布的Switch Transformer，成为全球最大的预训练模型。

迅猛增长的参数体量，也意味着更高的计算需求——有的可能需要数千块GPU来提供必要的算力。张云泉说，类似GPT这样的巨模型，对算力的需求“不是闹着玩的”。

难道，没有巨大算力就无法发展AI吗？

张云泉认为，在目前AI的发展阶段(感知智能和认知智能)中，算力仍然是第一位的。

他的理由是，发展AI可以通过算力提升、算法革命等途径进行，但在“资本导向”的现阶段，相比不确定性的算法模型突破，算力提升是个更容易的选择。

但必须要指出的是，运用巨大算力并不是人工智能发展的唯一方向，GPT-3这样的巨模型同样存在缺陷，如缺乏常识等；而探索人脑奥秘机理，实现小数据学习、迁移学习也是重要手段。

毕竟，大脑的功耗只有20瓦左右，创建低能耗的智能系统或许是更重要的努力方向。

### 前沿

本报讯(记者袁一雪)近日,《自然—通讯》刊登了来自英国格拉斯哥大学的一篇论文。该论文作者之一Lee Cronin阐述了其研发的“机器人化学家”,能够通过机器学习算法,自主决定将18种化学物质中的哪些化学物质组合到反应器中,以及如何设定发生反应的条件。这意味着该机器人能够自行运行实验,而无需人工监督。

研究人员称,这项研究旨在通过让机器在数周内不间断实验,为诠释地球上复杂的有机生命是如何从简单、无生命的化学物质发展而来提供新见解。

在开发过程中,研究人员通过测量每个实验产物的质量指数,让机器人了解每个反应产生的分子的复杂性。这些信息有助于它学习如何改变实验,以在随后的反应中创建一个更复杂的分子。该团队希望这是一个数字化版本,用以探索自然选择的复杂性。

通过数百次的实验,研究人员不仅发现了复杂的分子可以被创建,还发现一些新分子内部存在生命周期。这表明可能正在发生其他过程,如催化和复制。

在发表这篇论文之前,该团队还发表过一篇关于组装理论的论文。组装理论是研究人员开发的用以量化分子复杂性的工具,还可用于识别生命化学组成部分的迹象。而机器人化学家通过这些理论,不断重复过程,让研究人员看到了这一过程中更多可能性。

很少有实验能够持续与生物系统发展时间一样的时间长度。但机器人化学家则能够弥补这一点,在数周内进行数十次递归实验,然后最终持续数月甚至数年,让研究者了解从生命诞生之初到复杂性的变化过程。

相关链接:

<https://doi.org/10.1038/s41467-021-23828-z>



生命起源实验反应过程。

### 速递

## 计算机领域多位华人学者当选加拿大工程院院士

本报讯 当地时间6月14日,加拿大工程院官方网站公布的52名新院士中,有13位华人入选,包括微众银行首席人工智能官杨强,香港中文大学(深圳)数据科学学院教授张大鹏,香港城市大学校长、中国工程院外籍院士郭位在内的计算机科学领域的知名学者在列。

其中,杨强是国际领先的大型人工智能和数据挖掘解决方案工程实践专家、国际人工智能界“迁移学习”领域的发起人和带头人,也是国际“联邦学习”的发起人之一及带头人,为全球数百万用户和企业开发了用于普惠金融、健康和电子商务应用程序的开源系统;香港中文大学(深圳)数据科学学院教授张大鹏被公认为生物识别工程的先驱,在多个研究领域取得了成绩,包括掌纹认证、医学生物识别传感器和系统;中国工程院外籍院士郭位则为环境可持续性发展和能源系统的最佳可靠性设计做出了根本性的贡献。(袁一雪)

## 上海大学着力推进微电子学院建设

本报讯 日前,“新发展格局下的集成电路交叉创新论坛暨上海大学微电子学院建设推进会”在上海大学嘉定校区召开。

会上,上海大学与中国科学院相关研究机构签署合作备忘录,与代表性企业签订战略合作协议;上海微技术工业研究院资助上海大学8英寸MEMS中试线每年四分之一总机时,用于微电子学院教学实践,为期五年。此外,上海大学成立了集成电路教育基金支持人才培养,中天恒星(上海)科技有限公司首家捐赠1000万元人民币。(秦志伟)

## 七大技术将持续影响未来金融行业

本报讯 日前,全球管理咨询公司麦肯锡发布《Fintech 2030:全球金融科技生态扫描》报告。报告认为,回顾近年来金融科技发展,展望未来趋势,包括人工智能、超自动化、区块链、云计算、物联网、开源软件、无代码开发平台在内的七大技术将持续影响金融行业的未来。

麦肯锡全球资深董事合伙人曲向军介绍:“上述关键技术并非孤立存在,而是相互交织、嵌入和融合的,从而迸发出巨大能量并引领行业变革。目前这些核心技术组合已经在金融细分领域逐渐实践应用,未来将发挥更大价值,并影响金融科技乃至金融行业的市场竞争格局。”

曲向军表示:“未来十年,是科技驱动和科技赋能金融行业的黄金时期,也是金融行业业务模式变革和竞争格局变化的关键阶段。中国的金融科技企业、投资机构 and 传统金融机构须时刻关注全球金融科技的技术趋势和业务模式创新,结合中国实际,深挖中国庞大的市场潜力,围绕金融场景,打造金融生态,持续建设自身能力与核心竞争力,推动转型与开放合作,迎接即将到来的金融科技浪潮。”(郑金武)

# 机器人化学家洞察生命起源