DeepMind、苹果爆发隐私数据泄露案

数据保卫战如何打？

本报记者 李佳师

不久前英国医疗服务体系（NHS）医院向Alphabet旗下DeepMind提供了约160万患者的详细资料，DeepMind因数据隐私问题被英国信息安全员判定为违法。与此同时苹果公司也被爆“内鬼”倒卖20万条信息，涉案金额超5000万美元。数据隐私安全再度被关注。我们正在进入数据社会，当一切变得越来越数据化时，我们应该如何规范数据隐私的保护，如何制定相关数据保护政策，如何利用技术来保护数据隐私安全？

数据监管
如何跟上数据创新的脚步

尽管英国医疗服务体系（NHS）医院向Alphabet旗下DeepMind提供患者的详细资料，是用于开发和完善急性肾损伤（AKI）诊断和检测系统，与苹果公司“内鬼”倒卖20万条信息出发点完全不一样，但它依旧被裁定违法。英国最高隐私保护监管部门认为，这些医疗记录数据使用时并没有告诉医疗患者数据将被使用的方式。

眼下，我们正在利用人工智能技术来攻克一个又一个的难题，而在解决这些难题的过程里，除了算法，非常重要的一个维度就是需要大量的数据。但是这些数据的归属权界定与去隐私化问题就需迫切需要解决了。

近日，IBM Watson和云平台高级副总裁David Kenny向美国国会议员发出了一封公开信，概述IBM就人工智能技术兴起所引发的相关政策问题的看法，谈到数据隐私与归属权时他表示：“IBM始终认为，个体数据为个人所有，相关的数据政策应公平合理、优先考虑开放性，并尊重知识产权。”AI协作共事的人员需负责地管理公共和私人数据。

NHS与DeepMind的数据合作与苹果公司“内鬼”倒卖信息，在阿里数据经济研究中心秘书长、阿里研究院高级专家潘永花看来，涉及数据监管的不同维度，一个是政府和行业监管，一个是企业内部的数据管理。DeepMind事件是由于数据流通和数据开发利用导致了隐私泄露，苹果公司内部“内鬼”事件是因内部数据管理制度不健全而致。

“从数据监管的维度看，个体数据其实是进行数据创新最有价值的数据，也是数据安全等级最高的数据，我们利用数据的同时必须保障数据隐私，数据监管如何跟上数据创新的步伐，在法律上、在管理模式上需要进行更多的探索。”潘永花在接受《中国电子报》记者采访时表示，就像个人医疗数据，不仅仅与个人身份有关，更与个人的身体状况有关，是我们进行精准医疗、个性化医疗，推进人工智能进行疾病诊疗的关键基础。如何将这些数据实现匿名化收集，进行匿名化处理，这涉及技术问题，也涉及监管模式和法律的问题。

赛迪智库互联网研究所副所长陆峰对《中国电子报》记者表示，这两个事件再次警示我们要加强对数据安全的全生命周期管理，数据从产生到消亡整个生命周期过程中都有可能发生数据隐私泄露，要求我们制定相关措施，从技术保障、规范管理、法律法规等多方面入手，加强数据采集、传输、存储、流通、交易、开发和利用等全生命周期管理。

中国信息化推进联盟委员、国标委金融标准化专家、数秦科技首席科学家王毛路在接受《中国电子报》记者采访时表示，数据隐私事件已成为社会的焦点，侵犯数据隐私问题的不断发生说明了几个问题：个人针对数据隐私的意识还不够完善，数据保护政策未能完全贯彻落实，监管层面暂时还未能实现隐私数据的安全监管。要想更好地保护数据隐私，需要从数据分类、数据采集、数据存储、数据流通等环节制定相应的措施。比如分类措施，需要将数据进行分类，明确数据类型，划定隐私数据边界。比如数据采集，要制定隐私数据采集规定，明确在何场景下通过何种方式进行数据采集。比如数据存储，凡需存储隐私数据的单位和企业，需制定严格的数据存储规范。比如制定数据流通规则，明确各类数据的流通标准。比如制定数据安全保护制度，构建物理安全保护体系及网络安全保护体系。

也就在记者稿件成文之际，外电报道印度电信运营商Jio超过1亿用户的信息遭到泄露，成为印度电信行业有史以来最大规模数据外泄事件。这究竟是“内鬼”，是外部黑客，还是“内外勾结”所致，目前尚在调查中。

印度运营商用户信息泄露不是全球大规模数据泄露的第一个也不是最后一个事件。事实上，中国在此之前也爆发过多次互联网用户数据泄露案件，比如12306用户信息泄露、国家电网旗下App用户信息泄露等。

潘永花表示，过往几年中，基于数据买卖的地下灰黑产业非常猖獗，带动了消费者个人和国家对个人信息保护的关注，目前来看我国个人信息保护相关的法律规范尚不完善，只在一些法律中有零散规定，仍然存在效力层级低、法律法规协调性弱、保护内容片面等立法不足，有待于加强和完善。

企业内部
如何进行数据管理

在数据的全生命周期管理中，企业内部是非常关键的一个环节。陆峰表示。苹果“内鬼”事件中，如果我们采用相关的加密或认证技术对内部人员数据查看和拷贝采取严格的认证措施，就能从一定程度上降低数据被窃取的风险。不过，千万别迷信仅从技术入手就能保障数据绝对安全。

潘永花认为，苹果内鬼倒卖数据问题涉及的主要是在企业的内部如何进行数据安全管理的问题，企业内部的问题通过组织、规章与标准的规范以及技术的采用是可以避免的。

未来所有的企业都会变成数据企业，如何找到适合自己的数据管理的组织架构、规章制度、标准以及技术，需要进行更多的探索。互联网企业是目前拥有数据最多，也是基于数据进行业务创新和探索最早的一批企业。它们数据管理的方法、路径对其他企业进行数据管理有一定的借鉴意义。

“以阿里巴巴为例，在阿里巴巴这样的互联网公司，数据事实上是一切业务的来源，所以必须非常重视数据安全，会从组织架构、规章制度到标准以及技术等维度来规范和保障数据的安全。”潘永花说。

据介绍，目前阿里巴巴在集团层面设有CRO（首席数据风险官）以及数据安全管理小组来管理数据，除了顶层设计，在各个业务部门都设有数据安全保障岗位，每一个进入阿里的员工都要进行数据安全考试。在阿里，数据是按四个等级来进行分级的，涉及隐私的最高安全等级的数据谁都不能碰。涉及数据隐私的数据，需要利用技术的手段进行脱敏、打标等。据透露，目前阿里基于自己经验生成的“数据安全成熟度模型”已经开始对外输出服务，应用到国家电网以及蒙牛等企业。

技术如何能够
为数据保护发挥更多作用

王毛路认为，身份认证、数据加密、区块链等技术能够在保护数据隐私安全上起到很好的作用，其中区块链是新技术，可在保护数据隐私安全上起到积极的作用。因为区块链技术具备不对称加密、安全传输、不可篡改、可追溯等特点，可以很好地应用于保护数据隐私安全。通过区块链技术，在授权的前提下，可实现数据的加密传输，且能实现中间节点不留存数据；可实现数据的授权管理、传输管理、数据追溯等。

陆峰认为，区块链技术从安全、成本、效率三个角度考虑应用场景，最为适合银行间点对点支付清算系统、证券交易过程中券商点对点支付清算系统。目前上述两个系统都是点对点进行支付清算，都是央行和证交所集中式的清算。随着业务量的与日俱增，集中式清算模式瓶颈问题日益严重，加快区块链技术应用，推进银行间点对点支付清算、证券交易过程中券商点对点支付清算，能够提高效率。因为涉及点对点清算，需要防止参与清算的各个主体之间出现数据窃取、篡改和抵赖等问题，区块链技术正好能够保证这个应用场景数据安全，所以国外金融系统应用区块链技术保障数据安全的呼声很高。

但陆峰也强调，区块链技术不是万能的安全技术，针对具体的数据安全问题，需要我们对症下药。

英特尔推出至强可扩展处理器

本报讯 7月12日，英特尔公司宣布推出至强可扩展处理器。作为近十年来数据中心领域最大的技术进步，该处理器可为计算、网络和存储带来针对工作负载优化的性能，向下一代云基础设施提供坚实基础，并赋能数据分析、人工智能、高性能计算、网络转型等各类应用，以加速企业数据中心现代化及业务转型的实现。

29家来自全球各地的生态系统合作伙伴莅临现场。此外，来自腾讯云、苏宁云商、海鑫科金、国家气象局等不同行业、领域的用户代表分享了他们的最新用例。

英特尔公司数据中心事业部副总裁兼IT变革事业部总经理Lisa Davis表示：“为了满足企业应用、通信服务提供商、高性能计算、云服务提供商和人工智能技术等细分应用市场的多样化需求，全新的英特尔至强可扩展处理器不仅具备领先的性能和业务连续性，还采用创新的平台设计理念，提供铂金、金牌、银牌和铜牌四个品类、不同特点的产品系列，加速创新业务的实现与应用体验的改善，为未来数字服务基础设施平台提供全新动力。”

相比上一代产品，英特尔至强可扩展处理器的整体性能提升达1.65倍，OLTP仓库负载比当前系统提高达5倍——从而加速包括建模与仿真、机器学习、高性能计算和数字内容创建在内的工作负载。

借助新功能实现显著的性能提升，包括可提高计算密集型工作负载性能的Intel AVX-512、能够降低系统延迟的全新网格架构、用于加密和数据压缩硬件加速的Intel QAT以及集成IntelOPA（英特尔Omni Path架构）的高速互联，以部署更具成本效益的高性能计算集群等。

浪潮发布M5新一代服务器强调场景化

本报讯 7月12日，浪潮在北京正式发布新一代服务器M5系列产品，全线升级到Intel最新计算平台。浪潮M5新一代服务器针对智慧计算的时代需求所设计，聚焦云计算、大数据、深度学习（简称CBD）应用场景，强调极致的场景设计和应用优化，分为通用、融合架构、应用优化和关键业务4大系列35款产品，提供丰富的场景产品阵列。

浪潮集团副总裁彭震表示，智慧计算时代下，服务器需要打破传统设计思路，从均衡设计走向场景化的极致设计，即在关注服务器传统RASUM五大特性之外，对服务器空间、能耗、弹性、开放性提出新的要求，聚焦应用场景的极致设计和应用优化增强。彭震介绍，场景化设计是从部署环境和应用环境的具体需求出发，在密度、能耗等物理设计上，以及计算、存储和I/O配比等逻辑设计上，进行产品定义和产品实现。

浪潮M5新一代服务器根据不同的部署场景以及应用场景考量，将产品细分为4大系列，为用户提供更为优异的计算性能和可靠高效的业务保障。浪潮M5新一代服务器，计算性能峰值提升125%；多路服务器计算密度提升100%；存储服务器的存储密度是上一代的2.78倍；业界最全的AI的产品线，部署密度、峰值性能、硬件解耦处于业界领先水平，相比上一代产品性能提升1.5倍。通用服务器和关键业务服务器面向传统的企业应用，追求极致的RASUM特性，为ERP、CRM等传统企业应用提供强大、可靠、灵活的支撑平台。例如NF5280M5在2U空间内可扩展56个物理核心，24个内存插槽和10个标准PCI-E接口，可配置为30多种应用方案，覆盖通用计算、异构计算等5类应用场景。

产业观察

人工智能发展应高度重视大数据支撑作用

潘文

2016年以来，全球迎来人工智能发展新一轮浪潮，人工智能成为各方关注的焦点。从软件时代到互联网，再到如今的大数据时代，数据的量和复杂性都经历了从量到质的改变，可以说大数据引领人工智能发展进入重要战略窗口。

从发展意义来看，人工智能的核心在于数据支持。首先，大数据技术的发展打造坚实的素材基础。大数据具有体量大、多样性、价值密度低、速度快等特点。大数据技术能够通过数据采集、预处理、存储及管理、分析及挖掘等方式，从各种各样类型的海量数据中，快速获得有价值信息，为深度学习等人工智能算法提供坚实的素材基础。人工智能的发展也需要学习大量的知识和经验，而这些知识和经验就是数据，人工智能需要有大数据支撑，反过来人工智能技术也同样促进了大数据技术的进步，两者相辅相成，任何一方技术的突破都会促进另外一方的发展。其次，人工智能创新应用的发展更离不开公共数据的开放和共享。从国际上看，开发、开放和共享政府数据已经成为普遍潮流，英美等发达国家已经在公共数据驱动人工智能方面取得一定成效。而我国当前仍缺乏国家层面的整体战略设计与部署，政府数据开放仍处于起步阶段。在开放政府数据成为全球政府共识的背景下，我国应顺应历史发展潮流，抓住大数据背景下发展人工智能这一珍贵历史机遇，加快数据开发、开放和共享步伐，提升国家经济与社会竞争力。

从发展现状来看，人工智能技术取得突飞猛进的进展得益于良好的大数据基础。首先，海量数据为训练人工智能提供了原材料。据We Are Social公司统计，全球独立移动设备用户渗透率超过了总人口的65%，活跃互联网用户突破了40亿人，接入互联网的活跃移动设备超过了50亿台。根据IDC预测，2020年，全球将总共拥有35ZB的数据量。如此海量的数据给机器学习带来了充足的训练素材，打造了坚实的数据基础。移动互联网和物联网的爆发式发展为人工智能的发展提供了大量学习样本和数据支撑。其次，互联网企业依托大数据成为人工智能的排头兵。Facebook近五年里积累了超过12亿全球用户；IBM服务的很多客户拥有PB级的数据；Google的20亿行代码都存放在代码资源库中，提供给全部2.5万名Google工程师调用；亚马逊AWS为全球190个国家/地区超过百万家企业、政府以及创业公司和组织提供支持。在中国，百度、阿里巴巴、腾讯分别通过搜索、产业链、用户掌握着数据流量入口，体系和工具日趋成熟。再者，公共服务数据成为各国政府关注的焦点。美国联邦政府已在Data.gov数据平台开放多个领域13万个数据集的数据。这些领域包括农业、商业、气候、教育、能源、金融、卫生、科研等多个主题。英国、加拿大、新西兰等国都建立了政府数据开放平台。在我国，2011年香港特区政府上线data.gov.hk，上海率先在内地推出首个数据开放平台。之后，北京、武汉、无锡、佛山、南京等城市也都陆续上线数据平台。另外，基于产业数据协同的人工智能应用层出不穷。海尔借助拥有上亿用户数据的SCRM大数据平台，建立了需求预测和用户活跃度等数据模型，年转化的销售额达到60亿元；益海鑫星、有理数科技和阿里云数加平台合作，以中国海洋局的遥感卫星数据和全球船舶定位画像数据为基础，打造围绕海洋的数据服务平台，服务于渔业、远洋贸易、交通运输、金融保险、石油天然气、滨海旅游、环境保护等众多行业，从智能指导远洋捕捞到智能预测船舶在港时间，场景丰富。

综上所述，大数据为人工智能的发展提供了必要条件。现阶段，在大数据角度，制约我国人工智能发展的关键在于缺乏高质量大数据应用基础设施、公共数据开放共享程度不够、社会参与数据增值开发进展缓慢、标准缺乏时效性等。因此，需要从以下几个方面重点考虑：一是重点突破面向大数据应用基础设施。结合数据生命周期管理需求，培育大数据采集与集成、大数据分析与挖掘、大数据交互感知、基于语义理解的数据资源管理等平台产品。面向重点行业应用需求，形成垂直领域的大数据解决方案及服务。二是积极开展公共数据开发共享。国家要制定数据开放共享重大方针政策，加强统筹协调和分类指导。各地方要积极探索数据开放共享管理的新模式。鼓励有条件的地方探索建立数据开放共享管理部门，加强数据开放共享全过程的管理。三是鼓励社会力量参与数据再利用增值开发。建立数据社会化增值开放共享绩效评价制度，将数据社会化增值开放共享绩效评价列入电子政务效益评估的总体框架之中。设计可度量的指标，评估数据社会化增值开放共享的数量、质量、收费的合理性以及申请者的满意度。四是增强标准时效性。通过国家标准规定，要适应于移动应用的时代需求，提供相应的API，并规定API的基本格式，这样既能方便数据提供方进行API的开发，也大大降低了第三方软件开发者的开发复杂度，提高代码的重用率从而降低开发成本。

（作者为赛迪智库软件产业研究所所长）