让数据“可用而不可见”，人们不需要看到“黑箱”内是如何运作的，就能得到运算结果

隐私计算是打破“数据孤岛”的钥匙吗？

■本报记者 **沈湫莎**

数字经济时代，数据正越来越彰显其重要价值。当数据成为与土地、劳动力、资本、技术同等重要的生产要素，两个问题也随之而来。

一是当企业或机构通过大数据挖掘提供服务时，社会各方的数据权益如何保护；二是单独一家企业、科研院校或政府部门拥有的数据量毕竟有限，在某些情况下大家又难以坦诚地共享数据，客观上造就了一个个“数据孤岛”。

能否有一种机制，既能让彼此共享数据，又不会泄露核心机密呢？作为一种平衡数据利用和隐私保护问题的技术，隐私计算悄然兴起，正逐渐发挥越来越大的作用。

大数据“杀熟”，隐私计算说“不”

当下，各行各业都在推动数字化转型。数字化不仅是企业发展的必经之路，也成为经济社会发展的必然趋势。随着互联网、大数据、云计算、人工智能、区块链等技术加速创新并日益融入经济社会发展各领域全过程，数据的价值愈发凸显。

电商领域一个至今为人津津乐道的案例是，买了奶粉的家庭接下来大概率会买扫地机器人，这种被大数据挖掘出的正相关，预示着购买者家中可能刚刚添丁。金融机构也很早就开始用大数据勾勒用户画像，除了单位、收入、学历等基本信息，数据甚至覆盖到了出入场合、看书种类、浏览网页频率等几万个维度组合。类似的“大数据画像”固然能提高电商、企业、金融机构的营销效率，降低获客成本和运营风险，但数据挖掘的隐私边界在哪里，亦引发了诸多讨论。

中国互联网络信息中心（CNNIC）今年2月发布的第49次《中国互联网络发展状况统计报告》显示，截至2021年12月，我国网民规模达10.32亿，互联网普及率达73.0%；我国网民人均每周上网时长达28.5个小时，较2020年12月增加2.3个小时。显然，互联网已深度融入人们的日常生活，数据改变了人们的生产、生活和消费模式，这对社会各方的数据驾驭能力提出了更高要求。

在此背景下，在推进数字化转型过程中，信息安全和隐私保护被提到了前所未有的高度。2021年11月1日正式生效施行的《中华人民共和国个人信息保护法》，是我国第一部个人信息保护方面的专门法律。加上此前生效的《中华人民共和国数据安全法》《中华人民共和国网络安全法》，共同组成数据保护领域的“三驾马车”。

如何在数字化进程中兼顾信息安全和隐私保护呢？让数据“可用而不可见”的隐私计算或许提供了一种解决之道。上海树图区块链研究院研究总监杨光解释说，简单而言，隐私计算如同一个黑箱，人们不需要看到黑箱内是如何运作的，就能得到运算结果，“相当于人们共享了数据价值，又保护了各自的数字权益。”

在技术层面上，隐私计算几乎适配当前互联网的各个环节，让数据既能使用，又能保护隐私。因此它的出现，被业界视为“治疗”过度个人信息收集、大数据“杀熟”、隐私泄露等一系列问题的“一帖良药”。

想要大数据，但不想“被看见”

数据量越大、质量越高，越能进行准确的建模训练和预测。然而，并不是所有人都愿意公开自己的数据或模型，一来数据太容易被复制，担心拿出核心数据会降低自己的竞争力，二来某些领域的数据本就不适合公开。

比如，云计算本是一种节约成本、提高算力的方式，但某些企业不愿“上云”的原因之一就是，“上云”等于将自己的数据和算法模型一并交给了云服务商。如果有一种技术能限制云服务商查看运算数据的权限，企业“上云”的积极性将极大提高。

同理，在上下游企业或竞争对手之间，他们的数据往往高度相关且互补，如果能聚在一起，一定能发挥出比单打独斗更强的作用。但是，在保障自身数据安全的情况下，没有一家企业愿意拿出自己的核心数据。

上海交通大学网络空间安全学院长聘教轨副教授、光之树科技首席密码学科学家袁晨表示，破解“数据孤岛”的关键在于，在保障各方利益的情况下实现数据共享，而隐私计算是目前业界公认的方向之一，近年来发展迅速。

通俗地讲，隐私计算就是为解决“想要大数据，又不想‘被看见’”的需求而生。比如，寄快递时，“先寄后付”模式每单可节省少则5秒钟、多则一分钟。在疫情背景下，利用隐私计算技术，光之树科技将顺丰快递“先寄后付”的目标客户拓展了三倍，这是如何做到的呢？

袁晨解释说，“先寄后付”的目标客户是那些金融信用良好的用户，而顺丰并没有掌握这样的用户信息，但银行已经有相对完善的客户信用信息。然而，一般情况下，银行绝对不会将该核心信息交给别人。作为国内领先的隐私计算公司，光之树此前曾为中国银联提供了隐私计算解决方案，使得他们可以在不透露用户信息的情况下，只告诉顺丰该用户是否可以成为“先寄后付”的目标客户。由此，同一份数据发挥了更大的作用。

在产业数智化和精准营销领域，隐私计算也大有可为。比如，一家企业希望通过以数据库比对的方式，找出与现有客户相似的潜在客户，进行客户群体拓展。然而，企业都不希望将自己的核心用户名单开放给第三方。“你不给我，我怎么找？”隐私计算就能帮助企业解决这一两难问题。

互联网应用繁荣催生新赛道

每一种生产要素必须有与之相匹配的共享、交易机制，蓬勃发展的隐私计算无疑为数字经济的底层逻辑提供了一种技术支撑。

袁晨介绍，隐私计算分为硬件方案和软件方案。硬件方案就是可信计算TEE（可信执行环境），软件方案分为安全多方计算和联邦学习的隐私计算。目前，中国在该领域处于第一梯队。

“在同等条件下，软硬件结合的计算模式优于单靠软件计算。”近年来，袁晨在国际顶尖密码学会议上保持平均每年被收录1–2篇论文的频率。他也是目前为止，中国在国际三大密码学顶级会议上发表隐私计算相关论文数量最多的学者。

在硬件方面，支持隐私计算的机密计算服务器应运而生。它是指在芯片改造和硬件支持下，可以支持在运算过程中看不到其中的运算数据、只输出运算结果的服务器，适用于大规模数据运算。

作为异军突起的细分赛道，近年来隐私计算在中国的发展尤为迅猛。全球范围来看，不仅互联网巨头纷纷开出相关部门，初创企业的数量也十分可观，他们常常作为各单位的可信第三方存在。比如，为全球顶尖制药公司提供隐私计算平台的初创企业Owkin，实际上就充当了链接起医院、科研机构和制药公司的中间节点。

“隐私计算是一种面向应用的技术，需要繁荣强大的数字产业作为支撑。在这方面，中国有强大的数字支付、发展充分的互联网应用，保障数字权益的法律法规也相继出台，理应在隐私计算领域走在前列。”在袁晨看来，隐私计算在国内的蓬勃兴起，正是中国数字经济发展的生动写照。

“上海发展隐私计算拥有得天独厚的条件。”光之树科技创始人张佳辰说，这也是公司将总部设在张江的原因。去年底，上海数据交易所揭牌成立，光之树科技成为该交易所的首批技术服务方。在她看来，上海集聚了大量半导体企业和互联网产业，还有大量金融机构和生物医药企业，在隐私计算的“护航”下，将催生出更多新业态、新模式，让数字经济跑得更快、更稳。

专家视点

隐私计算如何解决“百万富翁难题”

杨光

根据数据和计算任务是否集中，目前隐私计算可分为三个主要方向和多种技术路线。一是安全多方计算，这是针对数据和计算都不集中的情况，主要技术路线有混淆电路和秘密共享两种；二是数据不集中、计算集中，这个方向的主要技术有数据脱敏、差分隐私保护、同态加密等；三是数据和计算都集中，这个方向包括可信执行环境和数据沙箱等。

安全多方计算这个研究方向起源于图灵奖得主、中科院院士姚期智于1982年提出的“百万富翁问题”：两个百万富翁相互不服气，想比较一下到底谁更有钱，但是比较的过程还不想泄露家底，他们该如何操作呢？

在这道“烧脑题”中，如果有一个可信的第三方，问题就能迎刃而解，安全多方计算研究的就是如何以密码学技术代替这样一个可信的第三方，在保护参与者隐私的同时完成计算任务。

姚期智先生在提出并解决“百万富翁问题”后，又于1986年提出了混淆电路技术。混淆电路在计算过程中始终处于加密状态，不泄露参与计算双方的任何私有信息，但能计算出正确的结果，在理论上解决了两方参与的隐私计算问题。

秘密共享方案则以秘密共享的方式，将每根线上的值共享给所有参与者，每个参与者都被分配了一份秘密份额，只有将足够多份不同的秘密份额组合在一起，才能恢复出完整的信息。

2008年，丹麦部署了世界上首个实际应用的安全多方计算系统，由丹麦唯一的甜菜处理商丹尼斯克、丹麦甜菜种植者协会、丹麦政府下属机构SIMAP等三方共同计算来年的甜菜种植计划和收购价格。

随着近年来技术方案（特别是布尔电路编译技术）的迭代优化，以及通信基础设施的快速发展，采用秘密共享方案的安全多方计算平台变得越来越实用，距离大规模应用仅一步之遥，被视为未来隐私计算最有潜力的发展方向。

第二条技术路径是“数据不集中、计算集中”，其核心思想是对数据进行变形、扰动、加密等操作，保障无法从流出的信息中恢复出原始数据。具体的技术主要有三种：数据脱敏、差分隐私、同态加密。

如果说数据脱敏是通过对敏感信息“做减法”的方式实现隐私保护，那么差分隐私就是以“掺沙子”的方式，在数据或计算结果上添加一定强度的噪声，保证传出的信息不能精确反映用户的隐私信息。比如在统计平均身高的场景下，每个人在提交身高数据之前，加上一个随机的误差，根据统计学即可估计随机误差对计算结果的影响，最终校正后的结果仍能较好反映实际的平均身高。

数据脱敏和差分隐私的技术方案非常简单，已被苹果、谷歌等公司用于收集用户使用情况的统计数据。但是这两种方案都会降低数据的质量，因而其应用范围很受限制，通常只用于统计类的计算任务。

同态加密则是用技术方式，在不影响数据运算结果的前提下，将数据变为密文，然后在密文上进行运算，最终的计算结果对应于先在明文上进行相同计算后再加密所得的结果。因为计算时看不到数据的明文，所以不会泄露隐私。

第三个方向“数据和计算都集中”，其核心思想是通过工程手段构建一个可信的计算平台，将其作为一个“可信的第三方”来使用。具体来说，就是通过隔离机制构建出一个安全可控的区域，数据在这个足够安全的空间中被集中使用且不流出，从而保证数据的隐私性和计算结果的正确性。

这一方向主要的技术方案是可信执行环境（TEE）。该技术通过软硬件隔离安全机制建立一个安全隔离的执行环境，从而防止外部攻击者（包括系统管理员）窃取TEE内部运行的数据。

TEE具备支持多层次、高复杂度的算法逻辑实现，具有运算效率高、可信度量保证运行逻辑可信等特点，是目前效率最高的隐私计算解决方案。然而，由于TEE依赖于CPU等硬件来实现，必须确保芯片厂商可信，其安全性存在一定的单点风险。

（作者系上海树图区块链研究院研究总监）

人工智能新药研发探索数据共享模式

本版图片均来源于视觉中国

■本报记者 **沈湫莎**

新药研发领域有一个广为人知的“双十定律”：研发一款新药平均需要花费10年时间、投入10亿美元。有没有更高效的办法？随着人工智能的发展进步，数据对于药物设计、发现、临床验证等各个环节的重要性日益凸显。

不过与全行业相比，任何一家制药企业、研发机构所拥有的数据量都十分有限，出于竞争的原因，他们只能依靠各自的数据进行药物研发。数据量已经成为影响人工智能药物研发成败的关键因素之一。

日前，一个囊括了全球10家顶尖制药企业、由17家合作伙伴共同参与的数据平台宣布成立。在该平台上，相互之间存在竞争关系的企业能够在不暴露他们所拥有数据的前提下获得计算结果。这一人工智能新药研发项目，旨在打破不同主体间的“数据孤岛”，探索一条数据共享的全新模式——利用多家制药企业的数据，创建更准确的模型，为药物开发筛选最有效的化合物。

该项目汇集了安进、阿斯利康、拜耳、勃林格殷格翰、葛兰素史克、杨森制药、默克、诺华等10家全球顶尖制药企业，合作伙伴包括一家人工智能企业、四家初创企业和两所欧洲大学。其中，隐私计算公司Owkin提供了基于区块链的数据模型平台，中央调度程序允许每家制药企业共享同一个联邦学习模型。简单来说，制药企业在不泄露自身数据的前提下，能够调用更多数据在模型中进行计算，海量数据大大提高了模型预测的准确性和适用性。

在此之前，Owkin在《自然·医学》杂志上发表过题为《基于深度学习的间皮瘤分类可改善对患者结局的预测》的论文。论文分析了近3000名患者的数字活检图像，数据来自多家法国机构。此次，10家制药企业承诺，在项目的安全性和隐私保护得到证明之后，他们将投入前所未有的海量数据。可以预见，如果该数据共享模式被证明可行，那么更多药物研究相关的数据持有者将通过该模式共享数据，“数据孤岛”将不复存在。

除了药物研发，隐私计算在城市治理中也将发挥越来越大的作用。去年，郑州遭遇百年一遇特大暴雨，此后各地政府都对治理内涝积水十分重视。如何提高内涝预测的准确性，把灾后响应转为灾前预警？首批入驻上海数据交易所的隐私计算领军企业光之树创始人张佳辰表示，用过去的计算方式，一些敏感却实用的数据无法用到，比如自来水管网的数据，但通过隐私计算技术，它可以和地形、地理、天文、水文等与气象相关的数据一同汇总，训练出精度更高的预测模型，并集成到整个水务系统的应急管理解决方案中。

在武汉东湖高新技术开发区，上海区块链企业零幺宇宙正在参与打造国内最大规模的“数字底座”，其中也有隐私计算的身影。与传统数字城市相比，利用区块链和隐私计算技术构建的“数字底座”拥有自学习能力，当某节点的算法进化，可在边缘侧节点之间自学习推广，一体化的系统设计为后续场景的加入预留了接口。

以往，不同部门间的数据格式和结构各不相同，交互起来很不方便。零幺宇宙（上海）科技有限公司执行总裁上官芸说，从“条块式”升级到“底座式”的数字片区建设，就是要打破过去各部门数据不相通的情况，实现数据的统一规划和接口标准。而要让各部门放心将数据共享，必然要用到隐私计算。据悉，“底座式”的数字片区将大大降低数字系统的运行成本。据测算，完成相同任务至少可减少三分之一的设备投入。

此外，在此次“数字底座”建设中，基于区块链和隐私计算的“碳账本”技术得到了初步应用。上官芸介绍，“城市零碳”方案分为两步：一是记录，二是预测。区块链技术可以从采集源头记录数据信息、计算存证，并通过模型预测未来将产生的碳排放量。同时，并不是所有数据信息都需要披露，基于隐私计算可只显示最终排放结果，为节能减排提供量化行动方案。