隐私计算崛起

编者按/ 数据正在成为信息时代的“石油”。

随着《个人信息保护法》的颁布实施，中国个人信息和数据使用的野蛮生长时代正在面临终结，如何能够合规利用中国庞大的数据资源，关乎数据能否发挥生产要素职能的关键所在。

一套发轫于上世纪80年代的方法，由此进入中国数据使用者们的视野，那便是隐私计算，这套综合了多种学科方法解决方案，如今在中国似乎找到了更为现实的应用场景。而数据的所有权与使用权分离，也由此开始。

资本敏锐地嗅到了这样的味道。但是，隐私计算的赛道究竟有多大，距离真正崛起还有多远，这仍然是一个等待回答的问题。

A5~A6

本报资料室/图

一线调查

隐私计算崛起还有多远？

本报记者 裴昱 北京报道

如何在不暴露和不泄露各自财产的情况下，准确地比较两个百万富翁谁更加富有，即“百万富翁命题”。这一看似无厘头的命题实际上用到了现代密码学的重要内容。

世纪之交的2000年，华裔科学家姚期智凭借这一命题研究获得了“计算机领域的诺贝尔奖”——图灵奖，为日后隐私计算的应用打下了理论基础。

所谓隐私计算，即在保护数据本身不对外泄露的前提下实现数据分析计算的技术集合，达到对数据“可用、不可见”的目的，从而实现数据价值的转化和释放。隐私计算是一类既能保护用户隐私，又能实现数据计算效果的技术的统称。

时值壮年的姚期智那时恐怕不会想到，在他完成“百万富翁命题”近40年后，他的祖国已经成为全球第二大经济体和数字经济最为发达的国度。而他的模型和思想，如今在金融、政务、医疗等场景中得以商业化应用，并形成独特的赛道。

当下，恰逢中国开始进入对数据安全和个人信息安全严格监管的时代。当海量的、已经“升级”为生产要素的数据资源，面对严格的个人信息监管时，发轫于姚期智的隐私计算技术终于有了用武之地。

同时，隐私计算不仅是商业和产业层面的问题，在个人信息保护、数据安全与数据应用之间寻求平衡，以数据优势持续构建中国的国家竞争力，同样是事关国运的重大命题。在这条赛道上，既有互联网巨头公司竞争，也有长期从事大数据业务的企业参与，还有初创企业的加速入局，资本正在识别“隐私计算”赛道的长短与容量，投资人跃跃欲试。

一位天使投资机构的投资人对《中国经营报》记者表示，中国庞大的市场和数据使用场景，总是能够提供更多的可能性，现在一切仍未有明确的答案，投资人正在探索隐私计算是否能够成为一个独立的赛道。“不试一试，怎么知道呢？”他说。

何为隐私计算？

简单来说，隐私计算是一类既能保护用户隐私，又能实现数据计算效果的技术的统称。

近日，招商银行“慧点隐私计算平台互联互通项目”中标结果公布，这个名字看起来有点拗口的项目，实际上是国内首个由大型股份制商业银行牵头，与多家隐私计算服务商共同合作的跨平台互联互通项目。项目中标方之一的同盾科技公司相关人士告诉记者，该项目将协助招商银行完成企业级隐私计算互联互通标准的制定。

隐私计算，这个对大众来说有些陌生的概念，正在变得越来越重要，金融业对其更是颇为热衷。在招商银行的项目之前，浦发银行联合腾讯云计算公司等机构，启动了“多方数据学习‘政融通’在线融资项目”；交通银行则联合中国银联、华控清交启动了监管沙盒项目；中国工商银行、中国农业银行也不同程度地在相关业务中尝试性地应用隐私计算工具。

在知名咨询公司Gartner对重要战略科技趋势的预测中，“隐私计算”技术连续两年入围。Gartner还预测，到2025年，将有一半的大型企业机构使用隐私计算在不受信任的环境和多方数据分析用例中处理数据。那么，隐私计算究竟为何物？

简单来说，隐私计算是一类既能保护用户隐私，又能实现数据计算效果的技术的统称。

“百万富翁命题”中，富翁拥有的财富就是数据所有权，富翁公布财富数据就是数据使用权，是否有一种技术，能让他们向这个技术平台透露财富数据，经过一系列加密数据的计算，最终得出（谁更富有的）结果，而他们的财富数据本身不会泄露。对于需要数据的企业来说，他们获得的不是原始数据，而是经过加密的数据，以此来为计算结果提供服务。理解了这一假说，就能理解隐私计算的大概思路。

姚期智早在上世纪80年代提出的这一命题，被视为隐私计算的雏形。但是，在他提出这套算法的时代，由于计算机的算力有限，实际应用耗时颇长，在当时并未广泛应用。

目前，隐私计算包括多个“流派”。据记者了解，姚期智论文中提出的方法，在隐私计算领域中被称为“多方安全计算”，更多依托于密码学，是将一组相互不信任或者不信任任何第三方的独立数据所有者，通过一个函数得出准确的运算结果，同时各方输入的数据信息不会暴露且不可还原。

“用更通俗的话说就是，多方安全计算，是指每个人算一点，然后合在一起能够完成一个任务，它可以做一些相对简单的数据查询、统计等计算。”同盾科技合伙人兼人工智能研究院院长李晓林说。

除了多方安全计算，隐私计算还有“联邦学习”和“可信执行环境”两大流派。

“联邦学习”由国际互联网巨头谷歌提出，是一种分布式的“机器学习”，即多个参与方事先商定好分析模型，在数据不出各方“本地”的情况下，用各方数据对模型进行训练，而后得出结论供各方使用。“这一流派是从机器学习的角度出发，比多方安全计算又走近了一步，更适合目前人工智能的需求，模型训练和学习需要利用更丰富的多方数据，因此需要数据隐私的保护。”李晓林说。

无论是多方安全计算，还是联邦学习，都是软件层面的路径，而“可信执行环境”这个流派，则加进了硬件的因素。这个流派的核心思路，是通过构建一个独立于各方，且受各方认可的安全硬件环境，在安全、机密的空间内进行计算，得出结论。

“通过硬件打造出来的一个安全屋，在这个安全屋里的所有操作是受保护的。多方的数据放入这个可执行环境中，然后在执行环节里面得出结果。”李晓林说。

目前，这三种技术流派在隐私计算的商业层面都有应用。“具体应用并非泾渭分明，而是从适用的角度出发，相互融合。这不是华山论剑，一定要比出哪个门派是高是低，单纯的、脱离市场来争论哪条路线的技术水平的高低，也没有太大意义。”蓝象智联CEO徐敏向记者表示，蓝象智联是一家隐私计算服务商，主要为金融机构提供服务。

应用场景何在？

在防范化解金融风险和金融支持实体经济的过程中，隐私计算也在发挥部分作用。

一项技术的伟大之处，在于其解决了什么样的实际问题，没有产业场景的应用，技术也只是“英雄无用武之地”，隐私计算也不例外。用技术创新寻求突破并非难事，但如何找到产业场景的沉淀是摆在隐私计算企业面前的重要问题。

隐私计算如何推动数字经济的发展？蓝象智联创始人兼董事长童玲对记者表示，隐私计算本质上并不算一个行业，而是一个基础技术，可能和各行各业发生关联。（在不同的应用场景下）利用隐私计算解决数据安全流动的问题，并进一步挖掘数据价值。“在我们的团队中，除技术团队外，还有一个数据运营团队，这个团队的人既要懂技术，还要懂行业。”童玲说。

当海量用户数据诞生后，已经开始指导各种产品的设计生产，在药物研发、风险评估等领域需要大量的交叉数据作为参考依据。数据价值也从消费侧向生产侧转变，在这一背景下，隐私计算为安全合规地使用交叉数据提供了一种可能。如今，在防范化解金融风险和金融支持实体经济的过程中，隐私计算也在发挥部分作用。

彭凯是金诚同达律师事务所高级合伙人，帮多家跨国企业和大型集团公司处理数据合规业务，他告诉记者，目前，即使是在一些大型企业和集团公司中，他们的信息系统基础设施（在安全防控方面）并没有我们想象得好。这就带来一个问题，在数据使用、传输过程中，被外界攻击怎么办？数据泄露怎么办？甚至还有一些更为原始的数据传输方式，比如拿一个U盘拷贝了数据，U盘丢了怎么办？

“从这个角度看，隐私计算为数据商业化应用提供了一个技术层面的解决方案。”彭凯说，“因为数据‘不可见’，能在一定程度上规避数据传输过程中的攻击和泄露风险。”

目前，监管部门在反洗钱和反诈骗方面对银行和金融机构有较高的要求，隐私计算在金融机构的一部分应用市场即来自于此。除此之外，贷款发放的风险审核也是当下隐私计算在金融机构的另一个应用场景。

对于金融机构而言，运用隐私计算服务的优势，在于可以使用多方数据对贷款申请者的资信、还款能力等风险因素进行评估，从而实现“精准风控”的目的。徐敏告诉记者，目前较为主流的方式，是银行自身数据与运营商数据、政府公共部门数据综合运用，通过隐私计算，就能够得出较为精准的贷款申请者风险评估，并进行分类。

不过，法律和监管要求电信运营商不得对外提供涉及敏感个人信息的数据，这就为隐私计算提供了重要的应用场景。在得到授权的情况下，通过建立模型，可以保证各方数据不出本地亦不可见的情况下，通过模型运算得出金融机构所需的评级，作为风控参考。“银行对获得敏感个人信息并不感兴趣，其所需要的，只是能够精准评估风险的结果。”一位大型股份制银行风控部门人士说。

徐敏表示，总在深夜打电话，或者是频繁接听国际长途，这些情况发生在一个人身上，可能只是特殊的个例，但当数据量足够大之后，就会发现这种情况和电信诈骗、洗钱等非法金融交易，有一定的正相关性，通过这个逻辑，就可以进行风险评估。“不拿出、不显示具体的数据，但模型最后会根据计算结果，给出一个风险较高的评级，这对金融机构是很有用的。”

隐私计算之所以能够实现保护信息安全的目的，是因为其中存在一项“同态加密”技术。李晓林告诉记者，不同的数据所有方，按照一个相同的方式将数据加密，然后再由模型程序进行计算，计算结果解密后和明文（未加密的数据）计算结果一致。利用这项技术，可以实现让解密方只能获知最后结果，而无法获得每一个数据，这对保护信息安全有重要意义。

隐私计算的另一个应用场景，在大型金控集团。这类企业可能拥有多张金融牌照、开展多项业务、（不同板块、公司）各自累积了海量数据，但在监管要求和内部竞争格局的合规约束和制约下，数据在集团内部也不能随意流动。

彭凯举例称，一家大型集团公司，旗下拥有地产、金融、电商等多板块业务，如果集团公司想整合不同板块的数据资源，或者把地产板块的客户信息用于电商业务的精准营销，根据《个人信息保护法》需要取得客户单独同意的、履行个人信息保护影响评估等法定义务。“这样的法定要求在实践中是很难实现的，同时也对大型企业、金融机构的业务流程、合规要求、系统建设提出了更高要求。”彭凯说。

“未来拼的是数据资产能力，数据融合是绕不开的问题。利用隐私计算的方式，可以在合法合规的情况下，得出指标性的评估结果，对于各自（条线）的业务开展，都是很有用的。”一位全牌照金融机构的人士表示，隐私计算可以解决数据跨机构互联互通的难题。

数据从何而来？

从应用层面而言，目前隐私计算重点对接的数据来源包括三个：银行数据、运营商数据、电网数据。

在隐私计算业内有一句“谚语”，“心有余而数不足”，用来形容缺乏数据的业务运营就像“空中楼阁”，数据从何而来？特别是在《数据安全法》《个人信息保护法》等法规出台后，企业如何安全合规地利用数据来支撑业务运营，隐私计算为其提供了一种技术手段。

徐敏告诉记者，从应用层面而言，目前隐私计算重点对接的数据来源包括三个：银行数据、运营商数据、电网数据。形成这一结构的原因在于，这类数据规范性较强，标准化程度较高，合规运用的效率较高。

“运营商数据很丰富，用户通话信息、上网行为信息和位置信息等，运营商都有，这些数据对于评估分配一个人的资信情况很有价值，这些数据不能对外提供，但是通过隐私计算的方式，可以综合这些因素，再和银行的数据一起形成评级，来标定一个主体的风险等级。”徐敏说。

此外，电网机构也是重要的数据来源。一位隐私计算公司的技术人员告诉记者，企业用电负荷、缴纳电费等电网数据，可以反映企业的经营状况，但只有这一数据的价值非常有限。如果结合工商登记信息、司法履约信息、对外投资信息等外部数据，构建科学的算法模型，就可以形成企业的评级或征信信息，等于挖掘出了更多的数据价值。

这个思路已经用于实践，同盾科技的隐私计算服务中所使用的数据源，便包括部分电网数据。李晓林告诉记者，他们采用的数据处理方式叫作知识联邦模式，这种模式不同于使用本地的海量数据进行机器学习、模型训练，而是利用多方的部分数据来提炼知识达到训练机器学习的作用，从而撬动更大的数据资源来共创和共享知识。

据李晓林介绍，知识联邦模式是一个多层级的模型系统。可以先在内网层面操作，比如在一定范围内交换不同的电网数据，把这部分数据加密处理，进行模型训练，用于训练的数据不出“本地”，在模型的科学性和应用性达标后，用于训练的数据将被清除，而这个模型将用于电网数据和其他外部数据融合。“这种模式更安全，且不能反推出原始数据，已经被用于金融反欺诈、企业融资、营销、疫情预报等场景。”他说。

而拥有数据资源或数据使用权的一方，也有开发数据价值的诉求，换言之，就是他们也需要和外部数据互联互通，推动多数据资源整合。一位运营商人士表示，他们有专门的部门和团队开展合规数据经营业务。现在，隐私计算的服务商，也是他们的客户。在与客户的合作中，运营商遵循两个原则：其一，不外泄个人信息，具体包括电话号码、APP装机种类、开机时长等；其二，相关信息不能对应个人姓名、电话，且信息不能出“本地”。

但是，他拒绝向记者透露这种合作具体的费用标准。

记者了解到，国家电网也成立了商用大数据、征信公司，拓展电力大数据的应用场景，打破各行业间的数据孤岛，推动数字经济融合发展，隐私计算将发挥更大作用。

“隐私计算有助于将数据的所有权和使用权分离，并创造商业价值，这在数据成为生产要素后，有着非常重大的现实意义和作用。”徐敏说。

赛道有多大？

隐私计算作为数字经济底层基础设施被看好，但目前落地场景有限，有观点认为其商业化应用的前景尚待验证。

牵涉到算法、模型、数据这些颇具“技术含量”的语汇，隐私计算给人“高深莫测”的感觉，技术是隐私计算立命根本，但这个赛道前景如何？商业化道路能走多远？应用空间也是不可忽视的问题。

就商业模式，徐敏告诉记者，隐私计算一般有两种模式：一种是收取隐私计算系统搭建的费用，这种模式下，收入是以“单”计算的，根据客户需求的不同，有不同的报价组合；另一种是收取运营的费用，即搭建系统完成后，按照客户使用查询的次数收费。这部分费用先由隐私计算服务供应商向客户收取，而后隐私计算服务的供应商，再向数据方支付使用数据的费用。

至于处理海量数据和查询服务的系统运算速度如何，是否会影响“客户体验”，徐敏表示，这其实不是太大的问题，他以银联数据和新网银行牵头、蓝象智联参与的多家金融机构间数据共享平台为例，目前每天大约有60多万笔跨机构隐私查询，每笔大概只需要100~200毫秒。

下转A6