数字强国③

Web3：数据确权算法透明，生产力大释放

杨光提出，元宇宙、AIGC带来技术迭代呼唤理顺利益分配的新型生产关系

9月16日下午，文汇讲堂“数字强国系列”第三期暨163-3期《Web3：下一代互联网的生产力与生产关系变革》在上报大厦43楼融媒创新空间举办。喜马拉雅·听和文汇报视频号进行了直播。

上海树图区块链研究院研究总监杨光博士主讲，上海交大计算机科学与工程学系郁昱教授应邀担任对话嘉宾。50位现场听友获赠NFT数字徽章。

本系列讲座由文汇报与上海树图区块链研究院联合主办。上报集团融媒创新空间运营团队提供技术支持。

整理：李念 金梦 版式：李洁 摄影：周文强

①

嘉宾主讲

相信很多人都想像过未来的互联网、数字世界会是什么形态。2023年6月，苹果公司发布了重磅产品——MR头显设备Vision Pro，业界普遍认为这个硬件将成为未来元宇宙的入口；至2023年上半年，AIGC（人工智能自动生成内容技术）不断“涌现”，催生出如Stable Diffusion、Mid-Journey、DALL-E等AI绘图软件，像Github Copilot等自动编码插件，它们预示着大模型带来的数字生产力飞速提升。

但随之而来的是，新的生产范式带来的版权争议直接限制了应用场景的推广。数字领域生产力发展亟待建立新的生产关系，此时以区块链技术做底层支撑的Web3显示出优势，就在这个层面上，我们今天来讨论Web3这个下一代互联网所昭示的新型生产关系。

生产关系受困：迭代生产力需要理顺收益分配

AI代码生成软件问世之后，很快被人质疑代码涉嫌抄袭，因为在某些任务上生成的代码和开源的代码非常雷同，使得该软件目前很难大规模商用。绘画领域，版权通常只保护画作的具体内容，无法保护绘画的风格。而AI学习人类绘画风格这件事显然是很难评判的，但这样的软件会因此遭抵制而无法商用。

提供数据、训练的人群如同原材料提供者

为何数字世界里的生产会有如此之多的版权问题？这就要从整个数字世界新的生产范式说起。以典型大模型生产流程为例，最开始要有原始数据，接着对原始数据有标注，继而应用各种算法对标注好的数据进行人工智能的学习、训练而生成模型，再把模型做成产品，最后通过给这个模型下达一些指令和一些引导，模型会生成我们想要的内容。比如命令Stable Diffusion生成某个主题的图片，这个图片或被用于训练下一轮的模型。

在整个生产流程中，谁受益最多？一是科技公司，他们负责做人工智能的算法研究，对外提供服务可以收费；二是直接使用这个模型的用户，模型产生价值用户愿为此付费。但其他参与者，例如提供数据的人、提供标注的人，他们就很难从大模型的盛筵中分享到合理的收益。

不妨把生产流程套用一下：数据和标注都是整个生产过程中的原材料，算法和模型属于生产中的工具或者机器的角色，调用模型的用户更像是生产活动的管理者，他想做什么产品，就会去操作机器产出目标对象。现实生产关系中，提供原材料的一方是有收益的。但在数字世界里，产出的结果对于原材料而言，并没有给他们一个很合理的收益分配，所以才会有诸多侵权质疑。

数亿人创造的数据集很难理出收益分配关系

为什么新的生产方式会有这些问题呢？这和数据与算法自身的特点有关。从技术上来说，大模型AIGC这种先进的生产力创造的价值和原材料、原始数据之间的因果关系较难量化。可以说整个数据集创造了整个价值，但具体到某一次使用，用了哪些数据，那些数据权重是多少，有没有因果关系，这件事很难说清。

即便有了量化关系，把收益分配给众多提供数据的参与者也很困难。现实世界中的原料供应商数量相对有限，即便大飞机如此复杂的产品，它的供应商总共也就是几万个的数量级。但大模型动辄运用几十亿、上百亿甚至上千亿数据，即使科技公司主观上愿意将大模型的收益公平分配给所有提供数据的人，但在Web2现有的技术框架上也很难实现，因为分配人数和收益额度无法量化。

在没有理顺收益分配关系的情况下，谈AI对于生产力的提升，一定会有人认为在整个生产链中受到了不公平对待。比如实际上提供原创内容的人，他们会认为自己被AI剽窃了。这些困境都和数字主权的缺失有关。

无数字主权易引发版权讹诈，“大数据杀熟”难以自证

目前，数字主权在很大程度上存在一些缺失。一方面，很多数据是由互联网平台公司所控制，导致数据存在泄露的可能风险，这会侵犯用户个人隐私权。

另一方面，在数据主权不确定时，会出现版权讹诈的风险。微博上曾有一热搜，某摄影师将自己拍摄的照片发布在自己的公众号上，某天收到律师函，称其侵权要支付版权费并赔偿。这件事中涉及到多层版权代理问题，关键是很难验证。

更有甚者是“大数据杀熟”。打车平台或外卖平台通常会给用户发推送并告知价格，平台算法会通过大数据对个人过往行为做分析，判断用户对价格是否敏感，然后让付费能力更强、对价格不敏感的人多支付，让对价格敏感的人少支付，这就存在对公平交易权的侵犯。

现在个人的数据都在平台公司，推荐算法也在黑箱里运行。推荐算法到底是怎样生成推荐结果？如何计算价格？这一过程并不是特别透明。即便技术公司想表明自己的算法属于完全公平，在目前Web2的技术框架内很难自证。

如何解决收益分配问题？一是数据确权，二是算法透明

既然问题出在数据和算法方面，必然要从此破解。

第一，在数据方面要做确权，必须先确定每一段数据属于谁，为所有者建立账户，才有可能记录和分配相应的利益。为了保证确权和收益分配的公正性，这样的账户和记账不应该由互联网平台公司单方面控制。

第二，算法需要有可理解的透明性。如果算法是一个黑盒子，就无法让他人完全信服。为此需要让整个算法实现透明性，并且模型要具有可解释性，最终才能保护所有参与者的知情权以及在参与中的平等地位。平等地位是指让参与者知道使用者利用数据做了什么事，产生了什么效果。

综上所述，如何建立新型生产关系？需要用到数字主权。数字主权指个人或者组织对数字身份、数据、算法的所有权、控制权和管理权。数字身份就是对数据做确权时要有一个确权主体或权益载体。数据是整个数字世界生产的原材料。算法是生产的过程。需要在身份、原材料、过程都有一个明确的权属和关系后，才能实现整个生产过程中利益的合理分配。

新生产力突围：区块链和第三代密码学发力

要解决数据与算法带来的问题，最终还是要靠技术的发展。2023年6月，上海市科委发布的《上海市“元宇宙”关键技术攻关行动方案（2023-2025）》中，明确沉浸式技术和Web3技术是两大主攻方向。前者是能够构建三维虚拟互联网空间的技术，后者是保护数字主权所需用到的技术。区块链技术是Web3的底层支撑，而区块链的理念类似于科幻小说《三体》提出的“透明思维”概念。即人与人之间完全公开透明，互相没有隐私且互相信任，这种非常高效的协同组合在一起形成一个所谓的“人列计算机”，就像计算机里的元件一样。

拜占庭容错共识：Don’t Trust, Verify!（不信为信）

那么地球人能否用这种思维去构建一个互相信任的系统呢？答案是“可以”，这就是所谓的共识机制，即让系统中的所有节点达成一致共识。传统的分布式系统，主要研究如何让同属于一个机构的机器保持一致，就像思维互相透明的三体人一样可以被信任。此时只需要考虑机器宕机的情况，也即故障容错（CFT）。

而现在这些机器从三体人变成了地球人，思维不再透明，即这些电脑被不同单位、不同组织所控制，上面运行什么程序，是否被修改过，已经无法完全信任。同时被修改过的程序也可向别人撒谎，这种情况下能否实现一个系统且达成共识，即形成所谓拜占庭容错的共识呢？

对此，区块链技术可以做到。其根本思想是，在看到他人告诉我的结果之后，要通过自己的验证才能相信。区块链里有一个说法“Don’t Trust, Verify!”，中文可译为“不信为信”，即我不相信其他人，亦不需要考虑这个人是谁，他说的事情必须经过我自己的逻辑、技术的方式验证通过才能相信，最终目的是希望达到个体对整个系统的高度信任。

哈希函数：如现实中的“骑缝章”保障数据不被篡改

区块链中有一个很重要的概念是抗碰撞的哈希函数。任何一个数据无论多长，无论是图片还是视频，经过运算可以得到一个固定长度的输出，这个输出就叫做哈希值，可把它理解为原始文件的数字指纹。每个文件的指纹都是不同的，就像现实中每个人的指纹不会相同一样。区块链的可验证性，最基础层面就是通过哈希函数、通过数字指纹来实现。

在区块链的数据库，它的结构和之前数据库最大的区别是，每一块数据要对前面的数据块做一个哈希运算，得到的指纹放在后一个数据块里，下一块数据又把前一个的指纹放进去，形成所谓的区块链。如果篡改了一小部分数据，局部变化会影响到后续所有的区块，它的指纹就会改变，为了维持所有数据块之间的链状关系、维持其合理性，所有区块都要一起修改。

对区块链的数据库来说，技术上并不是无法修改，只是无法暗自修改一小块数据。要修改就要从修改的当下数据之后的所有内容都做修改，这种改动很难逃过他人的观测。

哈希函数起到的作用就像现实中的骑缝章，在数据世界里，所有数据都可以盖上“骑缝章”以保证数据的不可篡改性。

第三代密码学技术：可保护数据的隐私性和正确性

数据本身的隐私性和正确性如何解决？这需要依靠第三代密码学技术。它就像监督员的角色，会对其看到的内容进行保密，只告诉别人你做的事情合规、正确且流程完整。

不同的技术实现功能也有差异，这里存在许多应用场景。例如，第三代密码学最早的技术被认为是安全多方计算，其源自姚期智先生于1982年提出的“百万富翁问题”，即两个有钱人想比富又不想透露具体财产，怎么办？姚先生表示，密码学的技术可以解决此事。这其实就是比较数字大小，在此基础上可以做出很多更复杂的应用。

比如可以设计一个用于选举的应用，每个人投票后对他人保密，但又可以保证最终得到的结果是根据投票的情况计算出来的。或像电子拍卖一样，每个人可以分别出价，按照拍卖程序计算，最后得出谁的出价最高谁获得拍品。但其他人并不知道别人的报价与最后的成交价。

零知识证明、同态加密技术、差分隐私技术各有应用

密码学技术上还有不少有趣的技术。比如零知识证明，指的是证明者能够在不向验证者提供任何有用的信息的情况下，使验证者相信某个论断是正确的。举例，证明者向他人证明这个数独是有解的；证明一个方程是否有解；一个数字签名是否是本人生成的。

同态加密技术是指先加密后计算和先计算后加密，最后得到的结果相同。在这个过程中，计算者看到的所有东西都是密文（加密后的信息），同时计算者又可以把很繁重的工作完成，最后会向他人证明，该结果确实是经过这些计算生成的，而非凭空编造的。简言之，让别人帮我完成许多计算任务，但别人不知道计算的内容是什么。

差分隐私技术，现在很多互联网公司已经在使用了。比如询问是否作弊过？这是无法撒谎但又属于敏感问题，他们只能不回答，但也能得到结论。

构建新的生产关系：过程正确、权益确认、信任可期

Web3是技术迭代带来的生产关系变革。

在Web3时代，最核心的基础设施是采用区块链技术构建的中立计算平台，它不受任何机构和组织所控制，由密码学技术保障处理信息的过程正确、合规，并且允许所有人公开验证。以这样的中立平台为基础，数字身份、数字权益等诸多数字主权可以在Web3的世界里实现。

凡是数字世界里需要信任处都可用Web3来解决处理。比如非金融领域中比较典型的场景就是NFT数字藏品，它在国内的应用比较丰富；现实世界有物流、安防监控和各种手段进行信息存证，未来发生纠纷时，数字世界的存证留痕会成为证据。典型案例是，在电商平台下单采购，电商平台如删掉了这笔订单，区块链存证则可为购买者找回公道。

信任问题还可通过区块链里的数字化契约来实现，例如控制资金池的风险。一般我们会把资金托管给有资质、可信赖的人，像保险、信托或者版权代理等可信的中介，这个中介出售的就是合规性和信任度。这种信任在一些比较简单的场合可以被Web3里的技术直接替代，这个过程公开透明，且成本和风险都很低。

更进一步说，你想用数据和别人做一些交换或者帮别人做一些算力交易，在目前的互联网比较困难，但在Web3平台上辅之以一些密码学技术，即有望解决。

尽管Web3发展也会面临很多挑战，诸如技术发展、法律合规，以及教育普及等，但其未来发展趋势就是让信任的程度变得更高、让信任的范围变得更广。相信，下一代互联网世界会变得更公平、更透明、更好地释放数字技术的新生产力。

现场提问

如何理解算法黑箱？衍生出“可解释人工智能”研究

东华大学传播学院教师徐敏：对于很多普通大众，算法仍然是一个黑箱，有无可能将算法这种复杂程序解读为大众可理解的东西？

杨光：目前有两个层面的算法黑箱。一个层面是它不告诉你运行什么算法，代码也不给你看，这是完全的黑箱。

如果打开告知你算法，但非专业人看不懂，在技术上不叫黑箱。另一个层面的技术黑箱是，现在做大模型、人工智能的神经网络技术人员，把代码、参数都告诉你，但他也解释不清楚为何是这样的。因此，现在人工智能有个新研究领域就叫可解释人工智能，不仅要作出决策还要作出解释。

郁昱：那些需要门槛但通过学习后能看懂的，属于可解释范围。但像深度学习、神经网络的分层，它的功能在实践中甚至可做预测能赚钱，但就是解释不清楚。一旦这些原理都可解释时，或可看得更远，设计出更好的算法。

算法如有利益属性就易成信息茧房，Web3可破解

五十四中学物理教师李世新：下一代互联网能否有效破解“信息茧房”困境，以实现更好的社会教育？

杨光：之所以存在“信息茧房”，很重要的原因是目前互联网推荐算法是为平台和公司利益所服务，目的是让用户在它的平台上停留更长时间。而Web3的算法具备中立性、透明性，打个比方，它更像一位忠臣，为一个美好的公共利益，会无私地推荐更相符的内容。换言之，Web3是能打破“信息茧房”的技术基础——算法掌握在谁手里，就会为谁的利益服务。

照片说明：

① 杨光主讲。

② 华东师大音乐学院小提琴专业研究生王天一现场演奏，杨光阐释NFT数字藏品原理。

③ 听友关注度很高。

④ 友情主持王刊博士介绍我国区块链公司的国际地位。

②


③

④