

算法治理的技术迷思与行动选择

文/邱泽奇

1984年是桌面计算机进入流行的初始阶段，也是人们对算法开始警觉的起点。人工智能的新一轮爆发引发了近些年对算法治理的关注，且形成了一股风潮，大多数涉及算法治理的文献甚至不谋而合地形成了一种套路式表述。只是对一些关键问题尚缺乏有效的回答，如，到底什么是算法？算法是一个旧已有之的存在，如今为什么要特别关注？如果关注，又能关注什么？本文将阐述算法已经成为人类生产和生活的普遍环境，分析算法的技术逻辑，说明关注过程治理、试图论证对算法进行实时监测难以实现算法治理目标，探讨算法治理的可能路径及中美欧的行动选择。

进入算法治理时代

数字技术的广泛应用与算法有什么关系？不用算法，行吗？即使运用算法，不治理，行吗？算法古已有之。为什么直到现在，人们才那么关注算法治理？

归纳既有文献，我们观察到三种主要理由。一种理由认为，算法侵害是触发人们关注算法治理的根由。人们认为，现实生产和生活中已经出现了算法歧视、算法偏见、算法垄断、算法操纵、算法剥削等与算法关联的社会、经济、甚至政治侵害。另一种理由认为，算法风险是触发算法治理的根由。人们认为，在个体层次，算法强化着信息茧房；在市场层次，算法酝酿着监控资本主义；在国家层次，算法被特定利益集团用于政治权力再生产。从个体层次到国家层次的风险给人类社会的政治、经济、社会生活带来巨大的不确定性。第三种理由认为，算法权力、算法规则是触发规则竞争、权力竞争甚至规则垄断、权力垄断的根由。值得特别注意的是，这也是西文文献探讨算法治理的主流。

我们认为，社会关注聚焦于算法治理其实是由两个因素推动的，除了算法影响的广度和深度，还有算法影响的覆盖面。那么，算法影响的覆盖面到底有多大呢？

在经济领域，数字技术快速迈过其扩散临界点，进入变革的起飞阶段。中国信通院的数据显示，2020年中国数字经济增加值为39.2万亿元

人民币，占GDP的比重跃升至38.6%，即数字经济已越过起飞临界点而进入起飞初始阶段。在社会领域，数字终端已经成为人们的必备工具。在中国，90%的人使用即时通信、80%的人使用政务服务、70%的人使用数字支付。在政务领域，审批等行政许可事项实现网上受理和“最多跑一次”的比例达82.13%，政府事项网上可办率超过了90%。哪一项都依赖算法。

任何一项终端应用的背后都有算法，算法成为人类新的生产工具、生活工具。我们甚至可以认为：无算法，不生产；无算法，不生活。人类正加速进入算法时代，我们无法回避算法给人类带来的本质性影响。

算法治理的技术迷思

算法到底是什么？在不同历史阶段、不同应用场景，算法有不同所指。可在本质上，算法只是一种求解逻辑。

第一，算法是面对问题的。让出租车尽快到达目的地是需要解决的现实问题，如何把某个商品卖给愿意出高价的人是需要解决的理论和现实问题，人们为什么在搜索引擎上搜索感冒药也是需要解决的理论和现实问题。人类需要解决的问题无处不在，故算法从不缺乏需要解决的问题。

第二，算法是通过计算解决问题的。任何用到算法的问题都需要将问题数学化为可计算模型，其中又内含着两面。一方面，要解决的问题可以被数学化即用数学模型刻画，不能建模或暂时不能建模的问题是运用算法的。另一方面，建构的数学模型是可计算的。满足可计算的条件是，至少有用于计算的数据，至少计算可以获得解。通俗地说，运用计算机不是为了计算，而是为了运用计算机的算力更有效地获得计算的解。

第三，算法是一系列问题和求解数学模型的集合。从一个人的算法到一座城市的算法，在数量规模扩大的同时，算法也从一个数学模型变成了无数个数学模型，且无数个数学模型之间不是独立的，相互之间不仅有关系，而且关系还异常复杂，模型之间不仅有结构，还有时序，结构与

时序之间还不是确定的，且算法还可以算法为数据而改进算法。同时，算法还与人类行为或事物的发展变化交互。那么，如何让用于解决问题的模型不变成进一步需要解决的问题，而真正解决人类问题呢？算法对代码的需要出现了。

第四，算法是由代码连接且结构化的一系列问题和求解数学模型的集合。只是，再复杂的数学模型都需要将其转化为向计算机传达的简单指令。复杂的简单也因此变成了复杂的复杂的一部分，计算机代码便成了算法的有机组成部分。在算法代码化的进程中，代码（code）的确是指示计算机做或不做的规则。表面上，正是由代码携带的简单指令构成的复杂算法影响着计算结果的输出，带来了文献关注的算法侵害、算法风险、算法权力、算法规则。可是，一旦我们进一步了解了代码与算法的关系，或许会有不一样的观点。

算法治理的行动选择

如果说数字连接、数据丰盈、人类必备是算法带来影响的三项必要条件，问题求解对能力的要求超出人类生物属性的极限是算法影响的充分条件；那么，在人类跨入数字时代、中国进入数字发展加速期的当下，算法又如何产生影响呢？

既有文献对算法影响社会的机制认识大致可以归结为两个本质性的争议。第一，算法是中立的还是有偏见的？斯廷森（Catherine Stinson）指出，人们以为算法是中立的，其实算法自身也会产生偏见，算法在不断迭代中从用户对算法推荐的响应中进行学习，创建着针对用户的选择性偏差，即偏见。这些偏见不仅具有统计学意义，也会带来不同类型不同程度的算法侵害后果。第二，算法是人类意志的产物还是自主自为的主体？认同“算法即规则”观点的研究者认为，算法已经作为独立行为主体参与人类生产、生活，算法规则也外化为社会行为规则，对人类产生深度影响。有人甚至认为，算法的自主自为性形成了算法利维坦，成为建构人类环境的重要乃至决定性的力量。

在探讨算法治理行动之前，我们认为有必要针对算法影响机制的争议进行澄清和回应。一项对谷歌图片算法的研究显示，谷歌图片算法再现了白人男性父权制结构。有趣的是，进一步的工作还显示，经过再培训的算法呈现了对社会文化实践的刻画，改善了之前呈现的社会歧视。也就是说，如果真实地拟合社会现实，那么算法会携带偏见，进而制造算法侵害。对算法再培训的结果

则说明：第一，算法侵害实际是编写算法的人群制造的，算法只是做了算法可做的事，算法依然是人类生产的产品；第二，算法并不具有自主自为性，没有人类输入的第一行代码，便不会有算法的其他代码，算法的自主性自为性便失去了载体，也根本没有机会制造算法偏见；第三，支持算法偏见和规则影响力的不是算法本身，而是对算法的运用。

如此，算法治理的方向选择已然出现。第一，算法不具有自由意志，算法是人类行动的产出。算法治理是针对人类生产算法行动的治理。第二，算法治理既是针对制造算法偏见的治理，也是针对算法运用的治理。前者是为算法纠偏，后者则是约束算法运用的方向和范围，以期最大限度地减少算法偏见带来的算法侵害。

在中文语境里，算法治理有双重含义：第一，用算法进行治理（governance by algorithm）；第二，对算法进行治理（governance to algorithm）。英文语境则没有算法治理（algorithm governance）概念，也很难找到一个对应的概念，维基百科中甚至没有相应词条。纵观中文文献对算法治理的探讨，大多探讨的是第二重含义。

算法治理发端于美国，美国如何应对算法侵害呢？直接的答案是，问责模式，问责对象是有潜力让算法产生广泛且深刻影响的算法生产者 and 使用者，问责内容包括算法内容和非算法流程，问责的主体是政府或第三方。

从2017年1月12日，美国计算机协会下属的美国公共政策委员会（USACM）发布《算法透明度和责任声明》，到2021年5月美国参议院引入《算法正义与在线平台透明度法案》，美国把因算法偏见带来的不良社会后果直接归责到算法生产者和使用者，而没有归咎于算法自身。

在数字时代的三方国际力量中，欧盟是重要一方。欧盟没有针对算法的专门法律，可针对数据保护的一系列法律在实践中也直接约束了算法的生产和运用。从1995年颁布的《数据保护指令》、2018年颁布的《通用数据保护条例》（GDPR），到2021年提案的《人工智能法案》，建立了基于算法风险的算法治理改进路径，再次强调算法生产者和使用者的责任，强调问责机制。

欧盟虽然与美国一样在形成算法治理的逻辑闭环，却在实践中从数据保护入手，建立了保护个体、组织不受算法侵害的防御性治理框架和法律制度，在算法治理理念上已向美国靠拢，即采用政府和第三方问责模式治理算法可能带来的侵害，且将算法侵害责任归咎于算法生产者和使用者。

为什么美国的算法治理要采用问责模式？为什么欧盟会向美国靠拢也采用问责模式？简单的回答是，算法的复杂性使算法自身始终处于迭代之中，在生产和生活中运行的算法不是一个确定现象，而是一个动态。算法的实时迭代意味着对算法代码进行实时督导（monitoring）和审查（auditing）需要具备与算法生产和使用相当或超越的技术能力。这意味着在技术上不可行；或即使可行，至少也是艰难的和代价高昂的。对超出能力范围的技术过程进行监管，需要跳出技术和技术过程本身，寻求算法治理的关键节点和效率节点。

算法治理的社会特征在于过程治理的不可行性或代价高昂以及结果治理的简单易行，是故，算法治理的关键节点便已不言而喻，那就是，针对算法结果进行治理。

算法问责便是对结果的治理，也显然采用了删繁就简原则。第一，在法律上约定算法可追责性。要求算法透明、算法解释、数据证据化等，正是为结果可追责提供事实前提和法理逻辑前提。第二，在规则上约定算法责任实体性。尽管算法即规则，可算法规则只适用于计算过程，不适用于外化的社会过程，算法生产者和使用者才是算法的责任主体，即算法的责任主体是人或人的集合体，不是算法自身。

作为数字时代三方国际力量之一的中国，又在如何进行算法治理呢？从1994年制定第一部相关行政指令《计算机信息系统安全保护条例》到目前，中国共出台了60多部相关法律法规和行政指令。从时间顺序来看，可以认为中国的算法治理是从总体安全出发的。只是到此为止，中国的算法治理只有目标，相关法律法规尚没有像欧美那样形成从治理理念到判例实践的闭环，其中，尤其是还看不出算法治理的法治理理念，更没完成算法治理的司法证据化约定。

如果说算法侵害是算法治理面对的核心问题，那么，在给定算法侵害人因化或组织化，给定算法技术处于不断迭代且动态演化的场景下，实施算法治理的行动选择其实不多。

第一，鉴于算法的技术复杂性和动态性，约定算法透明是让算法可以作为算法审查和追责的前提，却不是实施算法监测的理由。算法的技术迷思已经宣告了过程治理不是算法治理的可选项。可是，针对过程的治理要求却是其他治理选项的必备条件。

第二，如果把算法安全风险作为算法侵害的一个类型，则结果治理即使不是算法治理的唯一

选择，也是有效选择。算法的无处不在意味着算法风险的时时刻刻，优化的行动选择是治理出现的侵害并将其作为未来可能的风险源，而不是治理出现概率极小的预设风险。

第三，算法侵害的人因化意味着治理行动的对象是算法的生产者和使用者，而不是算法。在通用人工智能可以生产算法之前，算法依然是人类的产品。算法治理的关键是约束算法生产者写对第一行代码，约束算法使用者用对第一行代码，问责模式至少是当下可行且有效的治理模式。

第四，鉴于算法的生产者和使用者以及对人类构成广泛且深刻影响的主要是平台组织，把算法治理纳入平台治理、作为平台治理的一项内容，在形成成熟的算法治理进程中至少可以作为一项过渡性的行动选择。

结论

伴随数字连接迈向人类整体连接的是各类数字终端包括移动终端和传感器的时时刻刻无处不在，如此广泛且深刻的数字连接因连接的“节点—中心”结构而形成的是数据的汇流与应用的分发模式。在汇流与分发之间，算法成为运用数据资源促进人类发展和平等的关键所在。

算法在带来数字红利的同时，也在带来算法侵害。人们关注算法治理也源于人们试图阻止算法侵害的普遍发生。算法侵害源于算法偏见。可是算法偏见不是算法自己的偏见，而是人类偏见在算法中的再现和放大。阻止乃至防止算法偏见还需首先纠正人类的偏见。

针对算法偏见来纠正人类偏见有两个基本路径，一是针对偏见发生的过程，二是针对偏见带来的后果。在算法治理中，前者是过程治理，后者是结果治理。算法的技术迷思决定了过程治理是低效率的和高成本的。针对算法偏见带来的结果进行治理在理论上是更加有效的行动选择。

在数字时代的国际三方中，美国的算法治理采用了针对算法侵害的事后问责模式，也形成了从算法治理法理到实践的逻辑闭环。欧盟的算法治理从保护数据的守势入手，也在迈向美国的算法问责模式。中国的算法治理以网络安全为起点，以总体安全为目标，尚未呈现清晰的法理，在实践上还没有形成自洽的算法治理逻辑。我们认为，针对算法偏见的事后问责或许是殊途同归的有效选择。§

〔作者系北京大学博雅特聘教授、社会学系教授；摘自《人民论坛·学术前沿》2022年第10期〕