主题研讨

算法透明原则的迷思

——算法规制理论的批判

沈伟伟

内容提要:随着近年来算法问题的大量出现,人们开始思考如何规制算法。算法透明原则是学理和实践中众所周知的一项算法规制原则,许多学者对算法透明原则十分推崇。但与事后规制相比,算法透明原则作为一种事前规制方式,其规制效力有着天然的缺陷。即使算法透明原则可被用来限制"算法黑箱"的不利后果,但在大规模通过立法、行政、司法措施规制算法的时代,算法透明原则通常既不可行,也无必要。因此,就算法透明原则在算法规制谱系中的合理定位而言,其应该处于非普适性、辅助性的位置。比起本质主义色彩浓厚、以算法透明为代表的事前规制,以实用主义为导向、以算法问责为代表的事后规制是更加得当的规制策略。

关键词:算法规制　算法透明原则　事前规制　事后规制　本质主义　实用主义

沈伟伟,中国政法大学法学院副教授。

环球法律评论 GLOBAL LAW REVIEW

引　言

近半个世纪以来,算法〔1〕正以前所未有的深度和广度,影响和改变着人类活动。依托这一技术革命情境,并伴随着网络空间和现实空间的加速融合,算法应用越来越广泛。

〔1〕"算法社会""算法时代""算法世界"等指示日常生活与算法紧密关联的新词汇,已逐渐普及。比如2016年美国皮尤研究中心就用"算法时代"(Algorithm Age)一词。参见 Lee Rainie and Janna Anderson (Pew Research Center): Code-Dependent: Pros and Cons of the Algorithm Age, 2016。学理上,杰克·巴尔金将"算法社会"(Algorithmic Society)定义为一个通过算法、机器人和人工智能来进行社会和经济决策的社会。参见 Jack M. Balkin, The Three Laws of Robotics in the Age of Big Data, *78 Ohio ST. L. J.* 1217, 1219 (2017)。有关算法社会的讨论,还可参见 Danielle Keats Citron and Frank Pasquale, The Scored Society: Due Process for Automated Predictions, *89 Wash. L. Rev.* 1, 3 (2014);左亦鲁:《算法与言论:美国的理论与实践》,《环球法律评论》2018年第5期,第122－139页;丁晓东:《算法与歧视:从美国教育平权案看算法伦理与法律解释》,《中外法学》2017年第6期,第1609页。

可以说,在当代社会,算法几乎无处不在、无所不能,算法应用在发展。与此同时,大数据和人工智能的兴起,使算法得以突破"波兰尼悖论"的束缚,通过基于自我训练、自我学习过程,实现自我生产和自我更新,〔2〕算法本身也在发展。

然而,算法是一把双刃剑。算法可以调节室内温度,但是也可以把房间变成冰窖火炉;算法可以自动开门,也可以把人们锁闭在屋内;算法可以自动驾驶,但也可以引发事故;算法可以治病救人,但也可以误诊杀人;算法可以帮助我们更高效地分配资源,但也可以在分配中歧视特定群体……随着算法共谋、算法失灵、算法歧视等问题的出现,"如何规制算法?"〔3〕这一命题在近两三年来,以一种近乎猝不及防的方式被推向前台,也一跃而进入主流法学界的视野。〔4〕

就像面对魔法一样,人们在直觉上对算法引发问题的第一反应,是搞清楚它到底是什么。于是,在规制算法的纷纭众说中,最广为熟知、且被普遍认可的,便是算法透明原则。〔5〕尽管各研究领域的学者对于算法透明原则的内涵认识不一,但大体上,算法透明原则被归为一种对于算法的事前规制模式,它要求算法的设计主体或者使用主体公开和披露包括源代码在内的算法要素。〔6〕让人颇感意外的是,虽然学界呼吁算法透明原则的声音不绝于耳,但却鲜有中文文献对其作理论性辨析,也没有对其在实践中的应用作归纳反思,更不用说其在整个算法规制图景中如何进行合理定位。在相关研究尚未展开的背景下,有些学者却已然将算法透明原则作为算法规制的首要原则,甚至乐观地认为,一旦透明,算法就可知,一旦可知,算法问题就可解。〔7〕本文可能就是想在对算法透明原则作出理论和实践辨析后,为这股乐观情绪泼上一瓢冷水。

在笔者看来,目前有关算法规制的讨论,夸大了算法透明原则的作用。本文旨在揭示,算法透明仅在有限的情境下适用,在多数情境下,算法透明原则既不可行,也无必要。依托对算法透明原则的批判,本文尝试回应一个理论问题:如何规制算法?本文结合学理上事前规制与事后规制、本质主义与实用主义这两对比照,对算法规制理论重构展开初步

环球法律评论 GLOBAL LAW REVIEW

〔2〕 参见贾开:《人工智能与算法治理研究》,《中国行政管理》2019 年第 1 期,第 17 – 22 页。

〔3〕 David Lehr & Paul Ohm, Playing with the Data: What Legal Scholars Should Learn About Machine Learning, *51 U. C. Davis L. Rev. 653* (2017).

〔4〕 在此,仅举几个典型案例:喜达屋 – 万豪、华住等酒店集团住客信息数据泄露;个人征信巨头 Equifax 信用数据泄露案;Facebook 千万用户数据失窃;夏威夷虚假导弹警报信息;自动驾驶失灵致死事件;波音 737-Max 飞机控制系统失灵空难等。算法本身引发了全球普遍质疑。参见 Pew Research Center: Public Attitudes Toward Computer Algorithms, 2018, pp. 2 – 7。

〔5〕 参见 Frank Pasquale, *The Black Box Society* 8 – 11 (Harvard University Press, 2015); Danielle Keats Citron, Technological Due Process, *85 Wash. U. L. Rev. 1249*, 1253 (2008); Paul Schwartz, Data Processing and Government Administration: The Failure of the American Legal Response to the Computer, *43 Hastings L. J. 1321*, 1323 – 25 (1992);郑戈:《算法的法律与法律的算法》,《中国法律评论》2018 年第 2 期,第 67 – 69 页;汪庆华:《人工智能的法律规制路径:一个框架性讨论》,《现代法学》2019 年第 2 期,第 54 – 63 页;蒋舸:《作为算法的法律》,《清华法学》2019 年第 4 期,第 67 – 69 页;张凌寒:《算法权力的兴起、异化及法律规制》,《法商研究》2019 年第 4 期,第 74 – 75 页。

〔6〕 基于这一界定,本文选择不对"算法透明""算法公开""算法披露"三者作严格区分,行文中,三词将交替出现。

〔7〕 参见张恩典:《大数据时代的算法解释权:背景、逻辑与构造》,《法学论坛》2019 年第 4 期,第 156 – 157 页;高学强:《人工智能时代的算法裁判及其规制》,《陕西师范大学学报》2019 年第 3 期,第 166 – 167 页;刘友华:《算法偏见及其规制路径研究》,《法学杂志》2019 年第 6 期,第 63 – 64 页;张淑玲:《破解黑箱:智媒时代的算法权力规制与透明实现机制》,《中国出版》2018 年第 7 期,第 51 页。

思考,并借此阐明以算法问责为代表的事后规制手段,可能才是更加得当的规制策略。而算法透明本身,只能在特定情况下,起到一定辅助效果。

一 算法透明原则

无论是在政治学、经济学还是法学领域,透明原则已成为现代政府规制的一条基本准则。早在 19 世纪中叶,杰罗米·边沁(Jeremy Bentham)和约翰·斯图尔特·密尔(John Stuart Mill)等思想家,就有针对性地讨论过透明原则。这样的讨论,逐渐成为西方自由主义视野的一部分。直至近现代,诸如德里希·哈耶克(Friedrich Hayek)和约翰·罗尔斯(John Rawls)等自由主义理论家,无一例外地都受到这些讨论的影响。在这些西方思想家看来,透明原则的民主政治,有着两大根本助益:其一,它可以增强公权力机关的可问责性;其二,它可以保护公民的知情权,保护公民免遭专权独断。〔8〕

具体到法学领域,透明原则也一直贯穿于现代法律制度之中。套用美国大法官路易斯·布兰代斯(Louis Brandeis)的一句流传甚广的名言——"阳光是最好的消毒剂"。在美国法中,透明原则不但是公法中形式正当程序(Procedural Due Process)的一个核心原则,〔9〕而且也在某种程度上,通过相关法律制度的构建,塑造了代议制民主制度。〔10〕与之类似,在我国,透明原则也成为公法领域的一项原则要求,并且在制度上有着多重体现,比如规制依据公开、行政信息公开、听证制度以及行政决定公开等。〔11〕

当然,讨论透明原则在规制理论或者政府信息公开中的正当性,已经超出了本文的范围。本文聚焦于透明原则在互联网时代的一个具体延伸——算法透明原则。之所以说是延伸,而非属于相应类目,是因为算法本身并不是由公权力机关所独享,更多地,也会被私营机构所使用。具体而言,民主政治语境下的透明原则,也仅仅是在公权力机关,或者部分带有"公共性"的私营机构使用算法时,才涉及到传统公法的透明与信息公开问题。而本文所指的算法透明原则,既适用于政府的算法规制,也适用于私营机构的算法规制;也正是在这个意义上,它有着更丰富的内涵。

虽说有关算法透明的讨论早已有之,但必须承认的是,21 世纪初的两次美国总统大选,大大推进了人们对算法透明的关注,可以称得上是"神助攻"。〔12〕2000 年大选,首次

〔8〕 Jeremy Bentham, An Essay on Political Tactics, in 2 *The Works of Jeremy Bentham* 551 (John Bowring ed., Facsimile Publisher, 2018); John Stuart Mill, *Considerations on Representative Government* 80 – 89 (Henry Regnery Co. 1962).

〔9〕 Martin H. Redish & Lawrence C. Marshall, Adjudicator, Independence, and the Values of Procedural Due Process, *95 Yale L. J. 455*, 478 – 489 (1986). 有关技术领域,透明原则与形式正当程序的讨论,参见 Danielle Keats Citron, Technological Due Process, *85 Wash. U. L. Rev. 1249*, 1254 – 1255 (2008)。

〔10〕 参见[美]迈克尔·舒德森著:《知情权的兴起:美国政治与透明的文化》,郑一卉译,北京大学出版社 2018 年版。

〔11〕 参见马怀德著:《行政法与行政诉讼法》,中国法制出版社 2015 年版,第 292 – 294 页。

〔12〕 在此之前,许多有关算法透明的讨论,都局限在技术行业内部,多与开源软件(Open Source)运动有关。其中,最经典的说法,是埃里克·雷蒙德(Eric S. Raymond)在他讨论软件工程的名著《大教堂和市集》提到的 Linux 定律,亦即"只要让足够多双眼睛盯着,所有漏洞都将无处藏身"。参见 Eric S. Raymond, *The Cathedral and the Bazaar* 9, O'Reilly Media, 1999。

环球法律评论 GLOBAL LAW REVIEW

采用电子投票器。最终,在沸沸扬扬的布什诉戈尔案(Bush v. Gore)中,投票设备(包括老式打孔机、光学扫描机和电子投票机)的透明性和公正性,成为了全社会关注的焦点。[13] 作为回应,2002 年美国国会通过了《协助美国投票法案》(*The Help America Vote Act of 2002*),着力推广电子投票机,并配套相应管理措施。之后,大量科技公司看到电子投票器的商机,纷纷涌入这一领域。然而,各类新开发的电子投票器的大规模应用,不但未消旧愁,反而又添新忧:选民们怎么知道这些电子投票机在何时将数据上报到计票中心?而计票中心是不是准确无误地记录下每一个人投出的选票?谁又能确保选票数据统计没有造假或者选票数据库不被黑客攻破?[14] 算法透明,被认为是投票监管的一剂良药,因而受到广泛讨论。[15]

其后,算法应用在广度和深度上的增加,也成为算法透明讨论的一个重要推手,算法透明逐渐成为算法规制领域的一个原则性提议。值得一提的是,学者们对算法透明原则的认识,存在不小差别,这在网络法这类交叉学科研究中,也是十分正常的现象。这种差别,用"言人人殊"来形容,有过于夸大之嫌,但换个说法,用口径不一来形容,应该是恰如其分的。虽说如此,大体而言,大家对于算法透明原则还是有普遍认同的最大公约数——即针对算法的事前规制原则,要求算法的设计方或者使用方,披露包括源代码、输入数据、输出结果在内的算法要素。[16] 帕斯奎尔(Frank Pasquale)对于算法透明的理解,更为复杂而深入,他在不同的著述中,曾把算法透明理解为综合源代码公开、算法分析、算法审计等手段合理促成的算法透明。他的这种理解,当然给他的理论带来更强的解释力,但是也在某种程度上模糊了算法透明与其他规制手段的边界,这可能会给理论和实务都带来很大麻烦。因此,本文取狭义上的算法透明概念。

算法透明原则最终的落脚点,是对于算法自动化决策的规制。而算法所主导的自动

环球法律评论 GLOBAL LAW REVIEW

〔13〕 Bush v. Gore, 531 U. S. 98 (2000).

〔14〕 COMM. ON FED. ELECTION REFORM, Building Confidence in U. S. Elections (2005), http://www.american.edu/ia/cfer/report/full_report.pdf; Jon Stokes, How to Steal an Election by Hacking the Vote, ARS TECHNICA, Oct. 25, 2006, http:// arstechnica.com/articles/culture/evoting.ars; Greg Reeves, One Person, One Vote? Not Always, Kan. City Star, Sept. 5, 2004, at 1A; Thad E. Hall & R. Michael Alvarez, Center for Pub. Pol'y & Admin. Univ. of Utah, American Attitudes About Electronic Voting: Results of a National Survey (Sept. 9, 2004), http://www.vote.caltech.edu/Reports/fall04survey.pdf,最近访问时间[2019-09-20]。

〔15〕 在 2002 年的《协助美国投票法案》中,就有诸多条款涉及投票机运行模式的披露(比如第 301 条款和第 303 条款)。同样地,电子投票专门委员会也在其指导手册中明文确立了透明原则。参见 Procedural Manual for the Election Assistance Commission's Voting System Testing and Certification Program, 71 Fed. Reg. 76, 281 (Dec. 20, 2006)。学界对于算法透明原则在电子投票程序的应用更是不胜枚举,比如 Bev Harris, Black Box Voting: Ballot Tampering in the 21st Century (Talion Publishing, 2004); Andrew Massey, "But We Have to Protect Our Source!": How Electronic Voting Companies' Proprietary Code Ruins Elections, *27 Hastings Comm. & Ent. L. J. 233*, 241-242 (2004); Lillie Coney, A Call for Election Reform, *7 J. L. & Soc. Challenges 183*, 188 (2005); Daniel P. Tokaji, The Paperless Chase: Electronic Voting and Democratic Values, *73 Fordham L. Rev. 1711*, 1773-1780 (2005)。

〔16〕 Paul Schwartz, Data Processing and Government Administration: The Failure of the American Legal Response to the Computer, *43 Hastings L. J. 1321*, 1323-1325 (1992); Danielle Keats Citron and Frank Pasquale, The Scored Society: Due Process for Automated Predictions, *89 Wash. L. Rev. 1*, 8 (2014); Frank Pasquale, Beyond Innovation and Competition: The Need for Qualified Transparency in Internet Intermediaries, *104 NW. U. L. Rev. 105*, 160-161 (2010); Frank Pasquale, *The Black Box Society* 8-11 (Harvard University Press, 2015).

化决策可以概括为:基于输入数据,通过算法运算,实现结果输出。从这个意义上看,如果对算法没有一个明确的认知,也就无从判断算法自动化决策是否公正。表面上来看,算法透明,就是打开黑箱而将“阳光”洒落整个自动化决策过程的理想手段。

与传统的透明原则能带来的优势类似,算法透明同样在可问责性和知情权两个维度发挥作用。其一,算法透明可以让算法操控者变得更具可问责性,一旦出现精确性和公平性的偏差,可以依据所披露的算法来主张算法操控者的责任。更甚之,较之人为治理的透明原则,算法透明原则还隐含着一个算法治理本身的优势,亦即,人类决策者的内在偏见和私念很难被发现和根除,但假如我们窥探算法的“大脑”,即整个决策和执行过程,就可以变得更透明、更容易被监督。〔17〕 其二,算法透明也赋予算法规制对象一定程度上的知情权,而这种知情权有利于第三方(尤其是专业人士)实施监督,也有利于算法规制对象依据所披露的算法,在事后对算法决策提出公平性和合理性的质疑。

正因为算法透明有着这些好处,许多论者对算法透明原则趋之若鹜。〔18〕 更有乐观的论者认为,只要算法透明,甚至只需源代码公开,就可以解决很多现实中的算法问题。可以说,在当前国内,算法透明原则俨然成为了算法治理实践和学术讨论首当其冲的基本原则。

二 算法透明原则可行吗?

算法透明原则本身,是不是一个不容置疑的金科玉律呢?算法透明原则真的那么有用吗?在算法运用越来越广泛、而由此引发的问题越发复杂的情境下,是不是可以说,算法越透明越好呢?答案并不是那么简单。如果单单从美国大选投票算法中的例子出发,我们会很自然地把算法透明原则与自由主义传统下的政治学、经济学和法学中的透明原则密切联系起来。然而,这很可能是以偏概全。一方面,算法透明原则——如果得以践行——无论在外延上,还是在内涵上,都与传统的透明原则有所不同。另一方面,虽然本文第一部分阐述了算法透明原则与传统自由主义下的知情权和可问责性之间存在交叉,但不能否认,比起传统自由主义的透明原则,算法透明原则蕴含着更大的内在张力和具体限制。接下来,本文将分别探讨算法透明原则的两个根本问题:算法透明原则是否可行以及算法透明原则是否必要。本部分通过具体规制情境,考察算法透明原则的可行性问题。事实上,算法透明原则作为一项带有普遍强制性的法律原则,它

〔17〕 美国有些法律和政策甚至直接将监督等同于透明,比如《自由信息法案》(The Freedom of Information Act),参见 5 U. S. C. § 552 (2012)。类似的立法还有 Federal Agency Data Mining Reporting Act of 2007, 42 U. S. C. § 2000 ee - 3 (c) (2) (Supp. III 2007)。

〔18〕 Tal Z. Zarsky, Transparent Predictions, *2013 U. Ill. L. Rev. 1503*, 1506 (2013); Todd Essig, “Big Data” Got You Creeped Out? Transparency Can Help, Forbes (Feb. 27, 2012). 这其中,最典型的应当是弗兰克·帕斯奎尔。当然,他本人对算法透明的研究更透彻,自然也对算法透明的局限性有着比较清楚的把握。参见 Frank Pasquale, Restoring Transparency to Automated Authority, *9 J. on Telecomm. & High Tech. L. 235* (2011); Frank Pasquale, Beyond Innovation and Competition: The Need for Qualified Transparency in Internet Intermediaries, *104 Nw. U. L. Rev. 105* (2010)。

环球法律评论 GLOBAL LAW REVIEW

有可能会与国家安全、社会秩序和私主体权利等法益相冲突，不具有作为与基本法律原则所匹配的普遍可行性。

（一）算法透明 vs. 国家安全

无论古今中外，公开和保密，一直是国家治理中至关重要的理念。[19] 具体到算法治理领域。哪些算法可以公开，向谁公开，公开到何种程度，都需要放在国家安全这一棱镜中，着重考察。而对于以国家安全为由的保密义务，许多国家在政策和法律层面都给予了高位阶保护。比如，我国的《国家安全法》《网络安全法》《保密法》以及美国的《国家安全法案》《爱国者法案》等。这些法律在很大程度上，都给相关的算法透明设置了障碍。换言之，当算法透明与国家安全相冲突时，算法透明的可行性必将遭受挑战。

举例而言，为了方便机场安检效率，全球大部分国际机场都采取了抽样安检策略，即在常规安检之外，抽取特定人群进行更严格繁琐的检查。如此一来，既可以保证机场安检的速度，又能给恐怖分子带来一定威慑力。抽样的程序，则由算法来执行。假设为了防止对特定群体的歧视性抽样，根据算法透明原则，公众要求公开抽样算法，那么，机场应不应当让算法透明呢？可以想见，一旦算法透明，恐怖分子有可能根据公开的算法进行博弈，谋划规避手段来避免被严格检查，或者根据算法所提供的随机性逻辑来合理定制所需样本试错数量。再比如，假设某次导弹试射演练后，制导系统的算法失灵，致使导弹偏离既定弹道，炸毁民用设施，并造成伤亡。那么，公众是不是可以就此要求算法透明，要求军方公开制导系统的算法呢？后文将提出更合理的解决方案，但就本节所讨论的主题而言，即便公众的诉求完全公平合理，但本案例中算法透明的可行性，也将在很大程度上受到限制。

在上述两个案例中，很显然，如果坚持贯彻算法透明原则，将有可能导致产生国家安全隐患（飞机航路安全与军事设施安全等）。换言之，对于算法透明原则而言，当其与国家安全相冲突时，不可避免地会受到国家安全的限制。比如美国 911 事件过后，以小布什总统为首的保守派政治家，强烈抵制政府在国家安全领域的透明化，声称赢下“反恐战争”的唯一手段，就是让美国变得和它的影子对手一样神秘。[20] 于是，以《爱国者法案》为代表、以国家安全为由对抗信息披露的法律政策，也就应运而生。同样，我国在《宪法》第 53 条、《国家安全法》第 4、19、28、29 条与《网络安全法》第 77 条，以及其他法律法规中，都对涉及国家安全、国家秘密的信息披露，进行严格限制。这些都是算法透明原则在不同的适用领域所需面对的重重关卡。

综上所述，由于通常国家安全往往比算法透明背后的考量有着更高位阶的权重，因此，一旦出现这一组对立，国家安全将对算法透明实施“降维打击”，这样一来，算法透明

〔19〕 参见张群著：《中国保密法制史研究》，上海人民出版社 2017 年版；[美]戴维·弗罗斯特著：《美国政府保密史》，雷建锋译，金城出版社 2019 年版。

〔20〕 Julian E. Zelizer, Arsenal of Democracy, Basic Books, 2010. 对于美国国家安全和信息保密的讨论，还可参见 Dana Priest and William Arkin, Top Secret America: The Rise of the New American Security State, Hachette Book Group, 2011。

环球法律评论 GLOBAL LAW REVIEW

原则的可行性就很难得到保证。这便构成了算法透明可行性的第一道也是最难逾越的一道障碍。

(二)算法透明 vs. 社会秩序

算法透明也可能与社会秩序背道而驰。我们以当前应用广泛的智能语言测试系统为例。[21] 智能语言测试系统的应用,为的是测试的便捷和标准化。语言测试系统的判分算法信息,具有很强的保密性,不能被随意披露。不难想见,一旦这类信息被披露,就很可能让不法分子钻算法的空子,与语言测试系统博弈,也让整个测试无法达到其应有的考察目的。类似的情况也会发生在抽奖活动中,如果抽奖环节所使用的算法一开始就被披露,那么,投机分子就可能采取各种手段——比如破解算法直接干预抽奖环节、选择算法抽奖所青睐的时机和频次进入抽签环节——博弈,以及操纵抽奖结果等。

当然,网络空间中最经典的例子,当属搜索引擎优化(Search Engine Optimization)。起初,搜索引擎服务提供商,曾乐于践行算法透明,将其搜索引擎算法公之于众。比如,谷歌早期的 PageRank 排名算法的排序标准就曾公之于众。[22] 然而,出乎谷歌意料的是,某些恶意网站(尤其是内容农场、[23] 商业广告网站、钓鱼网站、恶意代码网站等)利用这些被披露的排序算法,玩起了"猫捉老鼠"的游戏——采取搜索引擎优化来与谷歌排序算法展开博弈,让一些本不应被优先排序的网站,挤进了搜索结果的靠前位置。如此一来,人们也就更难通过谷歌得到理想的搜索结果。换句话说,谷歌 PageRank 排名算法越透明,其搜索结果排名就越容易被博弈和操控,最后影响到公众对于搜索引擎的体验。也正因如此,谷歌以及其他搜索引擎,逐渐收紧算法披露,到最后,谷歌几乎明确拒绝算法透明,甚至将已公开的算法作出秘密调整。就这样,谷歌搜索引擎算法彻底变成黑箱,而这个黑箱,反倒成了公众获得理想搜索结果的保障。

上述案例仅仅涉及算法程序披露,而对于输入数据(作为算法的一部分)披露的案例更是不胜枚举。屡屡出现的计算机考试漏题案件,就属于这类输入数据披露对于社会秩序的影响。[24] 篇幅有限,不一一赘述。由此可见,算法透明在实践中可能会与社会秩序发生冲突,这便是算法透明可行性的第二道障碍。

---

〔21〕 参见王金铨、陈烨:《计算机辅助语言测试与评价——应用与发展》,《中国外语》2015 年第 6 期,第 76 - 81 页;张艳、张俊:《我国计算机辅助语言测试研究现状》,《中国考试》2017 年第 5 期,第 47 - 53 页。

〔22〕 有关谷歌搜索引擎的技术细节和商业模式,参见 Siva Vaidhyanathan, The Googlization of Everything, University of California Press (2010); Amy N. Langville and Carl D. Meyer, Google's PageRank and Beyond, Princeton University Press (2012)。

〔23〕 内容农场(Content Farm)是纯粹以获得在算法排名高排位目的,雇佣大量人员来粗编烂造各类热门内容,以迎合搜索引擎算法需要的一类公司。有关内容农场以及谷歌与内容农场之间的博弈,参见 Daniel Roth, The Answer Factory Demand media and the fast, disposable, and profitable-as-hell media model, WIRED; https://www.wired.com/2009/10/ff - demandmedia/; Ryan Singel, Google Clamps Down on Content Factories, WIRED, https://www.wired.com/2011/02/google - clamp - down - content - factories/,最近访问时间[2019 - 09 - 20]。与 DuckDuckGo 和前两年刚刚被 IBM 收购的 Blekko 这类小搜索引擎不同,谷歌拒绝在其英文搜索引擎中设立黑名单,这也给内容农场及其派生网站留下了更大的博弈空间。

〔24〕 《托福考题疑泄露官方公布举报邮箱》,《新京报》2015 年 2 月 1 日第 A12 版。

环球法律评论 GLOBAL LAW REVIEW

（三）算法透明 vs. 私主体权利

算法透明原则，将不可避免地带来信息披露，而在遍布私主体信息的当代社会，信息披露将很可能与私主体权利（尤其是个人隐私、商业秘密和知识产权）相冲突。比如，在金融信贷、个人征信和医疗诊治等领域，算法已经得到普遍应用，这些领域中的法定保密义务和约定保密义务，会给算法透明原则的实现造成很大阻碍。这是因为在被披露的算法中，往往既涉及到敏感的个人隐私，也涉及到关键的商业秘密和知识产权。这些敏感信息或机密信息，可能作为算法程序的一部分，或者可能作为输入数据、输出结果，甚至可能兼而有之。

上述此类信息披露，势必与隐私保护、商业秘密保护、知识产权保护等法律法规〔25〕或合同约定相冲突，并受到后者的限制。这一现象在金融信贷领域最为典型，且不说用户个人隐私屡屡成为金融机构拒绝透明的挡箭牌，金融机构还常常利用专利权、版权、商业秘密甚至商标权等私权，来对抗算法透明。〔26〕当然，就如下文将要讨论的 States vs. Loomis 案那样，开发算法的公司所最常使用的抗辩，依然是将算法作为商业秘密来寻求法律保护。〔27〕类似的情况，不胜枚举。

本文可以继续堆砌案例，但上述案例足以表明，算法透明原则并不是一个普适原则。当然，反过来说，这并不表明算法透明原则在任何情境下都不可行；这也不表明，一旦出现与上述三种考量因素的冲突，算法透明原则就必然走投无路。即便与三种制约因素存在冲突，但只要冲突是在合理范围之内，其可行性也依然存在。比如，前文提到的投票机案例，如若将投票机的算法公之于众，无论是从国家安全、社会秩序、私主体权利等哪个角度来看，他们对可行性的阻碍均很难成立。唯一可能存在的隐患是，假如投票机的算法公开，会增加不法分子侵入系统篡改投票结果的风险，但是这样的风险，可以在技术上和监管上加以限制。〔28〕

综上所述，本部分从国家安全、社会秩序和私主体权利等三个方面，质疑算法透明原则的可行性。换言之，算法透明原则至少会受到上述三方面考量的限制，并非放之四海而皆准。

三　算法透明原则必要吗？

本文第二部分论证了算法透明并不是一个普适原则，在一些情况下并不可行。接下来要回应的问题是：即便是在算法透明可行的情形下，算法透明原则是否具有必要性？显

〔25〕 例如《网络安全法》第 45 条、《民法总则》第 111、123 条、《著作权法》第 3 条第 8 款和《反不正当竞争法》第 9 条。

〔26〕 Brenda Reddix-Smalls, Credit Scoring and Trade Secrecy: An Algorithmic Quagmire or How the Lack of Transparency in Complex Financial Models Scuttled the Finance Market, *12 U. C. Davis Bus. L. J. 87*, 91 (2011).

〔27〕 State v. Loomis, 881 N. W. 2d 749 (Wis. 2016).

〔28〕 换句话说，选民们本身并不因为算法透明，就可以在投票环节博弈去操纵结果。这与智能判卷算法有所不同，这是由于答卷人对于系统的投机性博弈（比如对于答卷模式进行调整，以迎合算法评分需求），超出了系统控制范围之外。

环球法律评论 GLOBAL LAW REVIEW

然,比起可行性问题,更麻烦的问题是,当人们好不容易克服可行性障碍而最终实现算法透明时,却发现算法透明仍然无力兑现其规制承诺。对于算法透明必要性这一问题,本部分将从两个方面分别展开论述。

正如本文第一部分所提到的,算法透明就是打开黑箱、洒下“阳光”。那么,我们首先要回答:算法透明是不是就等于算法可知?如果这一前提条件不能成立,或者不能完全成立,如果黑箱套黑箱,或者“阳光”洒落在一块谜团上,那么,算法透明原则所能带来的诸多益处,也就仍然无法兑现。

(一)算法透明≠算法可知

在一些学者看来,算法透明就足以帮助我们了解算法的所有奥秘。如果说在早前技术尚未精进的时代有这种说法,倒可称得上是值得商榷,[29]但在现如今还秉持这一观点,则就让人难以理解。在笔者看来,算法透明不等于算法可知。在它们之间,至少存在如下四道障碍:披露对象的技术能力、算法的复杂化、机器学习和干扰性披露。

披露对象的技术能力这一问题,是比较好理解的。当披露对象是非计算机专业人士时(比如与公共政策和法律裁判关系密切的法官、陪审员、执法官员和普通公众),算法本身是难以辨识的。他们的技术能力有所欠缺,因此,即便向他们披露源代码和相关技术细节,可对他们而言,代码即乱码、算法像魔法,可能还是无法搞清自动化决策究竟是怎么做出的。外行只能看热闹,内行才能看门道。不可否认,外行可以借助内行来帮忙(比如专家证言),但这其中,可能会有成本和偏差。

如果说上述第一个障碍是阻挡外行的门槛,那么后面三个障碍,就把外行内行统统拒之门外。先说算法的复杂化。事实上,即便是简单的算法,也存在不可知的情况,比如计算机领域著名的莱斯定理(Rice's Theorem),就证明了某类算法的不可知属性。[30] 随着技术的不断演进、算法分工的不断精细以及社会生活对于算法需求的不断提升,大量算法变得愈发复杂。此处之所以着重强调复杂性,是因为复杂算法的不可知情况更具代表性——它既包含了单一算法本身的原因,也包含了更普遍的、多组算法模块交互的原因。而算法的复杂化,会给算法的解释工作带来很大难度。[31] 当然,这在计算机科学发展史上并不新鲜。计算机工程师应对这一问题的通行做法是:将算法系统模块化。[32] 对于模块化后的算法,计算机工程师再分别解释各部分子算法,各个击破,最后通过重新组合,解

〔29〕 [美]劳伦斯·莱斯格著:《代码 2.0》,李旭、沈伟伟译,清华大学出版社 2009 年版,第 154 - 167 页;Danielle Keats Citron, Technological Due Process, *85 Wash. U. L. Rev. 1249*, 1308 - 1309 (2008); David M. Berry & Giles Moss, Free and Open-Source Software: Opening and Democratising e-Government's Black Box, *11 Info. Polity 21*, 23 (2006)。

〔30〕 参见 H. G. Rice, Classes of Recursively Enumerable Sets and Their Decision Problems, *74 TRANSACTIONS AM. MATHEMATICAL SOC'Y 358* (1953)。

〔31〕 Katherine Noyes, The FTC Is Worried About Algorithmic Transparency, and You Should Be Too, PC World (Apr. 9, 2015).

〔32〕 Edsger W. Dijkstra, The Structure of the "THE"——Multiprogramming System, *11 COMM. ACM 341*, 343 (1968).

释整个算法系统。[33] 虽然通过模块化的分工,可以解决一部分复杂算法的解释问题,[34] 但即便如此,就连计算机工程师也承认,算法复杂化模块化,会令各个部分算法之间的相互反应变得不可预测。[35] 与此同时,如果要保证模块化处理运行顺畅,就需要在算法系统设计之时,进行整体规划;否则,复杂算法的模块化解释,也很可能达不到预期效果。在很多情况下,复杂算法应用和交互(比如 API 和云计算)无法确保我们从多个模块解释的组合中,或者与其他算法的交互中,对算法进行准确解释。[36] 简言之,算法的复杂化加大了我们理解算法的困难;而模块化这一解决进路,如果不是在算法系统设计之初就事先规划,也不能很好地解决复杂算法的解释问题。

相比算法的复杂性,机器学习对于算法可知的挑战,吸引了更多关注。[37] 传统算法要求计算机工程师事先指定一个表示结果变量的运算模式,作为以特定方式选定解释变量的参数,以此来决定输出结果。与传统算法不同,机器学习,作为一种更智能、更动态的算法,其运算不受固定参数所控制,也正因此,机器学习并不要求工程师事先指定运算模式。[38] 当然,"不要求"不等于"不能够",机器学习的门类中,也存在计算机工程师事先指定运算模式和控制学习材料的监督学习,与之对应的是运算更为自由而不可控的无监督学习和强化学习。对于这三种机器学习算法的通行分类,笔者无意展开技术分析。唯一与本部分论证有关的是,相对于后两者而言,计算机工程师对于监督学习的把控度更高。对于后两者,只要机器学习算法正在动态运行,我们就无法控制他们如何组合和比较数据,自然也无法顺利地解释机器学习算法本身。

而与算法可知直接相关的是,对于机器学习算法,其运算的函数关系不一定是固定清晰的数据集合。我们既无法保证机器学习过程代表任何一组真实关系,也无法通过此刻的因果关系,来推导未来的因果关系,因为算法本身不断学习、不断变化,在算法披露的那一刻过后,披露的算法就已经过时。古希腊哲学家赫拉克利特那句名言"人不能两次踏进同一条河流",在机器学习中找到了最好的印证。最典型的例子,便是智能广告推送算法,上一秒出现的推送结果,算法会根据你是否在页面停留或点击推送,进而计算出下一秒的推送结果。再比如,大部分垃圾邮件过滤算法,都使用邮件地址和 IP 地址的黑名单,应用最为广泛的,便是 Spamhaus,其邮件地址和 IP 地址也是根据用户举报和自身机器学

环球法律评论 GLOBAL LAW REVIEW

〔33〕 Id. at 344. Helen Nissenbaum, Accountability in a Computerized Society, *2 Sci. & Engineering Ethics 25*, 37 (1996).

〔34〕 [美]卡丽斯·鲍德温、金·克拉克著:《设计规则模块化的力量》,张传良译,中信出版社 2006 年版,第 131 - 172 页。

〔35〕 [美]卡丽斯·鲍德温、金·克拉克著:《设计规则模块化的力量》,张传良译,中信出版社 2006 年版,第 222 - 225 页。

〔36〕 Sendil K. Ethiraj & Daniel Levinthal, Modularity and Innovation in Complex Systems, *50 MGMT. SCI. 159*, 162 (2004); Richard N. Langlois, Modularity in Technology and Organization, *49 J. Econ. Behavior & ORG. 19*, 24 (2002).

〔37〕 Will Knight, The Dark Secret at the Heart of AI, *MIT Technology Review* (April 11, 2017); Andrew D. Selbst & Solon Barocas, The Intuitive Appeal of Explainable Machines, *87 Fordham L. Rev. 1085* (2018).

〔38〕 Richard A. Berk, Statistical Learning From Regression Perspective 13, Sprirger (2018); Cary Coglianese & David Lehr, Regulating by Robot: Administrative Decision Making in the Machine-Learning Era, *105 Geo. L. J. 1147*, 1156 - 1157 (2017).

习实时更新,换句话说,其这一刻不在黑名单上的邮件地址和 IP 地址,很可能在下一刻就会上黑名单。[39]

由于机器学习的决策规则本身,是从被分析的特定数据中不断生成的,因此,除了极少数被严格控制的监督学习以外,我们根本不能考察静态的源代码或原始数据,无法用这样一种刻舟求剑的进路,来推断机器学习算法的运算结果。也就是说,对于绝大部分机器学习的输出结果,无论输入和输出的因果关系在表面上看起来多么直观,这种因果关系都很可能根本无法被解释,其动态的变化也更难以把握。[40] 更重要的是,对于机器学习(尤其结合了强人工智能和神经网络等技术的机器学习)而言,输入数据的变化和累加,使得算法推算结果背后的深层原因,变得难以把握,在这个意义上,它本身就是一个无法实现透明的"黑箱"。而且,机器学习所推导的"因果关系",在很大程度上取决于输入数据,这类因果关系只能是统计意义上的因果关系,它与规范意义上的因果关系,存在一道难以跨越的鸿沟。例如,谷歌研发的强化学习算法——AlphaGo。设计 AlphaGo 的计算机工程师,都是棋力一般的业余爱好者,无法与柯洁、李世石这样的顶尖高手较量。但恰恰是这些工程师设计了 AlphaGo,把顶尖高手一一击败。[41] 可以想见,这些工程师本人是没有办法一一解释 AlphaGo 的每一步棋招——如果工程师真的能理解每步棋的奥妙,那么他们自己可能就是世界冠军了。换言之,AlphaGo 通过机器学习习得的竞技能力,工程师根本无法企及,他们的每一步棋,也自然超出了工程师的理解范畴。

最后一个阻碍算法透明向算法可知转化的障碍,是干扰性披露。与前三个与透明直接冲突的原则不同,干扰性披露本身,也可以被看成是算法透明的一种方式。它通过披露大量冗余干扰性数据,混杂在关键数据中,以此妨碍解释关键数据内容。也正是在这个意义上,干扰性披露是算法透明的一个典型悖论,亦即,公开的越多,可能对算法关键内容的理解就越困难。

其实,在《黑箱社会》一书中,帕斯奎尔就论述过这个现象,他称之为"混淆"(Obfuscation),其内涵与干扰性披露是一致的,就是指刻意增加冗余信息,以此来隐藏算法秘密,带来混淆。值得一提的是,帕斯奎尔的《黑箱社会》里,更多是指出黑箱社会或者说算法不透明带来的问题,而关于解决之道,他也并非一味奉行算法透明。[42] 哪怕极力主张算法透明的帕斯奎尔,也承认干扰性披露本身,也是算法黑箱的始作俑者之一。[43] 因为公开的算法内容越多、信息量越大,算法分析的工作量和难度也会随之增加,在这个意义上,我们也与算法可知越来越远。这就好像有些公司为了妨碍会计审查,有意披露大量的冗

〔39〕 SPAMHAUS, https://www.spamhaus.org/sbl,最近访问时间[2019-09-20]。

〔40〕 Cary Coglianese & David Lehr, Regulating by Robot: Administrative Decision Making in the Machine-Learning Era, *105 Geo. L. J. 1147*, 1156-1157 (2017).

〔41〕 Eric Mack, Google's AlphaGo Zero Destroys Humans All on Its Own, CNET, (Oct. 20, 2017); David Silver et al., Mastering the Game of Go with Deep Neural Networks and Tree Search, *529 NATURE 484*, 484 (2016); David Sliver et. al., A General Reinforcement Learning Algorithm That Masters Chess, Shogi, and Go Through Self-play, *SCIENCE 362*, 1140-1144 (2018).

〔42〕 Frank Pasquale, The Black Box Society 6-8, Harvard University Press (2015).

〔43〕 Frank Pasquale, The Black Box Society 8, 16, Harvard University Press (2015).

环球法律评论 GLOBAL LAW REVIEW

余材料,让调查人员不得不在几万份材料里大海捞针。而干扰性披露的存在,不但妨碍了算法可知,而且从另一个角度也强化了本文对于算法透明必要性的质疑。

综上,算法透明不等于算法可知,甚至有可能会妨碍算法可知。算法透明并不是终极目的,它只能是通向算法可知的一个阶梯。并且,这一阶梯也并非必由之路,这一点,将会在本文第四部分进行论述。因此,对于某些算法,即便算法透明,如果未能达到算法可知,也是于事无补,甚至适得其反。事实上,这是算法透明原则与传统公法上的透明原则所存在关键区别。传统公法上的透明原则,无论是立法上的透明,还是执法与司法上的透明,尽管不能百分百排除明修栈道、暗度陈仓的可能,但大体上,社会公众都能对所披露的信息(文本、音视频内容)有着较为明晰的认识。而算法透明原则却不尽然。一旦透明之后亦不可知,其透明性所能带来的规制效果也就无从谈起;更甚之,像干扰性披露那样误导披露对象,反而会减损而非增强规制效果。

(二)算法透明不能有效防范算法规制难题

对于算法必要性的第二个质疑,涉及到算法规制的实践。此处所要讨论的问题是:即便算法透明原则可行,那么,其是不是就像一些学者认为的那么必需,那么灵验,能防范算法歧视、算法失灵以及算法共谋等各类算法规制难题?本文认为,究其本质,算法透明原则仅仅是一种事前规制方式,我们不能夸大其在规制中的效用。

首先,算法即使透明、可知,也不意味着算法问题必然能被发现。单就算法漏洞而言,就包括了输入漏洞、读取漏洞、加载漏洞、执行漏洞、变量覆盖漏洞、逻辑处理漏洞和认证漏洞等。[44] 这些漏洞中的一部分,的确可通过算法透明来防患于未然,但另外的部分,却需要在算法执行过程中,才能被发掘并加以解决。比如,著名的 Heartbleed 安全漏洞,从程序开发到安全漏洞被发现,用时整整两年,而该算法是开放源代码,完全符合算法透明原则——算法透明原则并不能帮助工程师在两年间发现这一漏洞。[45]

其次,即便算法透明,计算机工程师也不能确切预测算法与外部运行环境的交互。对于一些算法而言,它们的运行,需要依赖于外部环境,比如外部软件[46]和外部客观条件等。例如,对于航空智能控制程序,需要根据特定时间的风向、风速、天气状况、飞机飞行角度等诸多外部客观条件,来决定具体输出的结果。而最近波音飞机由于算法失灵接连发生两起坠机事故,恰恰证明,即便算法透明,我们也无法有效避免算法失灵。而有赖于云计算、API 等技术,目前算法与外部环境的交互已变得越来越频繁,这种交互带来的情境变化,让算法透明更加无力承担化解算法问题的重任。

最后一点,也和算法透明的事前规制性质有关。即便算法透明,在执行算法的过程中,仍然无法保证排除第三方干预,从而影响最终结果。就像克鲁尔(Joshua A. Kroll)等人所指出的那样:“不管算法有多透明,人们仍然会怀疑,在他们自己的个案中,公开的算

〔44〕 参见尹毅著:《代码审计:企业级 Web 代码安全架构》,机械工业出版社 2015 年版。

〔45〕 Zakir Durumeric et al., The Matter of Heartbleed, *14 ACM Internet Measurement Conf. 475* (2014).

〔46〕 Managing Software Dependencies, GOV. UK Service Manual, https://www.gov.uk/service-manual/technology/managing-software-dependencies,最近访问时间[2019-09-10]。

环球法律评论 GLOBAL LAW REVIEW

法规则是否真的被用来做出决策。尤其是当这个过程中涉及到随机因素时,一个被安检抽查或被搜身的人可能会想:我难道是真的是被公平的规则选中了吗? 还是决策者一时兴起,挑中了我?"[47] 比如,在 State v. Loomis 一案中,一位名为卢米斯(Eric Loomis)的犯罪嫌疑人,被 COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)算法裁判为"累犯风险较高"。[48] COMPAS 通过算法计算出罪犯在前次犯罪后两年内的"累犯风险",而算法所依据的是罪犯的各项生理特征和社会背景。COMPAS 通过算法,可以给每一位罪犯计算出他的"累犯风险指数"。诚如卢米斯的诉状所主张的,不管 COMPAS 算法有多透明,他仍质疑,在自己的案例中,公开的算法规则是否真的被用来作出决定。再比如电子酒精测试仪的算法。算法的披露,并不能保证测试结果的公正。在执行过程中,探头可能老化失灵、执法人员可能因操作失误、受贿、种族性别歧视而有意控制探测部位等等,规制程序的诸多环节,都可能使透明算法规则导出不公正的裁判。换句话说,如果我们在算法公开和被披露之后,在执行算法的环节,受到算法之外的第三方因素介入,就像电子游戏的"外挂"或者黑客入侵程序一样,仍然可能导致算法得出不公正结果。[49] 而这些算法规制的问题,是所有事前规制手段的一个盲区。

在此,笔者并不是想证明,除了算法透明原则之外,其他的规制手段在应对执行环节的问题时,就能无往不利。本文只想指出,算法透明原则作为一项事前规制,有着它自身的局限,它并不能提供解决算法问题的万灵药方。而算法不透明也可能有其自身的价值(比如隐私保护、国家安全等),一味强调透明,非但不能保证解决现有问题,还可能带来新的算法规制问题。

四 算法透明的合理定位和算法规制的重构

从算法透明的可行性和必要性两个维度而言,该原则在算法治理中存在缺陷和不足。尽管如此,我们不能否认,算法透明原则仍然在某些情境下,有其适用的可行性和必要性。于是,本部分结合其他相关规制模式,探讨算法透明原则在算法规制中的地位问题,并进一步重构目前的算法规制理论。

(一)计算机科学角度的算法透明

首先,我们来考察一下计算机科学角度的算法透明。美国计算机协会(Association for Computing Machinery)作为算法治理的业界权威,在 2017 年,公布了算法治理七项原则(见下表)。[50]

---

〔47〕 参见约叔华·克鲁尔、乔安娜·休伊、索伦·巴洛卡斯、爱德华·菲尔顿、乔尔·瑞登伯格、大卫·罗宾逊、哈兰·余:《可问责的算法》,沈伟伟、薛迪译,《地方立法研究》2019 年第 4 期,第 102 - 150 页。

〔48〕 State v. Loomis, 881 N. W. 2d 749 (Wis. 2016).

〔49〕 James Grimmelmann, Regulation by Software, *114 Yale L. J. 1719*, 1741 - 1743 (2005).

〔50〕 Association for Computing Machinery Public Policy Council, Statement on Algorithmic Transparency and Accountability (Jan. 12, 2017), https://www. acm. org/binaries/content/assets/public - policy/2017_usacm_statement_algorithms. pdf,最近访问时间[2019 - 09 - 10]。

环球法律评论 GLOBAL LAW REVIEW

表　美国计算机协会(ACM)算法治理七项原则

序号	原则	基本内容
1	知情原则	算法所有者、设计者、操控者以及其他利益相关者,应该披露算法设计、执行、使用过程中可能存在的偏见和可能造成的潜在危害
2	访问和救济原则	监管部门应该鼓励落实相关机制,确保受到算法决策负面影响的个人或组织,享有对算法进行质询并救济的权利
3	可问责原则	即使使用算法的机构无法解释算法为何会产生相应结果,它们也应对算法决策结果负责
4	解释原则	我们鼓励使用算法的机构解释算法运行步骤以及具体决策结果
5	数据来源处理原则	算法设计者应该说明训练数据的采集方法以及数据收集过程中可能引入的偏见;对于数据的公共监督最有利于校正数据错误;处于隐私保护、商业秘密保护、避免算法披露后的恶性博弈等事由,可以只对适格的、获得授权的个人进行选择性披露
6	可审计原则	模型、算法、数据和决策结果应有明确记录,以便必要时接受监管部门或第三方机构审计
7	检验和测试原则	使用算法的机构应该采取有效措施来检验算法模型,并记录检验方法和检验结果;使用算法的机构尤其应该定期采取测试,来审计和决定算法模型是否将会导致歧视性后果,并公布测试结果

从上述列表中,我们可以得到四个有关算法透明的教益。

第一,知情原则对应的是算法透明中算法规制对象的知情权这一面向。但是,计算机工程师对于算法透明中的“知情”有更务实的把握——直接公开源代码不等于知情;而且,我们还需关注更深层次的“知情”,亦即“算法设计、执行、使用过程中可能存在的偏见和可能造成的潜在危害”。

第二,计算机工程师对于算法透明的功用,有着更为清醒的认识,他们认为即便是公开和披露算法,也无法确切把握最终运算结果。于是,他们使用了“可能存在的偏见”(第1项和第5项)和“可能造成的潜在危害”(第1项)这样的模糊字眼,其所隐含的信息是,我们对算法的认知,只能力图接近,但很难确切把握。这与部分法律人对算法透明脱离实际的期许,形成鲜明对比。

第三,计算机工程师明确意识到,算法披露本身,也受到其他条件的制约,比如第5条提到的隐私保护、商业秘密和恶性博弈。而这些制约,正如本文第二部分论述的那样,与算法透明的可行性有着持久的张力。尽管限制披露对象(只对适格的、获得授权的个人进行选择性披露),可以缓和这种张力,但这也无法根本解决所有冲突。

第四,对于前文讨论的算法规制的两大类别,计算机工程师所关注的,是事后规制,而非事前规制。除了第1项的部分内容和第5项之外,其余手段大体上均为可以纳入事后规制范畴。

从上述分析我们可以看出,计算机工程师——作为对算法技术比较熟悉的专家——对算法透明的局限,有着清醒的认识。一般而言,工程师更关心技术的细节,而法律人更

环球法律评论 GLOBAL LAW REVIEW

关心技术所带来的权利、义务和责任。照此逻辑,比起法律人,工程师应该更关注算法透明所能带来的对于技术细节的理解及其对算法规制的意义。然而,在计算机工程师眼里,算法透明却并不处于算法规制的核心地位,这很能说明问题——要么就是算法透明,由于客观原因而难以实现,或者即便能够实现,也无法确保他们对于技术细节的理解;要么就是算法透明本身不足以让我们能够解决相应的算法规制问题。或许正是因此,以美国计算机协会为代表的业界,并未对算法透明报以奢望,而是倾向于事后规制(如救济、审计、解释、验证、测试、问责等)为主的规制策略。[51]

(二)算法透明原则的合理定位

算法透明原则仅仅是一种事前规制方式,尽管在某些情形下有可能实现"防患于未然"的作用,但是,我们并不能夸大其在规制中的效用。算法透明并不是终极目的,它只能是通向算法可知的一个阶梯。而算法可知,最终也要服务于其他规制手段。这一点与上述计算机工程师对算法透明的定位相吻合,也可以呼应透明原则的传统政治学定位。

更重要的是,算法透明所能带来的规制效用,在很大程度上,可以被以算法问责为代表的事后规制手段所涵盖。算法规制最成熟的实例之一,便是美国对于 P2P 算法在音视频内容分享领域的规制。P2P 算法本身只是一种更为高效的文件传输技术,但在它问世之后迅速被用来传播音视频文件,其中大部分都是盗版内容。为了治理这类算法滥用,音乐电影产业和互联网公司的合力推动了版权立法和司法,而这种规制,更多地是以事后算法问责的形式出现。对于版权领域的算法问责机制,美国法传统有着多个层级的民事或刑事责任可以被运用,比如法人责任(Enterprise Liability)、替代侵权责任(Vicarious Liability)、帮助侵权责任(Contributory Liability)和产品责任(Product Liability)等。[52] 这一系列算法问责机制,对于算法的设计、执行和使用各个环节,都具有规制力。而算法本身,或者说算法透明所指向的算法可知,对于厘清侵权事实或许有一定帮助,但却不是问责机制的重点。哪怕曾被 P2P 技术案件中所关注的"中心服务器模式"和"去中心服务器模式"的区分——可以通过算法透明来厘清——也可以在随后的判例中被消解,法官后来只看重的是算法在后果上,是否构成法律意义上的"帮助侵权",而不是技术层面的"中心服务器模式"和"去中心服务器模式"的区分本身。[53]

正如前文分析所示,无论是从技术现实角度,还是从法理逻辑角度,算法透明都难以承担算法规制基本原则这一定位;充其量它也只能扮演一个辅助角色。打个比方,算法透明原则在算法规制中的地位,就类似于《福尔摩斯》中的华生医生——他对于简单的案件

---

〔51〕 类似地,国内业界对于人工智能和深度学习软件进行规制时,主要也采取了事后规制的手段。参见中国人工智能开源软件发展联盟:《人工智能—深度学习算法评估规范》,2018 年 7 月 1 日。

〔52〕 Alfred C. Yen, Internet Service Provider Liability for Subscriber Copyright Infringement, Enterprise Liability, and the First Amendment, *88 Geo. L. J. 1833* (2000).

〔53〕 有关 P2P 技术的几个经典判例,参见 A&M Records, Inc. v. Napster, Inc., 239 F. 3d 1004 (9th Cir. 2001); In re Aimster Copyright Litig., 334 F. 3d 643 (7th Cir. 2003); Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 545 U. S. 913 (2005)。到了 2005 年的 Grokster 案,法官已经摒弃了原有的技术层面的"中心服务器模式"和"去中心服务器模式"的区分,而将案件的焦点放在帮助侵权责任与替代侵权责任的问题中。

环球法律评论 GLOBAL LAW REVIEW

事实调查和分析可能对福尔摩斯办案有帮助,但不是每个案子都派得上用场。弄清了算法透明作为华生医生这一定位,下文将给出线索,帮助我们寻找算法规制领域真正的"福尔摩斯"。

(三)算法规制的重构

正如前文所述,传统政治经济学对于透明原则的考量,出发点都和限制公权力密不可分。一方面,透明原则可以加强对于政府的可问责性;另一方面,透明原则也可以赋予公民更大的知情权。然而,传统透明原则与本文所讨论的算法透明原则,在内在逻辑和实际应用方面,都有所不同。尽管政府也开始逐步使用算法施政,但目前大部分算法(包括大部分政府所使用的算法)都是由公司所开发,且这些算法的行为后果也不仅仅限于公民(也可能包括政府本身),因此,对于透明原则所能带来的强化政府可问责性和公民知情权两方面理据,并不能——至少不能完全——适用于算法透明原则。更重要的是,正如前文所述,比之传统政治经济学上的透明原则,算法透明原则在可行性和必要性上,有着很大瑕疵。换句话说,在实际应用层面上,算法透明原则也难以兑现我们对传统透明原则所期待的规制效果。

当然,本文前面的内容,集中讨论了算法透明原则在算法治理中的应用及其限制。可是,到目前为止,本文还没有具体展开"如何规制算法"这一核心问题。基于我们对算法透明的合理定位,接下来,本文将抛砖引玉,提出算法规制重构方面的一些思考。由于算法透明在规制效力上的不足和限制,它仅仅能在一些情境下作为辅助规制手段。在应用特定的技术措施来矫正算法问题之后[54]的事后规制,尤其是算法问责,应该是法律人所更应关注的重点。[55]

通常而言,事前规制注重于损害发生之前的防范,而事后规制则注重损害发生之后的解决。就像 P2P 算法规制所揭示的那样,对于这两种不同规制进路的强调,有着强烈的现实意义。并且,如果我们从成本收益分析的维度切入这一现实意义,就可以看得更加清晰。事前规制往往在损害防范成本低于损害发生成本时,被优先采用。[56] 在算法规制这一领域,如前所述,算法透明作为事前规制模式的一种,其防范损害发生的成本太高(尤其在面对机器学习和人工智能之时),而同时收效也没有保证。必须承认,技术发展是一个动态、多维度的过程。如果未来可以回到我们在算法原初之时对它的把握和认知,那么算法透明的成本是可以降低的。但目前我们看到的趋势,正好与之相悖。2019 年图灵奖就颁给了研究人工智能和深度学习的几位科学家,而他们的研究成果,恰恰是增加算法透

〔54〕 对于具体技术措施,可以参考克鲁尔等人的文章,其中提及四种常见的矫正算法规制问题的技术措施,亦即软件检验、加密承诺、零知识证明和公平随机选择。参见约叔华·克鲁尔、乔安娜·休伊、索伦·巴洛卡斯、爱德华·菲尔顿、乔尔·瑞登伯格、大卫·罗宾逊、哈兰·余著:《可问责的算法》,沈伟伟、薛迪译,《地方立法研究》2019 年第 4 期,第 102 - 150 页。这一部分前置程序,并非本文讨论的重点,但需要强调的是,多种事后规制手段,都可能反过来倒逼相关技术措施的开发与应用。

〔55〕 参见中国人工智能开源软件发展联盟:《人工智能—深度学习算法评估规范》,2018 年 7 月 1 日。

〔56〕 Steven Shavell, Foundations of Economic Analysis of Law 87 - 91, 428 - 430, 479 - 482, Harvard University Press (2004).

环球法律评论 GLOBAL LAW REVIEW

明的成本。即便损害发生的成本很高(比如飞机失事),也不能保证算法透明这一事前规制模式,是经济学上的更优选项。而事后规制在成本方面的好处主要有两点:其一,事后规制把一些很难获知且不一定有用的技术细节,利用事后规范或者追责的方式抹平——我们把注意力集中到通过责任分配等手段来解决,而从成本收益角度跳出了泥潭;其二,比之事前规制,事后规制在信息成本方面有着天然优势〔57〕——行为和后果往往在事后更容易得到明确,这点对于复杂算法所引发的后果尤其显著。本文限于篇幅,无力对算法规制作出细化的成本收益分析,但总体而言,笔者认为,事前规制在多数情况下,并非算法规制的更优选项,而作为事前规制手段的算法透明,更由于其在可行性和必要性上的不足,比之其他事后规制手段,其成本收益更显劣势。

除了成本收益考量之外,这两种进路的对比,也在某种程度上,折射出更深层次的两个算法规制理论面向:本质主义(Essentialism)和实用主义(Pragmatism)。这不禁让人想起几年前,雷恩·卡洛(Ryan Calo)和杰克·巴尔金(Jack Balkin)关于机器人规制的辩论。

对于机器人的规制,卡洛秉持本质主义进路,关注其机器人的技术特性,认为我们一定要先搞清楚机器人的技术特性,然后再根据这些技术特性,来实施对技术的规制。〔58〕巴尔金对卡洛本质主义的批判,非常有力也富有启发。他指出,包括卡洛本人在内的几乎所有当代美国法律人,都受到霍姆斯大法官的法律现实主义的影响。〔59〕而按照法律现实主义者对于法律与技术的理解,技术特点其实并不那么重要,真正重要的是技术的应用方式以及这些应用所带来的,以权力配置为代表的社会关系变化。这是由于技术的背后,还存在着人们怎么使用、博弈甚至规避技术这些具体实践。而就像乔纳森·兹特芮恩(Jonathan Zittrain)提到的创生性(Generative)技术那样,人们在使用技术的时候,往往会背离开发人员的初衷,也可以有很多变化,并在使用过程中不断地改进技术。〔60〕法律人应该关注这些技术变动背后的社会关系变动,而不是变化的技术本身。这显然是非常霍姆斯也非常实用主义的观点。

让我们回到 P2P 技术的例子。究其本质,P2P 技术就是一个共享文件的软件,但迅速被用来传播盗版音视频文件,并且依据这一特定需求,而开发出很多新的附带播放、缓存、去中心化等功能的盗版音视频共享“神器”。如果我们接受巴尔金的观点,把重点放在考察技术背后的社会关系,我们就能够跳出本质主义所设置的迷宫,更直接地回应具体的规

〔57〕 有关事后规制在信息成本方面优势的经典论述,参见 Richard A. Posner, Economics Analysis of Law 490 – 91, 8th ed., Wolters Kluwer Law & Business (2011)。

〔58〕 Ryan Calo, Robotics and the Lessons of Cyberlaw, *103 Cal. L. Rev. 513* (2015).

〔59〕 Jack M. Balkin, The Path of Robotics Law, *6 Cal. L. Rev. Circuit 45* (2015). 巴尔金将其文章标题取为“The Path of Robotics Law”,为的是呼应霍姆斯法官的经典文章“The Path of Law”。参见 Oliver Wendell Holmes, Jr., The Path of the Law, *10 Harv. L. Rev. 457* (1897)。霍姆斯法官在文章中强调:由于法律是社会生活综合力量所推动而成,我们应当从其社会功能和具体适用角度,来理解法律。事实上,不单单是美国法学界受到实用主义的影响,实用主义的痕迹遍及整个 20 世纪的美国社会科学界。参见[美]多萝西·萝丝著:《美国社会科学的起源》,王楠、刘阳、吴莹译,三联书店 2019 年版。

〔60〕 Jonathan Zittrain, The Future of the Internet: And How to Stop It 67, Yale University Press (2008).

环球法律评论 GLOBAL LAW REVIEW

制问题。不再过多纠结于技术本质,也可以帮助我们更好地考察与具体权利义务关系有着更直接关联的规制要素。比如对于 P2P 技术所引发的盗版问题的规制,与其纠结于技术本质,不如更多关注人们使用或规避 P2P 技术时,所引发的权利义务关系的变化。现如今 P2P 技术下载的盗版音视频作品得到遏制,除了法律规制以外,还要依赖于更便捷的流媒体(附带会员和广告营销)商业模式——既然获得正版的成本没有那么高,人们也就没必要承担 P2P 盗版的法律风险和麻烦。而这些都与 P2P 技术算法的具体细节,并没有直接关联。

重温卡洛与巴尔金的论辩,有助于我们理解以算法透明为代表的事前规制与以算法问责为代表的事后规制的区别,对算法规制理论的构建,有着重要意义。前者关注技术本质,后者关注技术所引发的后果,两种规制思路的分野,在某种程度上,恰恰折射出关于技术本质的算法透明和以算法问责为代表的、关注法律后果的规制模式的比照。算法透明,就是要规制者搞明白,目标算法究其本质是什么,根据算法的特性来施以规制。而以算法问责为代表的事后规制模式,就是要规制者去考察算法在实际运作中的具体结果及其背后的社会关系变化,针对它们来施以规制。[61] 这种学术讨论上的比附,也有助于我们反思当前算法透明原则在理论上的悖谬,以避免陷入“透明”“公开”“开放”等一些大词的迷思,而忘却法律人面对的具体规制问题以及其中可能存在的理论意义。换言之,法律人面对算法规制问题时,应当着重考量算法所引发的、以权力配置为代表的社会关系的变化(比如算法何以引发歧视性后果),而不是把关注点放在算法的技术本质(比如源代码是如何编写的)。

本文的论证进一步表明,带有强烈本质主义色彩的算法透明,在可行性和必要性上都存在瑕疵,只能作为算法规制的辅助手段存在。换句话说,算法本质应不应该被探究、能不能被探究清楚以及探究清楚之后能否保证有效规制,在本文看来,统统存疑。反之,实用主义导向的事后规制手段,较之算法透明有着更多优势,应该作为算法治理中的主要手段,而且也应当是法律人可能的理论贡献所在。后者,才是法学界应对算法问题的福尔摩斯。

当然,有些人可能质疑,我们一开始就把解决问题的重点放到了算法应用效果上,那么算法本质与算法应用之间转化的相关规制问题,可能就会在结构上被忽略了。这并非笔者本意。事实上,事后规制并不排斥在显现应用特定的技术措施来矫正算法问题,而且很多改造算法本质的技术措施,恰恰是由于事后规制倒逼而产生的。比如美国通过《儿童在线隐私保护法》(COPPA)及后续一系列判例形成对算法的事后问责之后,儿童保护网络内容软件也在不断改进迭代。

最后,我们再把这一规制进路,具象化地放到部分前例中。对于机场安检歧视,不应当算法透明,而更适合事后问责;导弹试射事故,不应当算法透明,而更适合事后问责;自

〔61〕 Deven R. Desai & Joshua A. Kroll, Trust but Verify: A Guide to Algorithms and the Law, *31 Harv. J. L. & Tech. 1*, 6 (2017).

环球法律评论 GLOBAL LAW REVIEW

动飞行事故,没必要算法透明,而更适合事后问责;酒精检测失灵,没必要算法透明,而更适合事后审计……而如何把后面这些具体的事后规制制度的设计得更好,恰恰是法律人理应关注的问题。篇幅有限,笔者在本文中无意也无法提供完整的算法规制图景,但就目前文章所论,至少揭示了算法透明的局限性,以及事后规制在实践中和学理上的优越性,为后续的讨论提供了基础。

结 语

为了应对当下算法在社会生活的应用中带来的一系列问题,法学界对于算法规制,有着迫切的需求。而学界对于算法透明原则的推崇,也是在某种程度上构成了算法规制问题及其制度回应的重要组成部分。然而,正如本文所揭示的,目前法律人所极力推崇的算法透明原则,作为事前规制的一种方式,其在可行性和必要性上,都存在瑕疵。本文无意完全否定算法透明在算法规制中的作用,但我们更应当充分认识算法透明的不足和适用的局限。而更为合理的规制手段,应当是实用主义导向的、以算法问责为代表的事后规制手段。

本文旨在进一步揭示批判算法透明原则的理论意涵。不可否认,在切入法律与技术这一交叉领域时,法律人当然有必要对技术有所了解,才能言之有物。[62] 然而,法律人对于技术本质的过分强调,可能会带来研究的困境和危险,体现在两个方面:其一是盲目夸大,由于自身技术专业能力不足,从而"神圣化"或"妖魔化"技术本质;其二是削足适履,过分纠结于技术本质,导致无法充分考察法律及其他规制要素对技术所引发的社会关系可能的回应。毕竟法律更应关注的,是算法失灵、算法歧视以及算法共谋等问题所带来的权利、义务和责任的关系,而不是这些技术问题本身,而后者是计算机工程师所关注的。

法律人一味强调算法透明,哪怕披上了一件漂亮的"科学"外衣,其在法律和制度层面上的意义,依旧是模糊的,甚至我们可以断言,单纯地探究算法透明,将限制法学界在算法规制领域的贡献。在笔者看来,那些一味强调算法透明的法律人,一方面,很可能是对算法技术本身一知半解,对算法可知以及算法透明的应用范围和规制效用,抱有不切实际的期待;另一方面,恐怕对网络法也缺乏深入理解,把本来可供法律人思考和探究的算法规制问题,推给了算法本身以及算法开发人员,用"透明""公开""开放"这样的大词来构造自己的理论。说到底,算法所引发的法律问题,无论在私法还是公法领域,都要求法律人在侵权法中的第三方责任理论、注意义务理论、因果关系理论、行政法中的正当程序理论、问责理论和法经济学中的成本收益分析等法学理论框架下,甚至在更广阔的社会科学理论框架下,来讨论类型化的应对,并借此尝试提出新的理论洞见。

一个多世纪前,工业事故危机引发了美国法律制度的大变革,包括霍姆斯在内的诸多美国法学家参与了这一进程,向美国的法律体系引入和构建了侵权法、事故法和保险法体

〔62〕 参见戴昕:《超越"马法"? ——网络法研究的理论推进》,《地方立法研究》2019 年第 4 期,第 1 - 17 页。

环球法律评论 GLOBAL LAW REVIEW

系,当时的许多理念和制度直至现在依然屹立不倒。〔63〕 现如今的算法规制危机,在某种程度上,也是向法律人开启的一个契机——这同样是法律人面对一个相对开放的领域,一个充满可能性的历史时刻。而正像本文所揭示的,实用主义进路,更可能帮助法律人跳出算法透明原则的迷思,更可能找到可以传世的理念和制度。

［**Abstract**］ In recent years, with the emergence of the problems of algorithms, people begin to think about the way to regulate algorithms. Among those thoughts, the Algorithmic Transparency Principle is, practically and theoretically, a well-known principle. It has attracted a significant number of scholars in recent years. However, as *ex ante* regulation, as opposed to *ex post* regulation, the Algorithmic Transparency Principle has its inherent limitations. Although the Algorithmic Transparency Principle can alleviate some "black box" problems, in the age of massive legislative, administrative and judicial regulation on algorithms, the Algorithmic Transparency Principle is normally neither feasible, nor necessary. As a result, the reasonable role of the Algorithmic Transparency Principle in algorithm regulation pedigree should be a non-universal and subsidiary one. Based on the critical analysis of the Algorithmic Transparency Principle, this article considers the reconstruction of the theory for algorithm regulation, and further argues that in contrast to essentialism-driven *ex ante* regulation such as algorithmic transparency, the pragmatism-driven *ex post* regulation such as algorithmic accountability should be a more appropriate regulatory strategy.

（责任编辑:姚　佳）

环球法律评论 GLOBAL LAW REVIEW

〔63〕 参见［美］约翰·法比安·维特著:《事故共和国》,田雷译,上海三联出版社 2013 年版,第 6－9 页。