# 人工智能算法黑箱的法律规制

## ——以智能投顾为例展开

徐　凤*

**内容摘要**：人工智能算法不公开、不透明，被称为“算法黑箱”。面对算法黑箱，不少人主张和呼吁算法透明。但绝对的透明是不存在的，即使透明也是相对的透明。换言之，在人工智能时代，算法的不公开是原则，公开才是例外。尽管如此，人们应有权要求算法公平。算法透明追求的其实是算法的简要说明，包括算法的假设和限制、算法的逻辑、算法的种类、算法的功能、算法的设计者、算法的风险、算法的重大变化等方面。算法透明的具体方法除了公开披露之外，还可以有诸如算法备案、算法解释权等替代工具，还应有算法审查、评估与测试，算法治理、第三方监管等保障算法公平的其他措施。

**关键词**：人工智能算法算法　黑箱　智能投顾

**中图分类号**：DF0　　**文献标识码**：A　　**文章编号**：1674- 4039-（2019）06- 0078- 86

DOI:10.19404/j.cnki.dffx.2019.06.002

算法〔1〕是人工智能的基础。“算法就是一系列指令，告诉计算机该做什么。”〔2〕“算法的核心就是按照设定程序运行以期获得理想结果的一套指令。”〔3〕所有的算法都包括以下几个共同的基本特征：输入、输出、明确性、有限性、有效性。〔4〕算法因数学而起，但现代算法的应用范畴早已超出了数学计算的范围，已经与每个人的生活息息相关。因此，“我们生活在算法的时代”。〔5〕随着人工智能时代的到来，算法越来越多地支配着我们的生活，也给现存的法律制度和法律秩序带来了冲击和挑战。人工智能算法不公开、不透明，被称为“算法黑箱”。这是人工智能给人类社会带来的重大新型问题之一。〔6〕法律制度如何应对“算法黑箱”的挑战？法律如何规制算法？这是法学研究必须面对的现实问题。人工智能的应用场景很多，笔者主要以智能投顾为例，来阐述算法黑箱的法律规制。相信它将对人工智能的其他应用场景提供有益的借鉴和启迪。

*中国政法大学外国语学院副教授。

本文系2018年度教育部人文社会科学研究青年基金项目（项目批准号：18YJC820072）的阶段性研究成果。

〔1〕算法（algorithm）一词来源于中世纪的拉丁语“algorism”。公元9世纪，波斯的一位数学家Al-Khwarizmi，他写了一本关于代数的著作。中世纪的学者用拉丁语传播Al-Khwarizmi的学说时，他的名字的拉丁语音译为“algorism”。到了18世纪，algorism演变成了algorithm。这个词汇就成了任何程序化运算或自动运算方法的统称。参见[美]克里斯托弗·斯坦纳：《算法帝国》，李筱莹译，人民邮电出版社2014年版，第42—43页。

〔2〕[美]佩德罗·多明戈斯：《终极算法：机器学习和人工智能如何重塑世界》，黄芳萍译，中信出版集团2017年版，第3页。

〔3〕参见前引〔1〕，斯坦纳书，第42页。

〔4〕徐恪、李沁：《算法统治世界——智能经济的隐形秩序》，清华大学出版社2017年版，第11页。

〔5〕参见前引〔2〕，多明戈斯书，第3页。

〔6〕邢会强：《人工智能时代的金融监管变革》，《探索与争鸣》2018年第10期，第21页。

## 一、算法透明之争

“黑箱”是控制论中的概念。作为一种隐喻，它指的是那些不为人知的不能打开、不能从外部直接观察其内部状态的系统。[7]人工智能所依赖的深度学习技术就是一个“黑箱”。深度学习是由计算机直接从事物原始特征出发，自动学习和生成高级的认知结果。在人工智能系统输入的数据和其输出的结果之间，存在着人们无法洞悉的“隐层”，这就是“算法黑箱”。[8]对透明的追求使人心理安定，“黑箱”使人恐惧。如何规制算法“黑箱”，算法是否要透明，如何透明，是法律规制遇到的首要问题。

（一）对算法透明的呼吁及其理由

面对算法黑箱，不少人主张、呼吁算法透明。总结其理由，主要有以下几点：

第一，算法透明是消费者知情权的组成部分。这种观点主张，因为算法的复杂性和专业性，人工智能具体应用领域中的信息不对称可能会更加严重，算法透明应是消费者知情权的组成部分。

第二，算法透明有助于缓解这种信息不对称。这种观点主张，算法的信息不对称加重不只发生在消费者与算法设计者、使用者之间，更发生在人类和机器之间，算法透明有助于缓解这种信息不对称。

第三，算法透明有助于防止人为不当干预。这种观点以智能投顾为例，认为算法模型是公开的，在双方约定投资策略的前提下，执行策略由时间和事件函数共同触发，执行则由计算机程序自动完成，避免了人为不当干预的风险，它比人为干预更加公平、公开和公正。[9]

第四，算法透明有助于防止利益冲突。这种观点以智能投顾为例，认为由于算法的非公开性和复杂性，智能投顾给出的资产配置建议有可能是推荐了与其自身利益高度攸关的产品，这就难以保证投资建议的独立性和客观性。[10]智能投顾可以通过对于推荐产品选项的特殊排列方式，把对自己最有利的产品排在最容易被选择到的位置。只有算法透明，才能防止这种利益冲突。

第五，算法透明有助于防范信息茧房。这种观点认为，算法可能形成信息茧房。算法科学的外表容易误导投资者，强化投资者的偏见，从而导致错误决策。算法技术为原本和普罗众生疏离的复杂难懂的金融披上了简单易懂的面纱，金融的高风险性被成功掩盖，轻松化的人机交互界面掩盖了金融风险的残酷本质。[11]

第六，算法透明有助于打破技术中立的外衣。智能金融给人以中立的感觉，而事实上，技术的背后是人。人类会将人性弱点和道德缺陷带进和嵌入算法之中，但它们却可能隐蔽于算法背后，从而更不易被发觉。

第七，算法透明有助于打破算法歧视。宾夕法尼亚州法学院的Tom Baker和荷兰鹿特丹伊拉斯谟大学的Benedict G.、C. Dellaert教授认为：公众不能预设智能投顾机器人没有人类所具有的不纯动机。因为智能金融算法存在歧视和黑箱现象，因此才需要算法的透明性或解释性机制。[12]

第八，算法透明有助于打破“算法监狱”与“算法暴政”。在人工智能时代，商业企业和公权部门都采用人工智能算法作出的自动化决策，算法存在的缺陷和偏见可能会使得大量的客户不能获得贷款、保险、承租房屋等服务，这如同被囚禁在“算法监狱”。然而，如果自动化决策的算法不透明、不接

〔7〕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50页。

〔8〕许可：《人工智能的算法黑箱与数据正义》，《社会科学报》2018年3月29日，第6版；兰亚妮、郑晋鸣：《让人工智能更有温度 》，《光明日报》2019年1月28日，第4版。

〔9〕参见宋湘燕、王韬：《机器人投顾——金融投资领域的新角色》，《金融时报》2016年5月9日第11版。

〔10〕参见伍旭川：《迎接金融科技的新风口——智能投顾》，《清华金融评论》2017年第10期，第87页。

〔11〕参见高丝敏：《智能投资顾问模式中的主体识别和义务设定》，《法学研究》2018年第5期，第43页。

〔12〕参见刘元兴：《智能金融的“算法可解释性”问题》，《金融科技观察》2018年第13期，第1页。

受人们的质询、不提供任何解释、不对客户或相对人进行救济，客户或相对人无从知晓自动化决策的原因，自动化决策就会缺少“改正”的机会，这种情况就属于“算法暴政”。〔13〕算法透明则有助于打破“算法监狱”与“算法暴政”。

第九，算法透明是提供算法可责性问题的解决工具和前提。有学者认为算法透明性和可解释性是解决算法可归责性的重要工具。明确算法决策的主体性、因果性或相关性，是确定和分配算法责任的前提。〔14〕

第十，算法透明有助于提高人们的参与度，确保质疑精神。这种观点认为，在卡夫卡的环境中，如果你不了解某个决定的形成过程，就难以提出反对的理由。〔15〕由于人们无法看清其中的规则和决定过程，人们无法提出不同的意见，也不能参与决策的过程，只能接受最终的结果。为走出这一困境，算法透明是必要的。还有人认为，质疑精神是人类前进的工具，如果没有质疑，就没有社会进步。为了保证人类的质疑，算法必须公开——除非有更强的不公开的理由，比如保护国家安全或个人隐私。

第十一，公开透明是确保人工智能研发、涉及、应用不偏离正确轨道的关键。这种观点认为，人工智能的发展一日千里，人工智能可能拥有超越人类的超级优势，甚至可能产生灾难性风险，因而应该坚持公开透明原则，将人工智能的研发、设计和应用置于监管机构、伦理委员会以及社会公众的监督之下，确保人工智能机器人处于可理解、可解释、可预测状态。〔16〕

（二）反对算法透明的声音及其理由

声音并非一边倒。反对算法透明的声音也不少，其主要理由如下：

第一，类比征信评分系统。征信评分系统不对外公开是国际惯例，其目的是防止“炒信”、“刷信”，使评级结果失真。很多人工智能系统类似于信用评级系统。

第二，周边定律。周边定律是指法律无须要求律师提请我们注意身边具有法律意义的内容，而是将其直接植入我们的设备和周边环境之中，并由这些设备和环境付诸实施。〔17〕主张该观点的人宣称，人类正在步入技术对人类的理解越来越深刻而人类却无须理解技术的时代。智能时代的设备、程序，就像我们的人体器官和中枢神经系统，我们对其知之甚少但却可以使用它们。同样，算法为自我管理、自我配置与自我优化而完成的自动计算活动，也无须用户的任何体力与智力投入。〔18〕

第三，算法不透明有助于减少麻烦。如果披露了算法，则可能会引起社会舆论的哗然反应，从而干扰算法的设计，降低预测的准确性。大数据预测尽管准确的概率较高，但也不能做到百分之百。换言之，大数据预测也会不准，也会失误。如果将算法公之于众，人们对预测错误的赋值权重就有可能偏大，从而会阻碍技术的发展。

第四，防止算法趋同。算法披露之后，好的算法、收益率高的算法、行业领导者的算法可能会引起业界的效仿，从而会出现“羊群效应”，加大顺周期的风险。

第五，信息过载或难以理解。算法属于计算机语言，不属于日常语言，即使对外披露了，除专业人士之外的大多数客户难以理解。换言之，对外披露的信息对于大多数用户来讲可能属于无效信息。〔19〕

第六，偏见存在于人类决策的方方面面，要求算法满足高于人类的标准是不合理的。〔20〕算法透明性本身并不能解决固有的偏见问题。〔21〕要求算法的透明性或者可解释性，将会减损已申请专利的软

〔13〕张凌寒：《商业自动化决策的算法解释权研究》，《法律科学》2018年第3期，第66页。

〔14〕Finale Doshi-Velez & Mason Kortz, Accountability of AI Under the Law: The Role of Explanation, https://arxiv.org/pdf/1711.01134.pdf.

〔15〕参见[美]卢克·多梅尔：《算法时代：新经济的引擎》，胡小锐、钟毅译，中信出版集团2016年版，第140页。

〔16〕金东寒主编：《秩序的重构——人工智能与人类社会》，上海大学出版社2017年版，第72页。

〔17〕参见前引〔15〕，多梅尔书，第123页。

〔18〕参见前引〔15〕，多梅尔书，第123页。

〔19〕参见前引〔7〕，张淑玲文，第51页。

〔20〕Joshua New and Daniel Castro：《算法可解释性与算法监管》，姜开锋译，大数据和人工智能法律研究院公众号，2018年7月3日。

〔21〕同上文。

件的价值。[22]要求算法的透明性还为动机不良者扰乱系统和利用算法驱动的平台提供了机会，它将使动机不良者更容易操纵算法。

第七，算法披露在现实中存在操作困难。智能投顾可能涉及多个算法，披露哪个或哪些算法？算法披露到什么程度？

（三）折中派的观点及其理由

有人认为，算法是一种商业秘密。“算法由编程者设计，进而给网站带来巨大的商业价值，因此其本质上是具有商业秘密属性的智力财产。”[23]如果将自己的专有算法程序公之于众，则有可能泄漏商业秘密，使自己丧失技术竞争优势。鉴于很多算法属于涉及商业利益的专有算法，受知识产权法保护，因此即使是强制要求算法透明，也只能是有限度的透明。

还有人认为，如何对待算法，这个问题并没有“一刀切”的答案。在某些情况下，增加透明度似乎是一个正确的做法，它有助于帮助公众了解决策是如何形成的，但是在涉及国家安全时，公开源代码的做法就不适用了，因为一旦公开了特定黑盒子的内部运行机制，某些人就可以绕开保密系统，使算法失效。[24]

（四）笔者的观点——算法的不公开是原则，公开是例外

绝对的透明是不存在的，即使透明也是相对的透明。在历史上，人类社会随着复杂性的增加，不可避免地产生以组织和技术形态出现的各类“黑箱”，它们的决定虽然影响着社会公众的利益，但仍然保持着某种程度的秘密性。[25]为了克服信息不对称带来的各种弊端，法律作出了各种回应，包括设计出某种程度的信息公开和透明化机制。例如上市公司强制信息披露等，以保障相关当事人和社会大众的知情权，避免恐慌，让社会大众保持一定程度的预测能力。[26]但是，尽管如此，上市公司就绝对透明了吗？事实上，绝对透明是做不到的。信息披露是有成本的，投资者的知情权也是需要保障的。为了平衡这种冲突，法律发展出了信息的“重大性”标准，只有符合这一标准的信息才应予披露，而不是所有的信息才能披露。[27]那么，在算法披露领域，是否要借鉴上市公司的信息“重大性”标准呢？如果要借鉴，算法的透明就是一种有限的透明。而且，就信息的“重大性”标准而言，实践中和学术界仍有“价格敏感重大性”和“投资决策重大性”之争。算法透明如果要借鉴，该标准该如何设定呢？这也是一个难题。

鉴于算法透明的利多于或大于弊，我们支持算法有限透明的立场。算法的完全透明是做不到的。在前人工智能时代，也有各种各样的算法，这些算法也在影响着人们的生活，但人们并未要求其完全公开。在人工智能时代，也可以做这样的推理：算法的不公开是原则，公开是例外。如果需要公开，也需要制定法律明确哪些算法应该公开，以及如何公开。

笔者认为，具有垄断地位的算法，或国家财政资金提供支持的、目的是提供普遍公共服务的算法，人们应有权要求其公开。因为具有垄断地位的算法限制了人们的选择权，对个人施加的影响巨大，人们应有知情权。而国家财政资金提供支持的、目的是提供普遍公共服务的算法，之所以需要公开，是因为这是纳税人知情权的组成部分。此外，对于歧视某一类人、侵犯公民平等权的算法，尽管它未必需要向社会公开，但人们有权提起诉讼，让其接受法官的审查。这是因为，从理论上说，私人的商业秘密作为个人利益，在涉嫌侵犯个人的权利时，是不能对抗法官的审查权的。

尽管算法不公开是原则，但人们应有权要求公平也是原则。美国《公平信用报告法》确保消费者可以看到某个数据档案对信用评分的影响，并且有权纠正档案中包括的任何错误，而《平等信用机会法》则禁止在信用评分中纳入种族或性别歧视。这种做法值得我国借鉴，即我国法律即使不要求算法

〔22〕前引〔20〕，Joshua New and Daniel Castro文。

〔23〕张凌寒：《风险防范下算法的监管路径研究》，《交大法学》2018年第4期，第56页。

〔24〕参见前引〔15〕，多梅尔书，第222页。

〔25〕参见胡凌：《人工智能的法律想象》，《文化纵横》2017年第2期，第111页。

〔26〕参见前引〔6〕，邢会强文，第115页。

〔27〕参见刘东辉：《谁是理性的投资者——美国证券法上重大性标准的演变》，《证券法律评论》2015年卷，第78页。

公开，但也应要求算法公平，并将其置于法官的审查之下。

## 二、算法透明的实践与内容

人们呼吁算法透明，却往往忽略了算法透明的具体内容，这将使算法透明的呼吁停留于表面上，而不具有现实的可操作性。揆诸当下各国的探索与实践，可以发现，算法透明的具体内容还没有真正付诸实施。

### （一）算法透明的探索与实践

2017年，美国计算机学会公众政策委员会公布了6项算法治理指导原则。第一个原则是知情原则，即算法设计者、架构师、控制方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以及可能对个人和社会造成的潜在危害。第二个原则是质询和申诉原则，即监管部门应该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疑并申诉的权力。第三个原则是算法责任认定原则。第四个原则是解释原则，即采用算法自动化决策的机构有义务解释算法运行原理以及算法具体决策结果。第五个原则是数据来源披露原则。第六个原则是可审计原则。仔细审视这6项原则，其要求的算法透明的具体内容主要是算法的偏见与危害、算法运行原理以及算法具体决策结果，以及数据来源。

发生于美国威斯康星州的State v. Loomis案所引发了美国社会关于算法透明的争论。在该案中，该州一法院使用“再犯风险评估内容”来进行量刑，被告Loomis认为法官违反了正当程序原则，他有权检查法律的算法，量刑法院应该公开算法。但该州最高法院认为，算法只是量刑的一个因素，而不是唯一因素，算法量刑没有违反正当程序原则，但法官应向被告解释其在作出量刑决定时所考量的因素并提醒法官警惕使用算法量刑可能带来的偏见。总之，在该案中，该州最高法院倾向于保护算法产品厂商的商业秘密，不会要求公开算法代码，也没有要求厂商用自然语言解释算法的设计原理、功能和目的。

2017年年底，纽约州通过一项《算法问责法案》要求成立一个由自动化决策系统专家和相应的公民组织代表组成的工作组，专门监督自动化决策算法的公平和透明。之前，该法案有一个更彻底的版本，规定市政机构要公布所有用于“追踪服务”或“对人施加惩罚或维护治安”的算法的源代码，并让它们接受公众的“自我测试”。“这是一份精炼的、引人入胜的、而且是富有雄心的法案”，它提议每当市政府机构打算使用自动化系统来配置警务、处罚或者服务时，该机构应将源代码——系统的内部运行方式——向公众开放。很快，人们发现这个版本的法案是一个很难成功的方案，他们希望不要进展得那么激进。因此，最终通过的法案删去了原始草案中的披露要求，设立了一个事实调查工作组来代替有关披露的提议，原始草案中的要求仅在最终版本里有一处间接地提及——“在适当的情况下，技术信息应当向公众开放”。〔28〕

在欧盟，《通用数据保护条例》（GDPR）在鉴于条款第71条规定：“在任何情况下，该等处理应该采取适当的保障，包括向数据主体提供具体信息，以及获得人为干预的权利，以表达数据主体的观点，在评估后获得决定解释权并质疑该决定。”据此，有人主张GDPR赋予了人们算法解释权。〔29〕但也有学者认为，这种看法很牵强，个人的可解释权并不成立。〔30〕

我国《新一代人工智能发展规划》指出：“建立健全公开透明的人工智能监管体系。”这提出了人

〔28〕［美］Julia Powles：《纽约市尝试对算法问责——政策有待完善，但行动敢为人先》，姜开锋译，大数据和人工智能法律研究院公众号，2018年12月29日。

〔29〕参见前引〔23〕，张凌寒文，第58页。

〔30〕参见刘元兴：《智能金融的“算法可解释性”问题》，《金融科技观察》2018年第13期，第2页。

工智能监管体系的透明，而没有要求算法本身的透明。

（二）算法透明的内容不是算法代码而是算法说明

人们呼吁算法透明，但透明的内容具体是算法的源代码，还是算法的简要说明？秉承“算法公开是例外，不公开是原则”的立场，即使是在算法需要公开的场合，也需要考察算法公开的具体内容是什么。

算法的披露应以保护用户权利为必要。算法的源代码、算法的具体编程公式（实际上也不存在这样的编程公式）是不能公开的。这主要是因为算法的源代码一方面非常复杂，且不断迭代升级，甚至不可追溯，无法予以披露；另一方面，公开源代码是专业术语，绝大部分客户看不懂，即使公开了也没有意义。

算法的透明追求的是算法的简要说明（简称算法简介）。算法的简介包括算法的假设和限制、算法的逻辑、算法的种类、算法的功能、算法的设计者、算法的风险、算法的重大变化等。算法简介的公开，也是需要有法律规定的，否则，不公开仍是基本原则。例如，美国《智能投顾升级指导意见》规定的与算法相关的披露内容包括：管理客户账户所使用的算法的说明；算法功能的介绍（如通过算法能对客户个人账户进行投资和重新调整）；算法的假设和限制（如该算法是基于现代投资组合理论，说明背后的假设和该理论的局限性）；对使用算法管理客户账户所固有的特定风险的描述（例如该算法可能不考虑市场条件而重新调整客户账户，或者进行比客户预期更频繁地调整以及算法可能无法应对市场条件的长期变化）；任何可能导致用于管理客户账户的智能投顾算法重写的状况描述（如智能投顾可能在紧张的市场状况下停止交易或采取其他临时性防御措施）；关于第三方参与管理客户账户的算法的开发、管理或所有权的说明，包括对这种安排可能产生的任何冲突利益的解释（例如，如果第三方以打折的方式向智能投顾方提供算法，那么此算法同样可能会将客户引导到一种能使第三方获利的产品上）。还如，新加坡金融管理局希望数字顾问可以书面向客户披露算法相关信息：首先，算法的假设、限制和风险；其次，明确数字顾问可以推翻算法或者暂停数字顾问的情形；再次，披露对算法的任何重大调整。总之，该指导意见要求披露的是对算法的说明而不是算法本身。

## 三、算法透明的替代或辅助方法

算法透明不能简单类比上市公司的透明，算法透明的具体方法除了公开披露之外，还可以有其他替代方法。当然它们也可以成为辅助方法。这些方法究竟是替代方法还是辅助方法，取决于立法者的决断。

（一）备案或注册

备案即要求义务人向监管机构或自律组织备案其算法或算法逻辑，算法或算法逻辑不向社会公开，但监管机构或自律组织应知悉。这种观点认为，智能投顾应向监管部门备案其算法（逻辑），监管部门应对智能投顾的算法进行大致分类，并采取必要的措施避免同质化，以免造成羊群效应。此外，还要明确要求智能投顾定期检查模型或算法的有效性，一旦有重要修改，应再次备案。〔31〕

其实，在程序化交易、量化交易或高频交易领域，备案已是常规做法。欧洲证券市场监管局要求从事量化交易的投资机构每年向其报备交易策略、交易参数的设定及其限制、核心风险控制模块构成及交易系统测试结果。〔32〕之所以如此，是因为算法备案一方面促使量化交易投资机构更为谨慎地使用和监控算法交易系统；另一方面也有助于促使监管机构掌握前沿的技术，以便更好地理解和评估算法交易系统，从而有助于改善和提高监管机构的监管能力。

〔31〕姜海燕、吴长凤：《智能投顾的发展现状及监管建议》，《证券市场导报》2016年第12期，第10页。

〔32〕金小野：《规范高频交易是国际证券业监管焦点》，《法制日报》2013年11月12日，第10版。

算法很复杂，很难用公式或可见的形式表达出来。算法的种类很多，一个人工智能系统可能会涉及很多算法，且算法也在不断迭代、更新和打补丁，就像其他软件系统不断更新一样。因此，算法本身没法备案，更无法披露。可以备案和披露的是算法的逻辑和参数。《关于规范金融机构资产管理业务的指导意见》（以下简称《资管新规》）第23条要求："金融机构应当向金融监督管理部门报备人工智能模型的主要参数以及资产配置的主要逻辑。"即是因为如此。但是，算法逻辑和主要参数的披露却可能引起业界的纷纷效仿，从而可能带来羊群效应。也正因为如此，算法逻辑和主要参数的备案，需要对金融监督管理部门及其工作人员课加严格的保密责任。

除了算法逻辑的备案以外，还可以要求算法开发设计人员的注册。2017年1月，SEC批准了对NASD规则1032（f）的修正案，该修正案扩大了需要注册为证券交易者的人员范围。具体而言，自2017年1月30日起，每个主要负责设计、开发或重大修改与股票、优先股或可转换债券有关的算法交易策略的人，或在上述活动中负责日常监管或指导的人，必须通过57系列考试并注册为证券交易者。美国自律监管组织——金融服务监管局的目标是确保公司识别并注册一个或多个相关人员，他具备交易策略（例如，套利策略）及实施该交易策略的技术实施（例如编码）的知识并对此负责，以便公司来评估相关产品的结果是否实现了其业务目标，且是否是合规的。如果智能投顾不是自行设计和开发算法，而是委托第三方设计和开发算法，则该第三方的设计开发机构中主要负责设计、开发或重大修改与股票、优先股或可转换债券有关的算法交易策略的人，也必须注册为证券交易者。这些经验也值得我国借鉴。

（二）算法可解释权

一旦人工智能系统被用于作出影响人们生活的决策，人们就有必要了解人工智能是如何作出这些决策的。方法之一是提供解释说明，包括提供人工智能系统如何运行以及如何与数据进行交互的背景信息。但仅发布人工智能系统的算法很难实现有意义的透明，因为诸如深度神经网络之类的最新的人工智能技术通常是没有任何算法输出可以帮助人们了解系统所发现的细微模式。[33]基于此，一些机构正在开发建立有意义的透明的最佳实践规范，包括以更易理解的方法、算法或模型来代替那些过于复杂且难以解释的方法。笔者认为，是否赋予客户以算法可解释权有待深入论证，但算法设计者有义务向公权机关解释算法的逻辑。

## 四、算法公平的保障措施

算法公开、算法备案等规制工具都属于信息规制工具，它们是形式性的规制工具。除了信息规制工具之外，还有其他实质性规制工具。形式性规制工具追求的价值目标是形式公平，实质性规制工具追求的价值目标是实质公平。在消费者权益和投资者权益保护过程中，除了保障形式公平之外，也要保障实质公平。因此，除了信息规制工具之外，还应有保障算法公平的其他实质性规制工具，这些工具主要包括三个方面，一是算法审查、评估与测试，二是算法治理，三是第三方监管。

（一）算法审查、测试与检测

在人工智能时代，算法主导着人们的生活。数据应用助推数据经济，但也有许多模型把人类的偏见、误解和偏爱编入了软件系统，而这些系统正日益在更大程度上操控着我们的生活。"这些数学模型像上帝一样隐晦不明，只有该领域最高级别的牧师，即那些数学家和计算机科学家才明白该模型是如何运作的。"[34]人们对模型得出的结论毫无争议，从不上诉，即使结论是错误的或是有害的。凯西·奥尼尔将其称为"数学杀伤性武器"。"算法就是上帝，数学杀伤性武器的裁决就是上帝的指令。"[35]

〔33〕参见［美］施博德、沈向洋：《未来计算》，北京大学出版社2018年版，第39页。

〔34〕［美］凯西·奥尼尔：《算法霸权——数学杀伤性武器的威胁》，马青玲译，中信出版社2018年版，前言第Ⅴ页。

〔35〕前引〔34〕，奥尼尔书，前言第Ⅹ—Ⅺ页。

然而，数学家和计算机科学家毕竟不是上帝，他们应当接受社会的审查。算法是人类的工具，而不是人类的主人。数学家和计算机科学家是人类的一员，他们应与我们普罗大众处于平等的地位，而不应凌驾于人类之上，他们不应是人类的统治者。即使是人类的统治者——君主或总统，在现代社会也应接受法律的规范和治理、人民的监督和制约，更何况群体庞大的数学家和计算机科学家。总之，算法应该接受审查。

算法黑箱吸入数据，吐出结论，其公平性应接受人类的审查。算法的开发者、设计者也有义务确保算法的公平性。在智能投顾领域，作为智能投顾的核心要素，算法利用大数据，基于各种模型和假设，将有关数据转化为适合特定投资者的投资建议。如果算法的设计有问题，则算法输出的结果可能会产生有较大偏差甚至是错误的结果，无法实现客户预期的投资目标。因此，有必要审查算法的有效性。这在国外已有实践。例如，美国对智能投顾算法的审查包括初步审查和持续审查。初步审查包括评估数字咨询工具使用的前提假设和相关方法是否适合特定的目标，评估系统输出是否符合公司的预期目标等；持续审查包括评估数字化建议工具使用的模型是否适用于持续变化的市场等。〔36〕澳大利亚明确规定智能投顾要有测试文档，说明算法测试的计划、案例、结果、缺陷及解决方法。目前，我国在智能投顾算法方面尚未建立起完整的监督和测试框架。因此，有必要借鉴发达市场的成熟经验，尽快填补智能投顾算法的监管空白。具体内容可包括：第一，智能投顾平台自身应充分理解算法使用的假设、投资者的偏好、模型以及算法的局限性；第二，为算法的设计、开发和运行建档，以便监管部门对算法进行检查和监督；第三，对算法是否适合特定的投资目标、是否符合客户预期进行测试和评估；第四，对算法的更新迭代进行严格监测。〔37〕

应该对人工智能系统进行测试。人工智能机器人目前尚未成为独立的民事主体，不能独立承担民事责任，但这并不妨碍对其颁发合格证书和营运证书。这正如汽车可以获得行驶证书和营运许可证书一样。自然人投资顾问参加资格考试，本质上就是对其“从业资格所需要的知识图谱”进行抽样评测。〔38〕智能投顾也可以参加“从业资格”考试。智能投顾参加从业资格考试，本质上是评测其知识图谱是否具有投资顾问服务所要求的功能。智能投顾参加从业资格考试的形式就是要对智能投顾系统和算法进行测试。2016年8月，韩国金融委员会出台了“机器人投顾测试床的基本运行方案”，通过三阶段的审核程序检验机器人投顾平台的实际运营情况，测试算法的稳定性、收益性和整体系统的安全性。〔39〕总之，我们可以通过评测智能投顾系统的知识图谱来判断它是否具备“从业资格”。智能投顾应该是可以被“评测的”。〔40〕只有检测合格的智能投顾才能投入市场，从事服务。

监测对于防范算法风险必不可少。由于智能投顾可能造成系统性金融风险，对此可以采取宏观审慎管理措施来监测，例如观测智能投顾是否存在大规模的一致性行为和协同行为等。〔41〕运用专业技术对算法的执行进行持续监测。如果发现算法存在严重错误，应及时中止系统服务，并采取有效措施予以纠正。〔42〕我国的《资管新规》也有这方面的要求。尤其是在智能投顾发展初期，需要审慎评估智能投顾对证券市场的影响，密切监测并加强对智能投顾算法的一贯性、中立性、合法性和安全性等方面的监管力度，及时跟踪市场变化，防范市场系统性金融风险。〔43〕

（二）算法治理

智能投顾应强化对智能投顾算法的组织管理。欧盟金融工具市场指令（MiFID II）要求，一家投资

〔36〕李苗苗、王亮：《智能投顾：优势、障碍与破解对策》，《南方金融》2017年第12期，第80页。

〔37〕前引〔36〕，李苗苗、王亮文，第80页。

〔38〕张家林：《人工智能投顾，需要从业资格考试吗》，《华夏时报》2016年12月5日，第34版。

〔39〕姜海燕、吴长凤：《机器人投顾领跑资管创新》，《清华金融评论》2016年第12期，第100页。

〔40〕前引〔38〕，张家林文。

〔41〕张家林：《人工智能投顾：21世纪的技术对应21世纪的监管》，《证券日报》2017年1月21日，第A03版。

〔42〕前引〔36〕，李苗苗、王亮文。

〔43〕参见前引〔31〕，姜海燕、吴长凤文，第10页。

公司应该确保其负责算法交易风险和合规的员工具有:(1) 充足的算法交易和交易策略知识;(2)跟踪自动警报所提供信息的能力;(3)算法交易造成交易环境紊乱或有疑似市场滥用时,有足够的权力去质疑负责算法交易的员工。[44]在澳大利亚,2016年8月正式发布《RG255:向零售客户提供数字金融产品建议》指南要求,智能投顾被许可人应确保业务人员中至少有一位了解用于提供数字建议技术和算法基本原理、风险和规则的人,至少有一位有能力检查数字建议的人,定期检查算法生成的数字建议"质量"。[45]质疑精神是人类社会前进的基本动力,必须将算法置于人类的质疑和掌控之下。人工智能的开发者和运营者应有能力理解和控制人工智能系统,而不能单纯地一味依赖于第三方软件开发者。

人工智能系统还应建立强大的反馈机制,以便用户轻松报告遇到的性能问题。[46]任何系统都需要不断迭代和优化,只有建立反馈机制,才能更好地不断改进该系统。

(三)加强第三方算法监管力量

为了保证对算法权力的全方位监督,应支持学术性组织和非营利机构的适当介入,加强第三方监管力量。目前在德国已经出现了由技术专家和资深媒体人挑头成立的名为"监控算法"的非营利组织,宗旨是评估并监控影响公共生活的算法决策过程。具体的监管手段包括审核访问协议的严密性、商定数字管理的道德准则、任命专人监管信息、在线跟踪个人信息再次使用的情况,允许用户不提供个人数据、为数据访问设置时间轴、未经同意不得将数据转卖给第三方等。[47]这种做法值得我国借鉴。为了让人工智能算法去除偏私,在设计算法时,对相关主题具有专业知识的人(例如,对信用评分人工智能系统具有消费者信用专业知识的人员)应该参与人工智能的设计过程和决策部署。[48]当人工智能系统被用于作出与人相关的决定时,应让相关领域的专家参与设计和运行。[49]

Abstract: Artificial intelligence algorithm is not disclosed or transparent, and that's why we called it "the algorithm black box". Confronted with the algorithm black box, many people advocate and call for transparency of the algorithm. However, absolute transparency does not exist. Even if it is transparent, it is only relatively transparent. In other words, in the era of artificial intelligence, non-disclosure of algorithms is normal, while disclosure is exceptional. Nevertheless, people should have the right to demand fairness in algorithms. The transparency of algorithms actually comprises a brief description of the algorithm, including the assumptions and limitations of the algorithm, the logic of the algorithm, the types of the algorithm, the functions of the algorithm, the designer of the algorithm, the risks of the algorithm, and the significant changes in the algorithm. In addition to disclosure, there are alternative tools such as algorithmic filing, algorithmic interpretation right and so on, and other measures to ensure algorithm fairness, such as algorithm review, evaluation and testing, algorithm governance, third-party supervision, etc.

Key words: artificial intelligence; algorithm; the algorithm black box; robo-advisor

〔44〕参见网页http://ec.europa.eu/finance/securities/docs/isd/mifid/rts/160719-rts-6_en.pdf。

〔45〕RG255.64。

〔46〕[美]施博德、沈向洋:《未来计算》,北京大学出版社2018年版,第34页。

〔47〕参见前引〔7〕,张淑玲文,第53页。

〔48〕参见前引〔46〕,施博德、沈向洋书,第29页。

〔49〕参见同上书,第34页。