智能投顾算法透明的价值辩证、规制导向与实现路径

刘博涵

(贵州大学法学院,贵州 贵阳 550025)

摘 要:运用人工智能技术进行大数据分析的智能投顾能够更好地落实投资者适当性原则,为投资者生成契合度更高的投资组合方案,却也暗藏了掩饰金融机构利益掠夺、致使投资方案风险错配、架空投资顾问法定义务等负面效应。算法透明有助于打破算法暴政、消除算法歧视,为算法问责奠定基础,而算法保密又是维持国家科技实力、保持科技企业市场竞争力之必需。面对算法透明与算法保密的价值冲突,彻底的算法透明和绝对的算法保密都失之偏颇,应根据规制算法的不同目的,以不同方式对不同主体实现不同程度、不同层次的算法透明。有层次的算法透明是比例原则之利益衡量功能、"目的—手段"合比例分析范式在智能投顾算法治理中的具体应用。要实现智能投顾的算法透明,应在事前利用"合规审查+算法备案"形成算法模型透明以防控系统性金融风险,在事中利用"算法告知+算法解释"形成算法逻辑透明以保障投资者的知情权利,事后评估算法对投资者信赖利益的影响程度,明确算法决策的主体性、因果性及关联性,以确定和分配算法责任。

关键词:智能投顾;算法透明;算法保密;价值辩证;规制导向;实现路径

中图分类号:D 922.294;D 922.16 **文献标识码**:A **文章编号**:1000-260X(2024)01-0103-11

随着新一代信息网络技术的加速迭代,人工智能算法持续得到优化。在人工智能算法的加持下,一种由计算机根据现代投资组合理论,针对投资者的特征或者偏好,自动计算并提供投资组合配置建议的投资理财顾问——智能投顾(Robo-advisory)应运而生。智能投顾具有降低投资门槛、提高投资理性、满足投资者个性化需求等优势[1],但也在某种程度上削减了金融机构的 KYC(Know Your Customer)义务,架空了金融产品推介中的 KYP(Know Your Product)原则①。投资者可能面临沦为算法"奴隶"、遭受算法宰制的风险。对此,算法透明作为规制算法最广为人知的手段,在学界受到广泛推崇②。然而,算法透明意味着算法秘密性的灭失,这将导致算法控制者在市场竞争中丧失技术优势,对算法开发和创新产生逆向激励,不利于市场竞争机制的维护和算法技术进步。

面对算法透明和算法保密③的利益冲突,有学者认为数字时代是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,算法透明将带来算法趋同、信息过载和贬抑创新等负面效应[2];另有学

收稿日期:2023-10-07

基金项目:贵州大学人文社会科学一般项目"数字经济时代金融数据的法律保护研究"(GDYB2023017);贵州省哲学社会科学规划课题青年项目"数字金融算法的法律规制研究"(23GZQN37);国家社会科学基金重大项目"数字经济的刑事安全风险防范体系建构研究"(21&ZD210)

作者简介:刘博涵,法学博士,贵州大学法学院副教授、硕士生导师,主要从事数字法学、金融法学研究。

者认为,只有算法透明才能揭示隐藏于技术中立外衣下的算法歧视,确保算法开发、应用不偏离“以人为本”的轨道[3]。但在算法透明的具体实现进路上,由于人工智能算法应用具有广泛性,当前的研究大多关注相对宽泛的利益权衡领域,止步于规制算法保密的方向性指引层面。有鉴于此,本文聚焦于智能投顾算法透明与算法保密价值冲突的具体语境,针对智能投顾算法给投资者适当性制度造成的冲击,辩证算法透明与算法保密的利益平衡,提出有层次的算法透明方案,探求智能投顾算法透明的实现路径。

一、大数据金融智能投顾的负面效应

基于大数据分析的智能投顾相较于传统的自然人投资顾问掌握了更多、更全、更新的金融数据,同时拥有人力无法企及的数据处理能力,因而能够更有效地生成投资者画像、了解投资者的风险承受能力、识别适格投资者,进而更好地落实投资者适当性原则[4]。然而,当下新兴事物都是价值与风险的矛盾统一体,隐藏在智能投顾背后的算法绝非人们想象的那般理性、中立、客观。算法既不可能抽象于运算模型设计者的系统设定而存在,也不可能抽象于运算模型所关联的人而存在[5]。相反,智能投顾可能会将传统人工投顾意图规避法律而实施的一些违法操作植入算法,造成对投资者利益的持续掠夺。同时,智能投顾还存在投资者风险错配、法定义务架空等负面效应。

(一)算法掩饰对投资者利益的掠夺

一般来说,金融机构在使用智能投顾为投资者定制个性化的资产配置方案时,投资顾问服务费是其最主要的收入来源。以美国智能投顾领域的明星公司 Betterment 为例,该公司根据投资者账户余额计算年费,藉此向投资者收取服务费。公司的投顾服务费大致是账户余额的 0.15%~0.35%,投资金额越大,费率越低。具体而言,当投资者的账户余额低于 100 美元时,公司每月收取 3 美元的服务费;账户余额为 100~10000 美元时,公司每年收取账户余额 0.35%的服务费;账户余额为 10000~100000 美元时,公司每年收取账户余额的 0.25%作为服务费;如果投资者的账户余额超过 100000 美元,公司每年收取的服务费仅为账户余额的 0.15%。美国 Wealthfront 公司甚至对账户资金余额低于 10000 美元的投资者免收服务费,对高于 10000 美元的部分每年仅收取 0.25%的服务费。同时,投资者还可以通过邀请新客户加入的方式获得服务费减免和转账费用补偿,以进一步降低服务费用[6]。

与低廉的投资顾问服务费相比,智能投顾系统的研发成本和日常运营维护费用十分高昂,单纯收取投资顾问服务费常常无法维持智能投顾公司的运营,更难以盈利。此种业态势必“逼迫”智能投顾公司从其他的渠道“开源”。在我国,弥财、蓝海智投、摩羯智投等头部智能投顾平台都策略性地不收取投资顾问服务费或者收取很低的服务费,那么它们靠什么维持运营和盈利呢?这些平台背靠成熟的金融产品公司,主要依靠销售金融产品获利。具体而言,智能投顾公司通过算法设计,在为投资者生成资产配置方案时有意识地抓取销售附加值高的金融产品数据,提出具有强烈自益导向的投资咨询建议[7]。在这样的算法设计中,投资者利益和投资者适当性并不位于最高优先级,智能投顾公司的销售提成反而被优先考虑。看似客观的算法实际藏于黑箱之中,沦为掩饰智能投顾公司掠夺投资者利益的工具。

(二)投资者画像失真的风险错配

在大数据时代,无论个人的出生日期、学历背景、婚姻状况、职业情况等身份信息,还是虹膜、指纹、面部图像等生物信息,抑或出行方式、购物频次、消费水平、运动强度等行为信息,都在时时刻刻被采集与数据化。“在大数据俯瞰的视野下,我们每一个人无异于‘裸奔’”[8]。数字时代赋予了每个人生物和数字的双重属性,投资者画像便是投资者在现实世界中投资活动的数字化表达。通过问卷“读取”投资者的性别、年龄、家庭情况、收入状况、可投资资产、风险偏好、投资期限、投资目标等特异性个体信息,使用一定的算法将现实世界的投资者“还原”成数据意义上的投资者,这些都是实现智

能投顾的前提。只有生成精准的投资者画像,才可能根据投资者的风险承受能力和投资目标为其匹配最优的资产配置方案。正是基于这个原因,投资者画像被认为是数字时代实现投资者适当性 KYC 原则的最佳方式。

然而,即便智能投顾系统掌握了某个投资者当前的财务状况及其过往的一切投资活动,它对该投资者的投资偏好仍然可能存在误判,生成的投资组合方案仍然可能与该投资者的风险承受能力不匹配。这是因为投资者的风险承受能力可以分为客观承受能力和主观承受能力,而投资者的财务状况只能反映其对投资风险的客观承受能力[9]。将投资者的风险承受能力等同于其财务状况是对"风险承受能力"的重大误解。对投资者风险承受能力之主观要素的忽视是导致投资者画像失真进而造成投资组合方案风险错配的重要原因。此外,智能投顾的算法也不可避免地存在缺陷甚至错误,并不能保证每一次、针对每一位投资者都能生成适配度最高的投资组合方案。例如,人工智能中广泛应用的贪心算法(Greedy Algorithm)④在求解每一个子问题时总是使用当前的最优方法来实现,但这并不能保证最后的解也是最优的。

(三)架空约束投资顾问的法定义务

在传统的自然人投资顾问模式下,投资者与金融机构、投资顾问的关系是委托人与受托人的关系,金融机构、投资顾问对投资者承担信义义务[10]。如果投资顾问疏于履行合规义务、审慎义务、投资者适当性义务等法定义务进而导致投资者损失,将需要承担对投资者的赔偿责任。智能投顾的出现使金融投资顾问活动变成人与机器混合的过程,其中既有幕后技术人员根据金融从业人员的先验常识、专业判断、逻辑推理和利益考量设计的程序,也有台前机器根据投资者输入的信息与算法生成的投资组合方案。人工智能的介入很可能使其成为人类错误的"背锅侠"。

在当前的智能投顾业务中,投资者"输入"自己的资产状况、风险偏好和投资目标并由投顾系统"输出"投资组合方案的过程仅需金融机构工作人员的辅助即可完成,甚至可以完全由投资者自助完成。这一过程看似改变了传统自然人投资顾问与投资者一对一、面对面的投资咨询过程,实际上依然是投资咨询机构通过智能投顾的长臂在接触投资者、收集投资者信息、获知投资者需求。智能投顾系统只是金融机构延伸出去的长臂,是金融机构的电子代理人。如果法律主体使用电子代理人作出意思表示,其行为的法律后果应归诸于使用者[11]。而随着强人工智能时代⑤的到来,人工智能设备逐渐具备深度自主学习的能力,能够自主思考和行动,不受开发者预设的指令约束,进而超出人类可以预见和控制的范围。强人工智能下的智能投顾通过神经网络自主学习、经过算法处理海量数据后生成的投资决策是程序设计者和运营者也无法预测的。如果智能投顾通过自主学习生成与设定目标相左的投资方案进而造成投资者损失时,投资损害的因果关系极易归咎于技术风险而导致"买者自负",现行法律约束投资顾问的忠实义务、谨慎义务和信义义务等都将被架空。

二、算法透明与算法保密的价值辩证

鉴于大数据金融智能投顾的负面效应,受算法影响的投资者权益得到了法律的紧密关切。例如,欧盟《通用数据保护条例》(General Data Protection Regulation,GDPR)第 14 条和第 22 条规定,如果个人数据被用于用户画像、自动化决策等场景,数据控制者须向数据主体提供算法逻辑以及此类处理对数据主体的预期后果,以确保涉及数据主体的决策合理、透明。尽管从消除智能投顾负面效应的角度出发,受算法影响的投资者有理由要求算法透明以保障其合法权益,但在利益冲突的另一端,算法多以商业秘密、专利加以保护,算法保密还是算法控制者在激烈的市场竞争中保持商业优势的重要保障,因此维护算法的保密性亦有其正当性。

(一)受算法影响投资者的利益诉求

在智能投顾的服务过程中,投资者只能看见数据输入和方案输出的客观事实和决策结果,对算法系统的设计原理、计算机制、程序运行和决策依据

等一无所知。存在于数据输入和结果输出之间的决策层深深地隐藏于算法黑箱之中。对此,人们直觉上的第一反应便是“开箱”。通过打开算法黑箱,将“阳光”洒落在智能投顾自动化决策的全过程,藉此识别智能投顾黑箱中的算法歧视和算法霸权,查明算法决策与投资者损害结果之间的因果关系。

当前,人们对算法透明原则的认知可谓言人人殊,其中最具代表性的是美国学者弗兰克·帕斯奎尔(Frank Pasquale)的论述。帕斯奎尔认为算法透明包括代码公开、算法分析和算法审计等,只有综合以上手段才能合理促成算法透明[12]。有国内学者批评帕斯奎尔的理论模糊了算法透明和其他规制手段的界限,会给理论研究和实务应用带来很大麻烦[13],但不得不承认,源代码披露、输入数据和输出结果公开等手段仍然是当前学界普遍认同的算法透明原则的要素集合。

算法透明显然符合受算法影响之投资者的利益诉求,因为算法透明有助于打破智能投顾的“算法暴政”。如果智能投顾的自动化决策算法不公开、不提供算法解释、不接受投资者质询、不进行权利救济,投资者便无从知晓算法中是否嵌入了操控性程序,无法确信系统生成的投资组合方案是否最合适,其权利救济更是无从谈起,只能接受算法的宰制和“暴政”[14]。只有通过算法透明,投资者才可能获知自身的财务信息和风险偏好如何与资产配置方案发生关联,才可能进行权利救济。

此外,算法透明也有助于消除智能投顾的“算法歧视”。在人们的普遍认知中,歧视、偏见专属于人类这种拥有主观立场的动物。以算法驱动的人工智能代表着冷静、理性和客观,算法决策似乎为实现人们苦苦追求的绝对公正带来了福音。而事实并非如此,人工智能的“成长”离不开数据的“投喂”,这种“投喂”如同人生阅历形塑人类主观立场的过程。也就是说,算法也是拥有主观立场的。不同的主观立场深刻影响着算法的决策结果,隐藏其中的偏见、歧视更难以察觉。因此,消除“算法歧视”也是算法透明的重要理据。

最后,算法透明还是对算法操控者问责的基本前提。如前所述,智能投顾的“人机混合”使得决策失误导致投资者损失的因果关系难以确定,投资者损失很容易被归咎于技术风险。对此,算法透明让投资者能够一窥算法的“大脑”,探知资产配置组合决策生成的全过程,算法运行过程中出现的精确性偏差和算法偏见都可能被发现和察觉。即便投资失败的损失已经“覆水难收”,投资者也可以依据算法透明所披露出来的因果关系向算法设计者或者算法运行者主张赔偿责任。

(二)算法控制者保持市场竞争力的现实需要

与欧盟治理算法强调私人权利保护、公平竞争秩序以及社会公共利益不同,美国采取了更加自由与开放的治理策略。搜索王诉谷歌案(Search King, Inc v. Google Tech, Inc)⑥奠定了美国司法部门对算法应用中利益冲突的基本态度。在该案中,法院认为每一种搜索引擎确定搜索结果的算法都不同,这意味着每一种搜索引擎都在表达各自不同的意见,而搜索引擎有权依据自身独特的算法“说”出自己的意见。根据算法呈现的搜索结果代表了谷歌公司的言论,而言论自由属于美国宪法上具有不可剥夺性的绝对权利,无论谷歌公司如何通过算法调整、呈现搜索结果,都等同于谷歌在发表自己的意见,任何人都不得干涉谷歌公司的“算法(言论)自由”。

而在威斯康星州诉卢米斯案(State v. Loomis)⑦中,法院则将算法认定为商业秘密。法院使用基于算法的 COMPAS 风险评估工具,根据卢米斯的访谈和犯罪史对其进行累犯风险评估,在此基础上对卢米斯判处刑罚。卢米斯认为法院仅根据评估方法并不公开的 COMPAS 系统就对其作出累犯风险评估,侵犯了其基于精确信息被量刑的权利和获得个性化判决的权利,因此要求公开 COMPAS 系统的算法。法院认为,卢米斯可以对 COMPAS 系统生成的报告进行否认或解释,从而验证法院判决依据的准确性,但不能要求公开 COMPAS 系统算法,因为 COMPAS 系统背后的算法属于商业秘密。法院最终驳回了卢米斯对量刑程序的异议。

被看作商业秘密的人工智能算法在美国法中受到严格保护,是因为作为知识产权保护客体之一的商业秘密一旦遭到侵犯,势必会阻碍科技进步,进而侵害人民的福祉。而无论被视为“言论”还是“商业秘密”,算法在美国都受到了法律的严格保护,

法官不会要求算法设计者或者运营者以公众能够理解的方式说明算法的工作原理,更不会要求其公开算法代码,算法黑箱被视为人工智能时代的正常现象,是人们享受算法带来的便利的同时必须承受的合理代价[15]。事实上,美国对算法的负面效应采取“绥靖政策”在某种程度上来说也是为了维护其自身在人工智能领域的国际领先地位,便于本国科技巨头在全球拓展业务[16]。

此外,在数字经济时代,算法也是科技公司的核心竞争力。将智能投顾背后的算法视作商业秘密加以保护同样符合金融机构的利益,是金融机构保持市场竞争力的现实所需。申言之,算法是设计者为解决特定问题设计的有限且明确的操作步骤,表现为源代码和目标代码,是源代码向目标代码转化的技术措施。源代码的秘密性是算法控制者拥有并保持独特市场竞争力的核心因素。源代码一旦失去秘密性,算法控制者的市场竞争力便不复存在。因此,通过法律保护算法、维持算法黑箱在一定程度上既是国家维持科技竞争力的需要,也是科技企业保持市场竞争力的需要。

三、算法规制目的导向:有层次的算法透明

出于维护算法控制者市场竞争力、激励技术创新的现实需要,将算法认定为言论或者商业秘密都将为算法提供强大的法律保护。而为了避免算法暴政对投资者权利的侵害,可能目前唯一的方法只能是“攻破”算法黑箱,尽量使算法透明化。在此之间的矛盾可以“转译”为算法保密与算法透明之间的价值冲突。由于算法保密和算法透明都具有一定的正当性,因此解决问题的重点便来到如何实现算法保密与算法透明的利益平衡。

(一)智能投顾算法透明层次性的概念阐释

作为算法规制理论的一项重要原则,算法透明要求算法开发者或者运营者披露包括算法源代码、输入数据和输出结果在内的一切算法要素。这种彻底的算法透明夸大了其对投资者权利保护的效用,忽视了金融机构保护商业秘密的现实需要。事实上,在算法广泛应用的数字经济时代,彻底的算法透明既不可行,也无必要。算法处理数据的能力远远超过了人类的运算能力,随着机器自主学习、深度学习的能力不断迭代更新,彻底的算法透明越来越难以实现。

在技术层面,不断增强的机器自主学习能力给算法透明带来了巨大的技术障碍,因为此刻透明的算法在通过自主学习实现更新迭代后便不再透明。在内容层面,自动化决策算法存在很高的专业壁垒,普通人无法理解复杂的数学模型,代码和技术的透明对其没有任何意义[17]。在智能投顾领域,让投资者理解杠杆比率、基本面分析和技术面分析等金融术语尚且有一定难度,倘若告知其智能投顾是通过随机森林、决策树模型、逻辑回归模型等算法进行金融分析的,那么这种算法透明对投资者而言也毫无用处[18]。

鉴于此,本文倡导一种有层次的算法透明,即根据规制算法的不同目的,以不同方式对不同主体实现不同程度、不同层次的算法透明。申言之,对于金融监管部门来说,捕获金融机构开展智能投顾业务潜在的算法风险,对风险进行事前评估并采取适当的防范措施是其核心职责所在。为了履行这一职责,金融监管部门需要对智能投顾的算法参数、代码结构和内置程序等内容进行审查,引导金融机构的算法开发、系统运营符合相关技术标准与行业标准。因而对金融监管部门有效的是“以模型为中心”(model-centered)的算法透明,这个层次的算法透明重在获取与算法模型密切相关的参数设置、特征权重、逻辑架构等宏观、全局性技术信息。

而对于投资者来说,作为智能投顾算法决策结果的直接利害关系人,其有权知晓算法的决策逻辑和运行机制。尤其是当自动决策系统生成的投资组合方案与投资者本人的资产配置诉求存在偏差时,投资者有权要求金融机构对自动决策系统算法的模型架构和数据分析进行解释,以便在投资者画像失真、风险错配或者权重失衡的情况下进行方案纠偏和权利救济。那么,对投资者有效的是“以用户为中心”(subject-centered)的算法透明,这个层次的算法透明重在告知投资者算法依据哪些个人信息进行决策,各项个人信息在算法自动化决策中的

权重,以及算法运行的逻辑等[19]。对于“以用户为中心”的算法透明来说,算法的源代码和数学模型并不重要,投资者更关心其个人信息如何参与算法的决策,以及算法按照怎样的逻辑进行决策。

(二)智能投顾算法透明层次性的理论依归

从工具价值来说,算法透明原则的提出在于对抗算法黑箱滋生的利益掠夺、算法歧视和算法暴政,实现对算法的规训。面对受算法影响投资者权利保护与算法控制者保持市场竞争力之间的冲突与张力,有层次的算法透明提供了富有弹性、容纳不同利益诉求的冲突协调方案。在智能投顾算法的开发和运营过程中,有层次的算法透明能够实现无利益冲突则各行其是、有利益诉求则“掀开最小缝隙”的特异性透明规范效果。算法透明的层次性立基于受算法影响投资者与算法控制者之间利益衡量、冲突化解的现实需要,是“目的—手段”理性的具体呈现,其理论依归可以从比例原则的功能定位和适用方式两个维度获得解释力。

诞生于德国公法中的比例原则以其旺盛的生命力实现了从国家、区域到全球的影响,以及从公法向私法和其他部门法的渗透,甚至在某种程度上嬗变为一般法理念和法律分析方法[20]。在功能定位上,原初的比例原则关注“公共利益与私人权益”之间的利益衡量。然而,由于利益是人们在一定社会情境下基于某种生存体验而形成的诉求[21],利益衡量在人们的社会生活中普遍存在,比例原则的功能定位完成了从防御性权利保障到一般化利益衡量的演变。就智能投顾算法开发、应用中的利益衡量来说,对算法的法律规制需要强调干预的适度性,根据不同主体不同的利益诉求采取不同方式、不同程度的算法透明,便是比例原则利益衡量功能的具体体现。

在适用形态上,比例原则作为法律原则的规范性命题伴随司法适用的推进实现了从规范立场向超越规范立场的转型,并且开始作为方法论成为“目的—手段”分析范式的基准[22]。具体来说,比例原则通过衡量手段和目的之间的比例关系来判断手段的正当性与合理性。如果欲达成某种目的所采取的手段适当、必要且达成目的所实现的利益大于该手段所侵害的利益,那么这一组“目的—手段”便是合比例的,该手段可以采取;反之,没有通过适当性、必要性、均衡性审查的手段由于不合比例则不可采取。在智能投顾受算法影响投资者与算法控制者的利益冲突中,为了达成保护投资者权利的目的,需要实现算法透明。但是,必须对算法透明作出必要限制,以符合比例原则所要求的实现手段的适当性、必要性和均衡性,这便是算法透明层次性背后的理论逻辑。

实行有层次的算法透明是当前调适算法保密与算法透明利益冲突的最优解,是比例原则在智能投顾算法治理中的具体应用。为了保护金融机构的商业秘密等私权利,激励金融机构创新、优化算法,在增强自身市场竞争力的同时增进投资者福祉,算法不公开是法律介入智能投顾算法的基本原则。在这个基础上,为了消除算法黑箱可能带来的种种负面效应,只有“掀开最小缝隙”实现算法对金融监管者、投资者及其他利益关切者最小程度、最特异性的“透明”,才能达成算法保密与算法透明两者间的利益平衡。

四、智能投顾算法透明的实现路径

在大数据时代,算法对于人们的生活越来越具有构成性。算法早已普遍而深刻地影响着人们的生活,决定着人们可选择的空间以及可能获得的结果。彻底杜绝算法黑箱意味着对数字化生活方式的远离,意味着对算法所蕴含便利和好处的远离。但是,对算法黑箱一定程度的容忍也绝不意味着人们应该束手就擒、任其宰割。由于算法透明具有层次性,我们需要在智能投顾算法设计、开发、应用的不同阶段,出于维护金融系统的安全稳定、保障投资者的知情权利、便利受算法影响投资者的权利救济等不同目的,对智能投顾算法实现不同层次的算法透明。

(一)合规审查+算法备案:防控系统性金融风险的事前透明

防范局部金融风险的产生及其向整个金融系

统的蔓延是金融监管部门的重要职责,对智能投顾中金融风险的防控也不例外。为此,金融监管部门需要详细了解智能投顾算法的核心思想、模型架构和服务模式,掌握智能投顾算法设计的底层逻辑,审查投资者个性化资产配置方案生成过程中算法在各个环节的功能和影响。为此,智能投顾系统的开发机构需要适度披露算法的模型,将智能投顾背后的金融数据、算法参数和代码结构等提供给金融监管部门,实现以模型为中心的算法透明。值得深究的问题是,应该利用怎样的制度安排去实现算法模型透明?当前,学界已提出算法开源、算法备案以及算法合规审查等多种算法透明制度,那么应当如何配置、组合这些强度不一的方案呢?

尽管算法开源有利于更广泛的社会主体通过检验代码捕获算法内含的缺陷及其潜在金融风险,但是算法源代码毕竟涉及开发者和运营者的商业秘密,法律不宜强制其开源[23]。在针对金融监管部门算法透明的实现方式上,可以区分成熟算法模型和新型算法模型,分别以合规审查和算法备案予以实现。具言之,对于那些经过长期研发和验证、在智能投顾中取得良好效果并得到广泛应用的成熟算法模型,由于大量实践应用能够证明其具有较高的准确性和可靠性,对这类算法模型应当以合规审查的方式进行监管,即审查该算法模型是否符合法律的强制性规范、金融业的行业标准以及相关技术标准。而对于那些为克服既有算法模型局限性而引入新思想、新技术、新数学概念的创新、改进型算法模型,由于相关的法规建设和标准制定滞后于技术创新,为了激励算法开发者积极解决复杂问题、提高算法效率,拓展新思路、开发新工具,金融监管部门应侧重于算法开发者的权利保护,以算法备案的方式掌握算法模型并持续监控可能潜藏其中的风险因素即可。需要特别指出的是,为了保护智能投顾系统开发机构的知识产权和运营机构的商业利益,需要对金融监管部门及其工作人员课以严格的保密义务。

从实践的情况来看,大数据金融算法备案包括算法逻辑本身的备案和金融从业人员的备案(注册)。例如,欧洲证券市场监管局要求使用算法进行高频交易的投资机构每年都向其报备算法交易策略、交易参数设定、风险控制模块以及系统测试结果等内容,藉此避免缺陷算法造成金融市场动荡[24]。由于智能投顾实际上模拟了金融从业人员根据先验知识、逻辑推理和专业判断提供投资决策方案的过程,因而参与算法开发、设计的金融从业人员也需进行备案。例如,美国 2016 年修订的《纳斯达规则》(NASD Rules)要求负责开发、设计或者重大修改金融产品算法交易策略的人员和日常监管、指导上述活动的人员都必须注册为证券交易者[25]。在我国,由中国人民银行、银保监会、证监会等金融监管部门在 2018 年出台的《关于规范金融机构资产管理业务的指导意见》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”。当然,该指导意见对金融科技算法备案的规定还十分模糊,算法备案的主管部门、备案内容、备案程序以及未履行备案义务的法律后果等事项还有待将来细化。

(二)算法告知+算法解释:保障投资者知情权利的事中透明

以算法备案破解算法黑箱的负面效应虽然必要但并不充分,因为算法备案并不意味着算法公开和算法可理解。姑且不谈算法备案对监管部门及其工作人员课加的保密义务,即便智能投顾的算法模型、主要参数、运行规则甚至源代码对投资者完全公开,金融企业的技术与行业的专业性仍然是投资者难以逾越的高墙。对于投资者来说,以自然语言对算法决策所依赖的逻辑和自动生成的资产配置方案进行解释才是有意义的[26]。因此,除了在算法设计、运行之初通过合规审查与算法备案对智能投顾算法进行监管之外,还需在算法的使用过程中增强算法自动化决策的可理解性,这就涉及规制算法的一组重要权利义务——算法告知义务和算法解释请求权。

受算法影响投资者的知情权和决定权是保护其个人信息权益免受侵害的基本权利,其中又以知情权更为重要,因为知情是允许、限制抑或拒绝他人进行信息处理的前提[27]。个人信息处理者的算法告知义务和信息权利人的算法解释请求权共同组成受算法影响投资者的知情权体系,二者在智能投

顾的金融服务中表现为解释算法运行逻辑的动态互动过程。由于算法控制者与投资者普遍的信息不对称,法律的规制方向通常倾向于先让算法控制者履行告知义务(投资者知情权的初次实现),在此基础上再以投资者的算法解释请求权作为补充(投资者知情权的补充实现)。就告知(解释)的内容而言,欧盟 GDPR 采取的是“基于逻辑的解释”⑧,即受算法影响者有权获得处理者对自动决策算法逻辑的解释,而美国采取的是“基于事实的解释”⑨,即金融消费者仅有权获得算法对其作出不利决策所依据事实的解释,而非算法逻辑本身。

我国对算法解释的规范经历了从信息处理公开透明的被动权利向算法决策解释说明的主动权利“进化”的过程。在信息处理公开透明方面,根据我国法律的相关规定⑩,网络运营商等实体在处理个人信息时必须在信息主体知情、自愿、同意的前提下遵循公开原则,明示其处理的目的、方式和范围;在算法决策解释说明方面,《个人信息保护法》第 24 条和第 48 条明确规定,通过自动化决策算法作出对个人权益有重大影响的决定,个人有权要求个人信息处理者对信息的处理规则进行解释与说明。可见,算法告知和算法解释在我国已经成为法定权利与义务。

算法告知和算法解释入法只是实现智能投顾算法事中透明的起点。算法告知义务和算法解释请求权还存在权利内容、解释程度、解释时间、解释方法等一系列问题有待厘清和细化[28]。但就制度效能而言,即便对算法解释的实际效果持怀疑态度的学者也认为不可能彻底抛弃算法解释权[29]。算法解释在保障投资者知情权的同时,作为外部监管的补充,还能够强化投资者与金融机构的信赖关系,共同限制金融机构的算法权力,一定程度弥合投资者与金融机构之间的专业鸿沟,实现投资者与金融监管部门对算法的合作治理[30]。

(三)评估算法的参与程度:维护投资者信赖利益的事后透明

在事前、事中的算法透明手段之外,事后的算法评估也是提升透明度的重要手段。由于人工智能算法时时刻刻都处于更新迭代之中,事前备案后实现相对透明的算法在运行过程中经过深度学习和自我纠偏,往往变得不复透明。有学者便指出,算法备案无法准确捕捉算法运行的内在机理,将逐渐异变为浮于形式的合规管理[31]。事实上,在“掀开最小缝隙”思想的指导下,智能投顾算法的事前备案旨在实现金融风险的可防、可控,而非算法的彻底透明。随着深度神经网络和动态算法技术的发展,即便事前的合规审查和算法备案转变为算法模型对金融监管部门的完全透明,算法投入运行后的透明度仍然会变得“忽明忽暗”。对此,算法评估无疑是改善投资者弱势地位、构筑社会公众算法信任的重要一环。

根据欧盟 GDPR 第 35 条的规定,如果使用人工智能等新技术处理数据将会提升侵犯自然人权利与自由的风险,除了考虑该处理的性质、范围、领域和目的之外,数据处理者应当在数据处理之前评估拟使用的处理程序对个人数据保护造成的影响。加拿大《自动化决策指令》(Directive on Automated Decision-Making)第 6 条规定,任何自动化决策系统投入运行之前都须发布算法影响评估的最终结果。美国《算法问责法案》(Algorithmic Accountability Act)第 3 部分规定,自动化决策系统在投入运营之前应当接受公正性、透明度和可问责性等方面的评估。该法案通过前两年内投入运营的算法系统也应纳入评估的范围。我国《个人信息保护法》第 55 条同样规定,利用个人信息进行自动化决策的,个人信息处理者应当事前进行个人信息保护影响评估。可见,自动化决策可能对个人权利和自由造成的影响已为美国、加拿大、欧盟和我国所重视,算法评估作为一项正式的法律制度在以上国家和地区都已经入法。需要特别指出的是,上述国家或地区对算法评估的规定都是出于保护自动化决策中的个人信息和数据权利,因而算法评估的启动节点都规定在算法程序投入运营之前。

然而,对于实现智能投顾算法的持续透明来说,评估算法对投资者信赖利益的影响更为迫切。在投资顾问活动中,法律要求金融机构严格遵守投资者适当性规则,如果金融机构违反投资者适当性规则给投资者造成损失,将部分或者全部赔偿投资者的损失。法律之所以如此严苛,是因为金融机构与

投资者之间存在弥深的专业鸿沟和严重的信息不对称。法律打破“买者自负”原则以投资者适当性规则对投资者进行“家长式保护”就是为了维护投资者对金融机构的信赖利益[32]。智能投顾中算法的介入显然会干扰投资者对金融机构的信赖利益,要了解这种“干扰”是否以及在多大程度上影响投资者的信赖利益,需要对不同的智能投顾算法系统进行个别化评估。事实上,从当前的立法情况来看,欧盟和加拿大法律规定的算法评估都具有周期性和动态性[33],美国的算法评估甚至有一定的溯及力。因此,我国不宜拘泥于事前、静态、封闭的合规型算法评估,应根据算法在智能投顾服务中的参与程度增加事后评估,以便于在投资损失发生后进行投资者、金融机构和技术风险的责任划分。

五、结　语

在大数据和算法的驱动下,各种算法模型被开发出来与具体的金融场景深度融合,不断提升金融产品质量、优化金融投资策略,实现金融服务的智能化、动态化和前瞻化。在此背景下诞生的智能投顾不仅能够精准分析投资者的资产配置需求,还极大提升了投资顾问业务的服务质量与服务体验。然而,算法在为金融赋能的同时,也加深了金融业的信息不对称,带来了更高的风险系数。对于传统金融监管来说,算法可能成为规避监管的工具;对于投资者来说,其与金融机构之间的信赖利益被算法的自动化决策切断,在投资决策与主体问责之间形成巨大的不确定性。

算法保持一定程度的秘密性既符合国家战略,也是算法开发者和运营者建立、维持竞争优势的核心所在。面对算法保密与算法透明之间的价值冲突,只有秉持“掀开最小缝隙”原则,在不同阶段、基于不同目的对不同主体实行不同方式、不同程度的透明,才能在算法福利与算法宰制之间取得平衡。就大数据金融中的智能投顾而言,为了防控系统性金融风险,需通过“合规审查+算法备案”向监管主体实现算法模型透明;为保障投资者的知情权,需以金融机构的算法告知义务为基础辅之以投资者的算法解释请求权实现算法逻辑透明;为缓解算法决策与主体问责之间的紧张关系,需通过算法评估判断算法对投资者信赖利益的影响程度,进而明确主体责任的划分。

注:

① KYC 义务要求金融机构建立适当的薪酬制度,防止业务员推介金融商品时单方面追求销售业绩引致的道德风险,同时落实金融消费者适当性原则;KYP 原则要求金融机构加强对创新金融产品的风险控制和监督,同时强化金融商品推介过程中的信息披露义务和风险提示义务,充分保障金融消费者的知情权。参见刘博涵.我国台湾地区金融消费者保护:历程、经验与省思[J].社会科学论坛,2021,(1):133–146.

② 关于以算法透明规制算法的讨论,参见 Frank Pasquale. The Black Box Society:The Secret Algorithms That Control Money and Information[M].Cambridge:Harvard University Press,2015;汪庆华.人工智能的法律规制路径:一个框架性讨论[J].现代法学,2019,(2):54–63; 张凌寒.算法权力的兴起、异化及法律规制[J].法商研究,2019,(4):63–75.

③ 人工智能算法不公开、不透明,在学界一般被称为“算法黑箱”。然而,“黑箱”与“黑幕”“暗箱操作”等语词直接关联,带有明显的贬义感情色彩。相较而言,在指代算法不公开、不透明时,“算法保密”更为中性,更有利于问题的理性讨论。鉴于此,本文不对“算法黑箱”“算法保密”作严格区分,行文中根据需要,二者将会交替出现。

④ 贪心算法的基本思路是:(1)建立数学模型来描述待解决的问题;(2)把求解的问题分成若干子问题;(3)对每一子问题进行求解,得到子问题的局部最优解;(4)把子问题的局部最优解合并成原来问题的一个解。参见尚荣华.智能算法导论[M].北京:清华大学出版社,2021:148–152.

⑤ 美国 OpenAI 公司于 2022 年 11 月推出的生成式预训练转换器(ChatGPT)一经问世就受到人们的广泛关注,该系统基于“自回归语言模型”(Auto–Regressive Language Model)、“人类反馈强化学习”(Reinforcement Learning from Human Feedback)技术架构与驯化模型,具备自然语言生成与错误更正能力。目前,ChatGPT 不仅可以同人类进行流畅的场景交流,还可替代人类从事文本撰写、方案制定、翻译等诸多事务。这预示着此前 300 多个人工智能科学家预测的至少 45 年以后才会到来的强人工智能时代正在加速到来。

⑥ 搜索王是一家将广告链接嵌入谷歌搜索结果网页的公

司。显然,在谷歌的搜索结果中排名越靠前,获得的点击量和流量就越高。搜索王的商业模式被谷歌公司发现,谷歌公司通过改变算法降低了搜索王相关网页的排名,甚至完全不予展示搜索王嵌入广告的网页。搜索王公司诉称,谷歌公司得知搜索王的公司业务高度依赖谷歌公司的搜索结果排名后有意降低相关网页的排名甚至移除相关网页,给公司带来了"无可估量的损失"。谷歌公司则辩称,谷歌没有义务将搜索王嵌入广告的网页排在其想要的位置,也有权不予展示搜索王嵌入了广告的网页。参见 Search King, Inc. v. Google Tech, Inc., No.02-1457,2003 WL 21464568(W. D. Okla. May 27, 2003).

⑦ 参见 State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

⑧《通用数据保护条例》序言第 71 条规定,数据主体有权就基于自动化处理而对其产生法律效力或类似重大影响的决定获得解释,并就该决定提出质疑。

⑨ 美国在《平等信贷机会法案》(ECOA)和《公平信用报告法案》(FCRA)两部法律中以"不利行动告知条款"保障金融消费者在算法作出拒绝提供信贷、拒绝提供保险服务等不利行动时获得解释的权利。

⑩ 参见《民法典》第 1035 条、《网络安全法》第 41 条、《个人信息保护法》第 7 条、第 14 条和第 17 条。

参考文献:

[1] 姜海燕,吴长凤.智能投顾的发展现状及监管建议[J].证券市场导报,2016,(12):4-10.

[2] (美)卢克·多梅尔.算法时代:新经济的引擎[M].胡小锐,钟毅译.北京:中信出版集团,2016.123.

[3] 金东寒.秩序的重构——人工智能与人类社会[M].上海:上海大学出版社,2017.72.

[4] 杨东,武雨佳.智能投顾中投资者适当性制度研究[J].国家检察官学院学报,2019,(2):3-18.

[5] Theodor D. Sterling. Guidelines for Humanizing Computerized Information Systems:A Report from Stanley House [J].Communications of the ACM,1974,(11):609-613.

[6] 廖岷.金融科技发展的国际经验和中国政策取向[M].北京:中国金融出版社,2017.88.

[7] 汪世虎,马瑞乾.金融数据安全背景下的智能投顾算法黑箱监管体系构建[J].社会科学辑刊,2022,(2):86-95.

[8] 刘博涵.信用图像与金融刑法重塑[J].西部法学评论,2020,(6):25-37.

[9] 何颖,阮少凯.论金融产品销售商的投资者适当性义务[J].财经法学,2021,(1):134-145.

[10] 甘培忠,周淳.证券投资顾问受信义务研究[J].法律适用,2012,(10):33-39.

[11] 高丝敏.智能投资顾问模式中的主体识别和义务设定[J].法学研究,2018,(5):40-57.

[12] Frank Pasquale.The Black Box Society:The Secret Algorithms That Control Money and Information[M].Cambridge:Harvard University Press,2015.8-11.

[13] 沈伟伟.算法透明原则的迷思——算法规制理论的批判[J].环球法律评论,2019,(6):20-39.

[14] 张凌寒.商业自动化决策的算法解释权研究[J].法律科学(西北政法大学学报),2018,(3):65-74.

[15] 陈景辉.算法的法律性质:言论、商业秘密还是正当程序?[J].比较法研究,2020,(2):120-132.

[16] 毕文轩.生成式人工智能的风险规制困境及其化解:以 ChatGPT 的规制为视角[J].比较法研究,2023,(3):155-172.

[17] 刘琳.算法解释权与商业秘密保护的冲突化解[J].行政法学研究,2023,(2):168-176.

[18] Davide Castelvecchi.Can We Open the Black Box of AI?[J].Nature,2016,(10):20-23.

[19] 安晋城.算法透明层次论[J].法学研究,2023,(2):52-66.

[20] 冷传莉.比例原则私法化的体系定位与调整对象[J].比较法研究,2023,(4):99-116.

[21] 卢永欣.利益和原则:马克思恩格斯思想的重要线索[J].贵州大学学报(社会科学版),2021,(3):10-15.

[22] 蒋红珍.比例原则适用的范式转型[J].中国社会科学,2021,(4):106-127.

[23] 许可.驯服算法:算法治理的历史展开与当代体系[J].华东政法大学学报,2022,(1):99-113.

[24] 金小野.规范高频交易是国际证券业监管焦点[N].法制日报,2013-11-12(010).

[25] 徐凤.人工智能算法黑箱的法律规制——以智能投顾为例展开[J].东方法学,2019,(6):78-86.

[26] 解正山.算法决策规制——以算法"解释权"为中心[J].现代法学,2020,(1):179-193.

[27] 万方.算法告知义务在知情权体系中的适用[J].政法论坛,2021,(6):84-95.

[28] 丁晓东.基于信任的自动化决策:算法解释权的原理反思与制度重构[J].中国法学,2022,(1):99-118.

[29] 辛巧巧.算法解释权质疑[J].求是学刊,2021,(3):100-109.

[30] Margot E. Kaminski.Binary Governance:Lessons from the GDPR's Approach to Algorithmic Accountability[J]. Southern California Law Review,2019,(6):1529.

[31] Ari Ezra Waldman.Power,Process,and Automated Decision-Making[J].Fordham Law Review,2019,(2):613-632.

[32] 胡改蓉,钱程.投资者适当性管理制度中金融机构的

免责机制——以“信赖利益”判断为核心[J].证券市场导报,2021,(10):61-70.
[33] 张恩典.算法影响评估制度的反思与建构[J].电子政务,2021,(11):57-68.

【责任编辑:周濛】

Value Dialectics, Regulatory Orientation and Implementation Path of Algorithm Transparency in Robo-advisory

LIU Bo-han
(Law School, Guizhou University, Guiyang, Guizhou, 550025)

Abstract: Robo-advisory using AI technology to analyze big data can better implement the principle of investor suitability and generate more suitable investment allocation schemes for investors, but they also conceal negative effects such as financial institutions' profit-grabbing, risk mismatch of investment solutions, and hollowing out of the legal obligations of investment advisors. Algorithm transparency helps to break the tyranny of algorithms, eliminate algorithm discrimination, and lay the foundation for algorithm accountability, while algorithm confidentiality is necessary to maintain national scientific and technological strength, and to maintain the competitiveness of scientific and technological enterprises in the market. In the face of the value conflict between algorithm transparency and algorithm confidentiality, complete algorithm transparency and absolute algorithm confidentiality are both biased, and different degrees and levels of algorithm transparency should be realized for different subjects in different ways according to the different purposes of regulating algorithms. The hierarchy of algorithmic transparency is a specific application of the principle of proportionality, the function of measuring interests, and the proportional analysis paradigm of "ends-means" in the governance of robo-advisory algorithms. In the realization path of algorithmic transparency of robo-advisory, "compliance review + algorithmic filing" is used to realize algorithmic model transparency in order to prevent and control systemic financial risks, "algorithmic informing + algorithmic interpretation" is used to realize algorithmic logic transparency in order to protect the investors right to know. Afterward, it evaluates the degree of influence of algorithms on investors' trust interests, and clarifies the subjectivity, causality and relevance of algorithmic decision-making, so as to determine and allocate algorithmic responsibilities.

Key words: robo-advisory; algorithm transparency; algorithm confidentiality; value dialectics; regulatory orientation; implementation path