理论前沿

算法信任的法律性质

黄伟文

内容提要:算法信任法律性质的讨论,需在法律维度、社会关系维度和概念维度所编织的论证网络中展开,并回答在法律意义上,信任算法是指信任什么、法律保障算法信任是在保障什么这两个核心问题。据此,作为法律的调整对象,算法信任既不是作为主观心理状态的信任感,也不是算法本身或算法治理制度的可信性;相反,它是一种特殊的委托关系,即在算法提供者与算法用户的交互行为中产生的回应性规范关系。在规范构造上,算法信任包含算法能力、算法意图和算法认知三个构成性要素。法律保障算法信任,就是要保障这三个要素的有效实现;算法治理制度的法律建构,也应据此而展开。

关键词:算法信任　回应性规范关系　构成性要素　算法治理

黄伟文,广东财经大学法学院副教授。

鉴于现代社会对智能算法的日益依赖与对算法的信任并不同步,〔1〕关于算法信任的法律规制,业已引起学界的高度关注。〔2〕但是,算法信任的法律性质为何,学界却缺乏清晰界定与共识,未能为相关法律制度的建构与反思提供充分理据,故有必要申论之。

算法信任法律性质的讨论,意在回答当我们在法律的意义上说算法信任时,实际上是在信任什么;或者当我们说法律应为算法信任提供保障时,实际上保障的是什么。对此,学界主要观点有三:一是主张算法信任即对算法的信任感;二是把算法信任理解为算法可信;三是认为算法信任就是对算法治理制度的信任。相应地,法律为算法信任提供保障,就分别是保障对算法的信任感、保障算法的可信性和保障算法的有效治理。本文将在前两个部分讨论这三种观点,并指出它们不符合法律调整与性质界定的基本原理。在第三

〔1〕 See Mariarosaria Taddeo & Luciano Floridi, The Case for E-Trust, 13 *Ethics and Information Technology* 1, 1-3 (2011).

〔2〕 较有代表性的中文文献主要参见丁晓东:《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》,《中国法学》2022 年第 1 期,第 99-118 页;袁康:《可信算法的法律规制》,《东方法学》2021 年第 3 期,第 5-20 页;张欣:《从算法危机到算法信任:算法治理的多元方案和本土化路径》,《华东政法大学学报》2019 年第 6 期,第 17-30 页。

部分,本文将为算法信任就其法律性质而言是一种回应性规范关系进行分析,并阐明其规范构造与具体内涵。第四部分就法律制度对此应作何回应,做一初步展望。最后是一个简单的结语。

一 作为法律调整对象的算法信任

(一)算法信任与法律

算法信任可为多学科的研究对象,对之既可作描述性研究,如社会学或心理学视之为一种社会的或心理的客观现象,[3]也可作规范性研究,如伦理学关于合理信任条件的讨论。[4] 本文是在法律视域下,讨论算法信任的性质。在法律的视野中,算法信任是一种社会关系,而算法信任的性质则指涉算法信任的概念或其构成性要素。因此,算法信任的法律性质便包含了法律的、社会关系的与概念的三个维度。本文的讨论将围绕这三个维度展开。

(二)算法信任与信任感

有一种观点认为,算法信任即算法信任感,就是指相信算法的一种情感或心理状态。因此,法律规制或调整算法信任,就是要促进人们对算法的信任感。[5] 这种理解,是将心理学意义上的信任概念,直接添加算法的前缀,移植至法律领域。这种做法忽视了法律调整的特殊性质。

信任与信任感之间的关联是显而易见的。当我们信任他人时,也就是说我们对他人有信任感,如果缺乏信任感,我们便无法信任他人。但是,在现实生活中,两者并非总是一致的。由此可见,信任与信任感并非同一概念,信任感是一种主观状态和心理事实,而信任则包含了两个方面:信任他人,且认为他人是值得信任的,后者强调值得信任的客观属性,具有明显的规范意涵。

法律对算法信任的有效规制,无疑有利于促进人们对算法的信任感。但是,法律是通过调整主体外在的交互行为从而调整社会关系的,单纯的内心行为和个人行为不是法律的调整对象。[6] 信任感作为一种主观的和单方面的心理状态,并非法律的调整对象。换言之,算法信任要成为法律适格的规制或调整对象,就不能与信任感等而视之。

法律是行为规范,法律规制算法信任,是指为了维护算法信任,主体应如何行为。法律规制算法信任,主要体现在算法应用场合。在此场合中,存在双方主体,即算法的提供者,例如算法运营商、算法平台等,以及算法的使用者,即算法用户。两者形成一种委托关系:算法用户委托算法提供者提供算法服务,以实现其特定目标;算法提供者则

〔3〕 参见刘特、郑跃平、杨学敏:《公共部门中的算法透明能否影响公众信任?——基于一项调查实验》,《电子政务》2024年第8期,第13-28页。

〔4〕 See Carolyn McLeod, Our Attitude towards the Motivation of Those We Trust, 38 *Southern Journal of Philosophy* 465, 465-479 (2000).

〔5〕 参见袁康:《可信算法的法律规制》,《东方法学》2021年第3期,第6页。

〔6〕 参见张文显主编:《法理学》(第五版),高等教育出版社2018年版,第72、79-80页。

根据算法用户的委托,提供相应的算法服务,以实现该目标。算法信任就是指这种基于信任的算法服务委托关系。在法律调整的视角下,法律对算法信任的规制,就是要确定双方主体应如何行为,以保证该关系的合理实现。显然,算法信任在此并非描述性概念,而是规范性概念;算法信任也不是单方面的个人行为或心理活动,而是关涉他人的交互性社会行为。

从法律调整的目的来看,算法信任作为法律的规范性目标,是受法律保护的法益。因此,如果算法信任就是算法信任感,那么,只要有助于产生、维持和促进算法信任感,便应获得法律的正面评价。但是,通过如欺骗、隐瞒、篡改与删除负面信息等不正当算法操纵手段,或者利用煽动与诱导等传销话术,也可能达到这种效果,而法律对此却应予以负面评价。由此亦可见,算法信任与算法信任感并非一回事。

二 算法可信、算法治理与算法信任

如果把思路从主观的信任感转到算法可信的客观属性,关于算法信任的第二种观点,即算法可信论,便呼之欲出了。该理论认为,算法信任就是对算法可信性的信任,法律规制算法信任,就是保障算法的可信性。算法可信论有两种不同的理论形态:一是技术性能可信论,认为算法信任就是对算法技术性能的信任;二是规范效果可信论,认为算法信任就是对算法运行所产生的规范性效果的信任。这两种可信论在本体论意义上都指向算法的可信性,但从前者到后者,在认识论意义上,却完成了从实体可信到程序可信、从直接论证到间接论证的范式嬗变。沿着程序性的间接论证思路,关于算法信任的第三种观点主张,算法信任就是对算法治理制度的信任,法律规制算法信任就是算法治理制度的自我建构,可称为算法治理制度可信论。接下来本文将论证,这两种观点都无法合理解释算法信任的法律性质:算法可信论不符合算法信任的社会关系维度,要么忽视、要么无法完整解释算法信任的规范性;算法治理制度可信论则不符合算法信任的概念维度,未能明确区分算法信任的必要条件与构成性要素。

(一)算法技术性能可信与算法信任

技术性能可信论强调算法技术性能的可信性,着眼于探究算法可信的客观属性及其法律保障。[7] 相较于算法信任感论,这种观点具有明显优势:矫正了算法信任感论依赖于主观感受之偏差,将算法信任纳入法律调整的范围;法律可以要求算法提供者为算法技术性能的可信性提供担保,类似安全生产等法律规定。

但是,另一方面,技术性能可信论也犯了和算法信任感论相同的错误,即都没有考虑到算法信任不是单方面的和事实性的,而是交互性的和规范性的。在传统技术应用场合,法律对其技术性能提出了要求,但法律规制的真正对象,其实是应用技术的人。换言之,主体和技术之间并不存在规范性的交互关系,所谓相信科技的说法,只是指科技性能安全

〔7〕 参见袁康:《可信算法的法律规制》,《东方法学》2021 年第 3 期,第 14-17 页;张欣:《从算法危机到算法信任:算法治理的多元方案和本土化路径》,《华东政法大学学报》2019 年第 6 期,第 26-30 页。

可靠,其实质是值得依赖或可放心利用,而非是法律意义上的信任。[8] 与传统技术不同,在算法应用场域,基于智能算法具有深度学习和自动化决策的能力,提供算法的并不是或并不仅仅是算法背后的人,而毋宁是智能算法本身或人机交互系统。[9] 也就是说,使用算法的人类主体与智能算法或人机系统之间,可以进行规范性互动并因此产生规范性交互关系。

技术性能可信论可能会提出两种辩护主张:一是认为算法与传统技术不存在本质差别,因而算法信任实质上就是算法依赖;二是认为存在单方面信任的情形,并无需对方的规范性回应,因而算法信任并不必然是交互性和规范性概念。

前一种主张的主要理由有三:其一,信任是一种人际关系,仅存在于理性行动主体之间,算法并非理性行动主体,因而只存在算法依赖,不可能存在算法信任;[10] 其二,算法系统缺乏信任所必需的意向性,既不能考虑自身利益,也不能考虑算法委托人的规范性诉求,更无法体验两者之间可能存在的冲突,因而算法信任不存在任何可以发展的空间;[11] 其三,算法系统是由数学事实、认知事实和物理事实构成的,三者都不涉及价值判断,所以算法与传统技术一样,只是一套价值中立的系统,不是作为规范性关系的信任之适格对象。[12]

但是,这种否定算法信任的主张及其理由是不成立的。首先,智能算法虽不具有人类的大脑和心灵结构,但可以从事原本由人类专属的理性活动,呈现出理性的功能,因此可成为拟制的主体。[13] 其次,智能算法虽不具有生物和化学结构,无法像人类那样产生原初意图,但是,智能算法却可以产生衍生意图,从而表现出意图性,自动化决策便是其典型例子。[14] 最后,算法并非技术中立的计算步骤,[15] 而是具有鲜明权力属性的价值系统。事实上,与科学不同,技术因为涉及人类应用,体现规范性目标,已非价值无涉。而智能算法从数据收集、数据分析到数据输出,每一个环节都渗透价值判断与选择,更是突破纯技术层面,成为一套知识和权力的生产和建构体系。[16] 综上可见,存在规范意义的算法信任并无不可逾越的障碍。但需说明的是,此处的算法信任,在当前技术条件下,实质是对人机交互系统的信任。把智能算法视为可被信任的主体,仅仅是基于智能算法具有某些

〔8〕 关于信任与依赖的区别之详细论述,参见[英]凯瑟琳·凯利著:《信任博弈》,唐甜甜译,东方出版社 2021 年版,第 9-11 页。

〔9〕 See Melvin Chen, Trust and Trust-Engineering in Artificial Intelligence Research: Theory and Praxis, 34 *Philosophy & Technology* 1429, 1437-1440 (2021).

〔10〕 See Diego Gambetta, *Trust: Making & Breaking Cooperative Relations*, Blackwell, 1998, pp. 213-238.

〔11〕 See Melvin Chen, Trust and Trust-Engineering in Artificial Intelligence Research: Theory and Praxis, 34 *Philosophy & Technology* 1429, 1432-1433 (2021).

〔12〕 See Melvin Chen, Trust and Trust-Engineering in Artificial Intelligence Research: Theory and Praxis, 34 *Philosophy & Technology* 1429, 1433 (2021).

〔13〕 See Melvin Chen, Trust and Trust-Engineering in Artificial Intelligence Research: Theory and Praxis, 34 *Philosophy & Technology* 1429, 1434-1435 (2021).

〔14〕 See Peter-Paul Verbeek, Morality in Design: Design Ethics and the Morality of Technological Artifacts, in Peter Kroes, Pieter E. Vermaas, Andrew Light & Steven A. Moore eds., *Philosophy & Design*, Springer, 2008, pp. 91-103.

〔15〕 See Tarleton Gillespie, Pablo J. Boczkowski & Kirsten A. Foot, *Media Technologies: Essays on Communication, Materiality and Society*, The MIT Press, 2014, p. 167.

〔16〕 参见林曦、郭苏建:《算法不正义与大数据伦理》,《社会科学》2020 年第 8 期,第 3-22 页。

特定类型的理性功能(例如生成式人工智能可进行深度推理),因而在日常用语和生活实践中可将其视为拟制的主体,但这并不意味着智能算法在法律意义上具有或应当具有拟制主体之地位。

后一种主张并不否认存在算法信任的可能性,只是否认算法信任必然是一种交互性的和规范性的关系,因而算法信任只依赖于算法的技术性能可信。其最主要和最有力的理由,便是存在单方面信任。也就是说,算法提供者可以在完全不知情的情况下,被算法使用者信任。应当承认,这种单方面信任的情形确实是存在的,但是,这并非信任的典型情形。而且,即使是在此种情形中,算法使用者也是作了如下假定的:如果算法提供者知道算法使用者的意图,算法提供者就会基于实现该意图而行动。更为根本的是,如果这种单方面的信任只是一种单纯的个人心理或情感,而不具有任何交互性行为的形式,那么,它就不是法律的调整对象。

综上所述,技术性能可信论的两个辩护理由难以成立。为了弥补技术性能可信论的不足,算法可信论将目光从技术性能转向规范性后果,提出了规范效果可信论。

(二)算法规范效果可信与算法信任

是否需要作出规范性回应,是智能算法与传统技术的根本区别之一。规范效果论关注算法运行的结果,要求该结果符合主体的规范性要求,而法律则应为此提供保障。根据关注的具体对象不同,规范效果论可以分为实体的规范效果论和程序的规范效果论。前者要求算法从数据输入、运算、再到输出结果的全过程都是可信的,算法信任就是相信算法应用不会做出侵害权利等违背规范性要求的行为。[17] 但是,因为存在无法完全消除的算法黑箱,难以对算法过程作出合理认知和判断,实体意义上的算法可信成为难以实现的目标。而且,信任并不要求完全的可知与确定,相反,信任总是与风险和不确定性相伴而行。因此,实体的规范效果论并非对算法信任的妥适理解。

与实体的规范效果论不同,程序的规范效果论不拘泥于算法过程,而重点关注算法输入和输出两端,尤其是输出端。[18] 如果算法的输入数据和算法结果均为正当,那么,无论算法过程是否透明和可理解,都可认为该算法可信。这可以说是一种算法统计与预测学,即通过对算法数据与结果的统计,计算其正当性的概率,只要该概率大于一定阈值,便被认为是可信的。相应地,算法信任的法律保障,就是要求算法提供者确保算法数据和结果达到特定的阈值标准。

然而,事实并非如此。例如,即使自动驾驶的安全性能已经远超人类驾驶,但是,人们对自动驾依然缺乏信任。这是因为信任不能还原为单纯的数据统计与结果预测。数据统计与结果预测是事实性的,而信任则是规范性的,程序的规范效果论没有说明算法效果与算法的规范性意图之间的关系,也就没有排除以下可能:即使一种算法的规范效果是达标的,但如果该算法的规范性意图不当,那么它就是不可信的。

---

〔17〕 参见袁康:《可信算法的法律规制》,《东方法学》2021 年第 3 期,第 17-18 页;张欣:《从算法危机到算法信任:算法治理的多元方案和本土化路径》,《华东政法大学学报》2019 年第 6 期,第 20 页。

〔18〕 参见陈景辉:《算法的法律性质:言论、商业秘密还是正当程序?》,《比较法研究》2020 年第 2 期,第 129-132 页。

问题在于,算法信任必然对规范性意图有所要求吗?如果有,是何种要求?是否要求出于良善的动机?在有些信任情形,似乎只在意行为的效果,而不问动机或意图。例如,消费者通过算法系统的排序功能,选择报价最低的商家,只是为了省钱,对于商家让利的动机与意图,则并不在乎。但是,在另一些信任情形,则对动机或意图有更高的要求。在此种情形下,如果仅仅从历史记录与相关数据的统计与预测结果来反推,并不能证明行为人的动机或意图,从而不能确定其是否值得信任。如果存在前述两种不同的情形,那么,算法信任是属于何种类型呢?对于这些问题,程序的效果可信论并没有给出论证。一个可能的原因在于,在程序的效果可信论看来,意图或动机存在于内心,即使对于人类而言也未能尽知,何况算法乎?由此便又只能回到程序论,根据后果来推论。沿着这个思路,第三种也是一种更为具体的理论模式被提出来了,这就是算法信任的治理制度可信论。

(三)算法治理制度可信与算法信任

算法治理制度可信论认为,如果关于算法治理的法律制度是可信的,则算法是可信的,因此,算法的法律保障就是算法法律制度的自我建构。在内容上,算法治理制度主要体现为算法透明、可解释、知情同意、可遗忘等一系列算法权利和问责制。需说明的是,在程序理论的视野下,这些主要体现为算法权利和问责制的算法治理制度,不应理解为实体性的,而更多地应视为是程序性的。[19] 制度可信论的观点认为,算法治理制度可信虽然严格而言并不等于算法可信,但是,算法治理制度却可为算法信任提供保障,在存在算法黑箱因而无法对算法是否可信作实质性判断的情形下,只要算法治理制度可信,便可认为算法是可信的。换言之,算法治理制度可为算法信任提供充分担保。

算法治理制度可信论正确地指出了,算法治理制度与算法信任之间,存在紧密的保障关系,但是,算法信任保障制度毕竟无法取代算法信任,正如朋友可信与朋友的朋友可信,虽非完全两回事,但也绝非完全一回事。两者至少存在两点区别:第一,一般的算法用户可以根据算法治理制度的可信性,来判断算法的可信性,但是,算法治理制度的制定者却必须正面回答何为可信算法的问题。第二,传统技术信任也依赖于制度保障,但是,正如前文所述,传统技术所谓的信任其实质是依赖,而制度可信论未能区分两者的本质差异。此外,制度可信论还忽视了一个重要的问题:算法权利和责任制是算法治理制度的主要内容,但是,并非所有的信任都依赖于权利和责任制,例如亲密关系中的信任。在亲朋、夫妻或恋人等亲密关系中,权利和问责不能成为信任的有效保障。

当然,制度可信论可能会辩护说,确实存在某些类型的信任,不依赖于权利和问责制,例如亲密关系中的信任,但是,算法信任并非这些类型的信任。亲密关系是一种私人关系,存在于相互了解的熟人社会,其信任关系往往依赖于社会舆论和内心道德律的约束,制度性的规制非为必要。但是,算法关系却是一种公共性关系,存在于彼此不熟悉的陌生人社会,而且,算法权力的滥用风险需要警惕与防范,因此,治理制度的保障必不

〔19〕 参见丁晓东:《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》,《中国法学》2022 年第 1 期,第 113-114 页。

可少。

制度可信论的这个辩护观点,仅就其本身而言是正确的,但却不是一个有效的辩护。因为即使证明了算法治理制度是算法信任的必要条件,也不是对算法信任法律性质的有效说明。其原因在于,性质的探讨实质上是一个概念论的问题,讨论算法信任的法律性质,就是讨论法律意义上算法信任的概念,即作为法律调整对象的算法信任,其构成性要素为何。必要条件与构成性要素,都具有不可或缺的含义,但两者是不同的概念。对于一个对象而言,必要条件是其存在的外在因素,但并非其成其所是的内在要素;而构成性要素则是其存在的和成其所是的内在要素。例如,阳光、空气和水分是生命的必要条件,但核酸、蛋白质等生命元素才是其构成性要素。如果算法治理制度的某些内容,对于算法信任而言,是不可或缺的,那么,这些内容是算法信任的必要条件,还是构成性要素?制度可信论未予以说明。

如果不区分必要条件和构成性要素,就无法区分某项制度要素,例如某项权利,究竟是算法信任在概念上的要求,还是基于其他原因例如必要保障或另一项权利的要求。区分必要条件和构成性要素的法律意义在于:如果是必要条件,算法信任的法律制度只需在整体上有所反映,但具有场域特征,因而允许根据具体情况作灵活调整,并非须臾不可或缺;但是,如果是构成性要素,那么法律制度就不是只需要在整体上有所反映,而是与算法信任共存亡,时刻不可分离。然而,制度可信论对此并无明确说明,因而在算法信任法律性质的概念维度上显示出明显缺陷。

三 作为回应性规范关系的算法信任

前文虽然未作综合阐述,但在多处提到,作为法律的直接调整对象,算法信任在性质上是一种回应性规范关系。结合本文开头提出的三个维度,讨论算法信任的法律性质,就是要揭示:作为一种回应性规范关系,算法信任的构成性要素是什么?前文已述,关于算法信任的法律性质的三种主要观点,与三个维度的要求均各有不合,有必要对这个问题作出新的回答。

(一)算法信任的规范构造:能力、意图与认知

法律意义上的算法信任是一个关系性概念。根据关于信任的通常观念,在算法信任关系中,存在算法信任是指 A(委托人,算法用户)信任 B(受托人,智能算法或人机交互系统)做 X(行动,自动化决策)。[20] 由此,我们可以将算法信任界定如下:

第一,A 信任算法 B,是指 A 相信算法 B 会去做 X。

在这个界定中,算法信任是指算法用户对算法的一种单方面的心理状态。但是,单纯的心理状态并非法律的直接调整对象,法律调整算法信任的目的在于引导理性的算法信任行动,形成理性的算法信任关系。理性的信任不单是委托人对受托人的主观态度,而且

〔20〕 See Warren J. von Eschenbach, Transparency and the Black Box Problem: Why We Do Not Trust AI, 34 *Philosophy & Technology* 1607, 1609-1610 (2021).

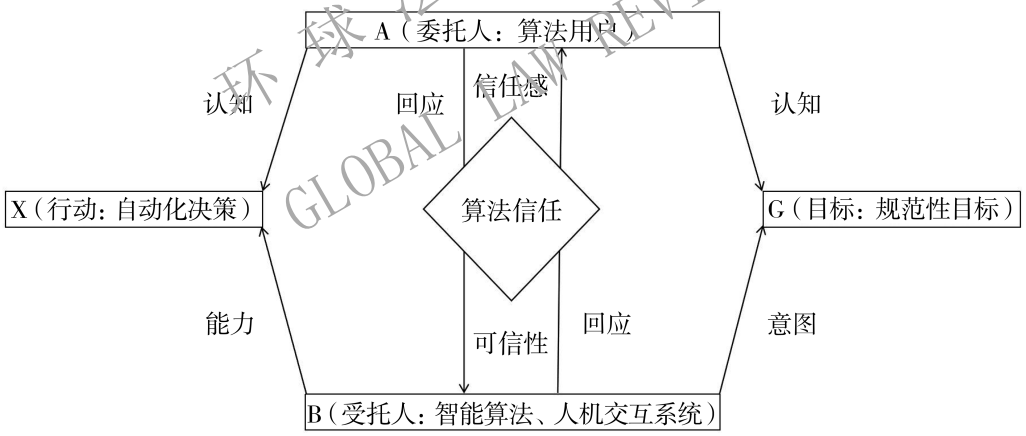
要求委托人有充分理由相信受托人。换言之,如果 A 要建立对算法 B 的信任,那么他就需要对算法 B 是否值得信任,亦即对算法 B 做 X 的可能性作出判断。[21] 只有当 A 有充分理由相信算法 B 会做 X 时才付出信任,这种信任才是理性的信任。据此,可以将算法信任进一步界定为:

第二,A 信任算法 B,意味着 A 有充分理由相信算法 B 会去做 X。[22]

根据这个界定,A 信任算法 B 是指 A 有充分理由相信算法 B 是可信的。那么,A 是根据什么来判断算法 B 是否可信呢? A 应当从能力和意图两个方面作出判断,即算法 B 是否有能力并有意愿代表 A 做 X 以实现其预期目标 G? 如果 A 对算法 B 的信任,是在充分判断且获得肯定回应后作出的,则该信任便是理性的信任。这里包含了 A 对算法 B 的能力和意图的合理认知,但依然是单方面的,忽视了算法 B 对 A 规范意图的认知与回应,不符合法律的调整对象为规范性交互关系之限定。因此,需补全算法 B 作为受托人之视角。这样,我们便可获得关于算法信任的一个比较完整的界定了:

第三,A 信任算法 B,是指 A 有充分理由相信算法 B 会去做 X。这意味着 A 有充分理由相信算法 B 是值得信任的,即相信算法 B 有能力去做 X,且算法 B 是为了 A 所期待的目标 G 而去做 X,且算法 B 就其能力与意图皆向 A 作出了合理的回应。[23]

根据这个界定,算法信任的构成性要素包括算法能力、算法意图和算法认知三个方面。[24] 据此,我们可以将算法信任的规范结构图示如下:[25]



(二)算法信任的构成性要素之一:实现规范性目标的能力

实现规范性目标的能力是算法信任的构成性要素之一。如果一个算法缺乏实现用户

〔21〕 See Diego Gambetta, *Trust: Making & Breaking Cooperative Relations*, Blackwell, 1998, pp. 213-238.

〔22〕 See Warren J. von Eschenbach, Transparency and the Black Box Problem: Why We Do Not Trust AI, 34 *Philosophy & Technology* 1607, 1609-1610 (2021).

〔23〕 See Warren J. von Eschenbach, Transparency and the Black Box Problem: Why We Do Not Trust AI, 34 *Philosophy & Technology* 1607, 1610 (2021).

〔24〕 需说明的是,这三个方面均包含了回应,为方便论述,本文不将回应当作一个独立的要素看待。

〔25〕 本文参考了陈梅文(Melvin Chen)的信任模型,并有所改动和补充,See Melvin Chen, Trust and Trust-Engineering in Artificial Intelligence Research: Theory and Praxis, 34 *Philosophy & Technology* 1429, 1430-1432 (2021)。

规范性目标的基本能力,该算法便不足以信任。需指出的是,虽然算法能力是算法信任的构成性要素,但应注意两点:第一,算法能力与算法信任不存在固定的对应关系,相反,算法能力与算法信任的关系具有群体性、个人性和场域性特征。〔26〕第二,算法能力与算法信任也不总是呈现正比关系。人工智能与算法技术的发展,并没有降低算法风险或增进算法信任。仅仅从技术能力的角度看,现代科技的发展水平肯定不是传统科技所能同日而语的。但是,现代科技也带来了迥异于传统技术的新型风险。〔27〕而且,这些风险更具根本性和整体性,对人类的主体性构成直接威胁,从而使人类陷入前所未有的生存性危机与恐惧之中。〔28〕技术的发展日新月异,但是,人类试图减少对风险的无助感、对风险来源的无知感以及对无法摆脱风险的无力感,却不仅丝毫没有改变,反而更加强烈。〔29〕然而,这并不影响算法能力成为算法信任的构成性要素。算法风险与技术进步如影随形,但是,除了寄希望于算法能力的持续改进,我们似乎也并无他法。

(三)算法信任的构成性要素之二:实现规范性目标的意图

法律调整的是意志行为,法律主体总是带着特定意图而行动。正如前文所述,智能算法可具有衍生性意图,因而算法与算法信任可成为法律的调整对象。但是,算法的意图是否依赖于良善动机,却存在理论分歧。动机理论认为,只有当算法是基于良善的动机去实现算法用户的规范性目标时,才是值得信任的,而非动机理论则持反对立场。

有两种理论方案支持动机理论。第一种可称为封装利益理论,其主要观点是:如果算法是出于实现自己利益的动机,且将算法用户的利益封装进自己的利益,以维持其与用户的关系,那么,该算法便是可信的。〔30〕但是,即便算法将用户的利益封装进自己的利益,也并不意味着他们的利益必然是一致的,例如当该利益的实现是以用户更大利益的牺牲为代价时。〔31〕封装利益理论可能会提出一个修正版本,来为自己辩护:算法不仅要封装用户利益,而且要封装其重大利益、长远利益和整体利益,甚至要求算法将用户利益置于自身利益之上,并应积极主动地谋求用户最佳利益。但是,该理论忽视了算法提供者与其用户之间可能存在的利益冲突,无法化解能否封装和如何封装的难题。而且,基于算法平台与用户在信息技术上天然的不平等,用户很难知道算法是否封装了其利益。

第二种方案是善良意志理论,认为算法必须出于善良的动机关心用户的利益,才是值得信任的。〔32〕在信任关系中,关心确实是重要的,因为可据此区分信任和单纯的依赖。〔33〕

〔26〕 参见[英]奥妮尔著:《信任的力量》,闫欣译,重庆出版社 2017 年版,第 10-11 页。

〔27〕 See Matthew U. Scherer, Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies, 29 *Harvard Journal of Law & Techology* 353, 353 (2016).

〔28〕 参见赵汀阳:《终极问题:智能的分叉》,《世界哲学》2016 年第 5 期,第 63-71 页。

〔29〕 参见[英]奥妮尔著:《信任的力量》,闫欣译,重庆出版社 2017 年版,第 19 页。

〔30〕 See Carolyn McLeod, Our Attitude towards the Motivation of Those We Trust, 38 *Southern Journal of Philosophy* 465, 465-479 (2000).

〔31〕 See Carolyn McLeod, Our Attitude towards the Motivation of Those We Trust, 38 *Southern Journal of Philosophy* 465, 465-479 (2000).

〔32〕 See Karen Jones, Second-Hand Moral Knowledge, 96 *The Journal of Philosophy* 55, 55-78 (1999).

〔33〕 See Annette Baier, Two Lectures on "Trust", in Grethe Peterson ed., *Tanner Lectures on Human Values* (Volume 13), University of Utah Press, 1992, pp. 109-174.

但是,关心是否出于善意,并不影响信任的合理建立。例如,对于商业平台而言,对用户的善意关心固然有助于提升企业商誉,但善意关心只是平台企业的道德义务,而非法律义务。况且,有时候算法的善意关心未必妥当与受欢迎。例如,在智能经济网络中,如果商业平台利用算法,禁止或限制"怀孕预测得分"高的客户购买酒精饮料,即便是出于对客户的善意和关心,也可能会引发他们对"算法家长主义"的排斥。〔34〕可见,善意的关心既并非算法信任的必要条件,亦非其充分条件。

鉴于动机理论的不足,学界提出了不同版本的非动机理论。在信任关系中,委托人总是对受托人将会做其所托之事有所期待。但是,这里的期待不是对结果的事实性预测,因为依赖关系也可以预测结果。相反,这里的期待,是希望受托人不仅仅按照我们假定的或预测的那样行事,而且按照他们"做其应做之事"那样行事。〔35〕这便是一种被称为规范期待理论所赞成的观点。需注意的是,规范期待理论所谓之"做其应做之事",可以容纳不同的动机,无论是出于良善的动机,还是仅仅是因为害怕惩罚,所以它是一种非动机理论。规范期待理论揭示了信任的规范性,但算法是基于何种动机真的不重要吗?如何理解这里的"规范性"?何为"应做之事"?对于这些问题,规范期待理论并没有作出回答。

但这些问题是必须回答的吗?有一种被冠名为承诺理论的观点就宣称,算法信任理论根本无需回答这些问题。它认为,受托人只需要按自己的承诺行事即可,无需向委托人作出规范性回应,而且,其承诺也不需要向委托人作出。〔36〕承诺理论是一种非动机理论,因为它对受托人是基于何种动机承诺均在所不问,它也无需了解委托人的期望,因而无需作出规范性回应。很明显,承诺理论是一种单方面的理论,不符合法律调整的基本原理。如果算法信任涉及的是交互性行为和社会关系,那么,一方面,算法受托人应该了解算法用户的规范性目标,并向其做出回应;另一方面,算法用户也应当有合理途径,了解算法受托人的能力及意图,否则算法信任便难以建立。即便承诺理论作出修正,加上回应的要素,也并未足够。因为算法仍需对何为真正的规范性需求进行实质性判断。例如,在某些信任情境中,为了实现委托人的规范期待而应当超越承诺。〔37〕我们期待真正值得信任的智能护理机器人,在发生火警时懂得协助阿尔茨海默病患者逃离现场,而不是机械履行防止老人外出之承诺。

由此,似乎又回到了规范期待理论,并需正面回答它试图回避的那些问题。如果回到规范性期待与回应,就难免涉及实质性价值评价,因此算法不能以眼前利益、局部利益与轻微利益之名,而让算法用户付出长远利益、整体利益与重大利益受损的代价;有时候算法要真正做到不负信任,可能需要超越字面意义上的承诺。这也说明,对规范性的理解,不能局限于算法用户的主观期望。但同时,也需要警惕管制型家长主义的陷阱,慎防世俗

〔34〕 参见[德]克里斯托弗·布施:《个性化经济中的算法规制和(不)完美执行》,《环球法律评论》2019 年第 6 期,第 10 页。

〔35〕 See Margaret Urban Walker, *Moral Repair: Reconstructing Moral Relations after Wrongdoing*, Cambridge University Press, 2006, pp. 77-79.

〔36〕 See Hawley Katherine, Trust, Distrust and Commitment, 48 *Noûs* 1, 1-20 (2014).

〔37〕 See Kirton Andrew, Matters of Trust as Matters of Attachment Security, 28 *International Journal of Philosophical Studies* 583, 584-588 (2020).

权力与资本通过算法之手,侵犯算法用户自由与权利的疆界。此外,还需考虑到算法及其提供者与算法用户之间、不同算法系统之间和不同算法用户之间可能存在的利益冲突。两相权衡的中庸之道,或许可以为算法设定应基于公平与互惠原则,作规范性考量之法律义务。亦即算法必须愿意遵循这个标准,做算法用户所期望之事,并且应以合理的方式,向算法用户作出积极回应。例如,即使不能直接禁止过度消费或禁止可能已经怀孕的妇女购买酒精饮料,但可以在知情同意的基础上发出提醒或警示。鉴于算法在特定情形中需遵循公平互惠的原则作实质性的价值考量,虽然受应避免管制型家长主义的约束,但显然对算法信任提出了颇为积极的要求。因此,可以把这种规范期待理论称为积极的或负责任的规范期待理论。

在不少国家或地区的立法中,都可以找到相关规定,可视为积极的或负责任的规范期待理论之制度体现。例如,欧盟《一般数据保护条例》(GDPR)第 6 条列举了数据处理的六种合法情形,其中第 d 和第 f 项分别提到,数据处理是为"保护数据主体或另一自然人的核心利益所必需"和"实现控制者或者第三方所追求的正当利益所必需";美国《关于安全、可靠、值得信赖地开发和使用人工智能的行政命令》第 1 节即提到,该行政命令之目的是为了促进负责任的人工智能;我国《个人信息保护法》第 5 条规定了处理个人信息应当遵循合法、正当、必要和诚信原则。但是,以上多为原则性规定,其目标如何实现与保障,仍缺乏具体规定。近年来,有学者提出数据信托或数据信义模式,是一个值得重视的方案。该模式认为,受托人应以谨慎和忠实的态度,全面充分地保护委托人之利益。[38] 但是,数据信托目前还主要停留在理论设想阶段,各国的立法经验亦尚未成熟,而且存在各有利弊的不同模式,其中,应由何种主体担任受托人之角色,分歧与差异甚大。[39] 数据信托是否可行,以及何种模式更适合我国,还有待进一步探索。但是,算法与数据的受托人应为委托人规范目标之实现,承担积极的或负责任的义务,当为相关立法题中应有之义。

(四)算法信任的构成性要素之三:对算法能力与意图的认知

算法信任作为一种社会关系,建立在双方主体的交互性行动之上,这依赖于算法用户对算法的能力与意图的认知,故算法认知为算法信任的构成性要素之一。算法认知属于算法信任的认识论,旨在回答信任的合理性问题,即何时应信任,何时不应信任。该信任时不信任,或者不该信任时却信任,都会造成不良后果。因此,何为合理的信任便成为重要的问题。对于合理信任的标准,在理论上存在两种不同的立场:在合理理由的鉴定问题上,存在真理导向和结果导向的对立;在认知责任的主体问题上,存在内在主义和外在主义的争锋。

首先,合理信任是否需要充分的理由与证据?真理导向的合理信任观支持一种认知理性,[40] 认为只有当人们有充分理由或证据相信时才相信,该信任才是合理的;如果缺乏

〔38〕 参见唐林垚:《人工智能时代的算法规制:责任分层与义务合规》,《现代法学》2020 年第 1 期,第 199-204 页;解正山:《数据驱动时代的数据隐私保护——从个人控制到数据控制者信义义务》,《法商研究》2020 年第 2 期,第 80-84 页;翟志勇:《论数据信托:一种数据治理的新方案》,《东方法学》2021 年第 4 期,第 61-76 页。

〔39〕 参见翟志勇:《论数据信托:一种数据治理的新方案》,《东方法学》2021 年第 4 期,第 72-75 页。

〔40〕 See Baker Judith, Trust and Rationality, 68 *Pacific Philosophical Quarterly* 1, 1-13 (1987).

事实上的可信性,即使后果有利,也不应信任。[41] 根据这种观念,合理的算法信任是建立在算法客观可信、且其可信性完全可知的前提之上的。但是,这既不符合信任的脆弱性,也不符合算法信任的广泛生活实践。

相反,结果导向的理性观则主张一种战略理性,[42]认为虽无充分证据证明,但当结果有利时信任便是合理的,而无论其事实上是否真的可信。[43] 但是,信任的目的并不能为信任提供理由,关于信任的有用性或价值考虑,与可信性或真实性无关。[44]

如果我们承认合理的信任不仅仅是一种主观态度,而且是一种客观属性,那么,就必须承认真理导向的观念有其正确的一面。[45] 事实上,虽然结果导向意义上的合理性不依赖于有充分证据证明的可信性,但包含着一种对规范结果和可信性的合理期待,即虽无充分证据证明其可信性,但依然可合理期待其可信。[46] 否则,便很难说是一种信任,而毋宁是单纯的依赖或利用。因此,真正的合理算法信任或多或少总是真理导向的,它依赖于理由,该理由来源于算法的可信性或对其可信性的合理期待。

其次,算法用户需要自行承担认知责任吗?内在主义认为一个人应对自己的信任承担认知责任,委托人应搜集并评估信任的直接证据以形成内在理由,特别是在信任将使其极其脆弱时,更应如此。[47] 在算法制度上,这一观念主要体现为个人赋权制度,希望通过赋予个体一系列数据权利,并经由个体的权利行动,营造干预自动化决策的合理规制空间。[48] 但是,这显然科予了算法用户过重的认知责任。假定数据主体是理性且自主的人,只是一个陷入"自治陷阱"而产生的"控制错觉",[49]将因遭遇个体能力不足、算法结构难以理解和权利抗辩等障碍,而难以实现。此外,该观念否认算法用户在无法或难以获得内在理由的情形下,仍可存在理性算法信任的可能性。

与内在主义相反,外在主义认为委托人无需收集和评估信任所需的直接证据并形成内在理由,相反,只要存在信任的原因,也即只要具有间接证据并可形成外在理由,便可合理信任。[50] 例如,基于委托人的历史评价或专家推荐而信任。正是基于这种理解,外在

〔41〕 See Hieronymi Pamela, The Reasons of Trust, 86 *Australasian Journal of Philosophy* 213, 213-236 (2008).

〔42〕 See Baker Judith, Trust and Rationality, 68 *Pacific Philosophical Quarterly* 1, 1-13 (1987).

〔43〕 See Frost Arnold Karen, Trust and Epistemic Responsibility, in Simon Judith ed., *The Routledge Handbook of Trust and Philosophy*, Routledge, 2020, pp. 64-75.

〔44〕 See Hieronymi Pamela, The Reasons of Trust, 86 *Australasian Journal of Philosophy* 213, 213-236 (2008).

〔45〕 See McGeer Victoria & Philip Pettit, The Empowering Theory of Trust, in Faulkner Paul & Thomas Simpson eds., *The Philosophy of Trust*, Oxford University Press, 2017, pp. 15-35.

〔46〕 See McGeer Victoria & Philip Pettit, The Empowering Theory of Trust, in Faulkner Paul & Thomas Simpson eds., *The Philosophy of Trust*, Oxford University Press, 2017, pp. 15-35.

〔47〕 See Elizabeth Fricker, Critical Notice: Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony, 104 *Mind* 343, 393-411 (1995).

〔48〕 参见张欣:《从算法危机到算法责任:算法治理的多元方案和本土化路径》,《华东政法大学学报》2019 年第 6 期,第 21-22 页。

〔49〕 参见解正山:《数据驱动时代的数据隐私保护——从个人控制到数据控制者信义义务》,《法商研究》2020 年第 2 期,第 78-79 页。

〔50〕 See Carolyn Mcleod & Emma Ryman, Trust, Autonomy, and the Fiduciary Relationship, in Paul B. Miller & Matthew Harding eds., *Fiduciaries and Trust: Ethics, Politics, Economics, and Law*, Cambridge University Press, 2020, pp. 74-86.

主义支持外部问责的算法治理思路,将对算法信任的认知诉诸专业性的行政机构或其他外部监督主体,[51]以应对内在主义和个人赋权模式的不足。但是,外在主义不能解决算法信任的核心问题,即算法的可信性最终只依赖于可信性本身。间接证据和外部理由虽然有时候确实可以成为信任的合理依据,但是,它们要具有证据和理由的效力,也必须和可信性本身建立起意义关联。而对间接证据和外部理由的证据和理由资质的判断,离不开内在主义的认知路径。于是,外在主义和内在主义在这里便合流了。但是,我们又该如何判断呢?内在主义的困境于此又再度浮现。

其实,外在主义虽并不要求委托人具有直接证据和内在理由,但对受托人依然抱有规范性期待,即认为有合理原因相信受托人会做其应做之事。这样,我们或许可以在内在主义与外在主义之间,找到一条综合主义的认知路径,即以下两种情况都属于理性的算法信任:算法用户基于内在理由而相信算法将做其应做之事;或者算法用户基于外在理由而相信算法将做其应做之事,且算法用户有充分的理由相信,该外在理由是建立在内在理由的基础之上的。在算法信任的制度设计上,可作二阶式的程序性规定:首先,算法提供者应提供内在主义的认知途径,即为算法用户提供可以获得和认知算法可信直接证据的简易途径,以帮助算法用户形成内在理由;其次,因为算法用户的能力不足或认知成本过高等合理原因,以致上述要求和目标难以达到时,则应提供外在主义的认知保障,例如向专家开放、公开审计和历史记录和接受政府与公众监督等。

在各国算法治理的法律实践中,多采此二阶相结合之规制模式。例如,欧盟《一般数据保护条例》第三章关于数据主体享有透明与告知、访问、纠正和删除、拒绝和自主决定等权利的规定,我国《个人信息保护法》第二章关于个人信息主体享有的知情同意与撤回同意等权利的规定,即属内在主义之认知路径;欧盟《一般数据保护条例》第三章第 31 条和第 33 条为控制者分别设定了和监管机构合作以及向监管机构通知个人数据泄露的义务、第四章第四节和第五节则分别设置了数据保护专员和认证制度等规定,我国《个人信息保护法》第六章规定了履行个人信息保护职责的部门,在各自职责范围内负责个人信息保护和监督管理工作,则体现了外在主义之认知路径。需强调的是,无论是内在主义还是外在主义抑或是两者的结合,要建构理性的算法信任,都必须证明与算法本身的可信性存在直接或间接关联。一旦这种关联丧失或减弱,理性算法信任也将随之丧失或减弱。

四 算法信任法律治理的制度设计

建立和维护良好的算法信任,是一项系统工程,需要多方协同。就法律治理而言,算法信任依赖于法律保障,算法信任相关制度的建构,需要遵循法律的制度理性与法律的规

〔51〕 参见张欣:《从算法危机到算法责任:算法治理的多元方案和本土化路径》,《华东政法大学学报》2019 年第 6 期,第 23-25 页。

制逻辑,并注重塑造制度的韧性空间。

(一)算法信任的法律建构应遵循制度理性

不同于亲密关系与私人之间的信任,算法信任作为一种公共性的社会关系,不可能寄希望于内心情感与伦理道德的力量,而必然仰仗社会制度,尤其是法律制度的保驾护航。因此,我们制定了许多制度和标准,例如算法和数据监管、算法权利和算法责任制等,为算法信任搭建了基本的制度框架。但是,需防止两种制度失灵。其一是制度不足,即法律应予规制而未规制。例如算法监管权力的缺位、算法权利的不足和算法责任制的虚弱。目前,各种算法操纵与算法侵权等算法风险无处不在;知情同意、算法透明、遗忘与可删除等算法权利保障乏力;算法审计与算法问责困难重重;凡此种种,皆说明算法信任保障制度之不足。其二是制度过度,即法律不应规制的却规制了。算法信任的法律制度不是越多越好、越严越好,制度的建构不等于制定更多更严的法律。事实上,我们正面临尴尬局面:付出了巨大和昂贵的努力,将越来越多的技术用于防范和侦测背信行为,但却效果不彰,不信任的气氛反而不断蔓延。[52] 我们应记取"法密如凝脂,犹漏吞舟之鱼"的古训,以免"法令滋彰,盗贼多有"。权力总有扩张的冲动与滥用的风险,这既适用于算法权力,也适用于监管算法权力的权力。制度的制定者与执行者应避免沉溺权力任性而陷入制度迷思,既不能过度监管,侵害算法提供者与算法用户的权利,又不能对资本和算法权力放任不管。

(二)算法信任的法律建构应遵循法律的规制逻辑

作为法律的直接调整对象,算法信任既不是指信任感,也不是指算法的可信性或算法治理制度,而是指算法提供者和算法用户之间的算法委托关系。这种关系是一种回应性规范关系,通过双方主体的外在的交互性行为而实现:首先,算法用户向算法提供者表达使用算法的规范性目标;其次,算法提供者对该规范目标作出回应,亦即向算法用户展示其实现该规范目标的能力和意图;再次,为顺利开展以上交互性回应行为,算法提供者应确保提供合理的认知途径和正当的操作程序;最后,基于前述三点,算法用户理性作出是否信任算法的决定。法律规制算法信任,就是保障这种回应性规范关系的有效实现。

需说明的是,这里所说的算法提供者,在当前智能技术条件下,主要指人机交互系统,即除了智能算法自身外,还包括与算法有关的人员,例如算法系统的设计人员、程序员、数据主体、操作员、检查人员、监管或审计人员等,他们通过彼此之间以及与算法系统的复杂互动,并基于"循环信息的因果关系",产生指向特定目标的行动、决策或效果。[53]

(三)算法信任的法律建构应塑造制度韧性

算法信任的法律治理制度,要想获得持久的有效性,应注重塑造制度的韧性空间。

---

〔52〕 参见[英]奥妮尔著:《信任的力量》,闫欣译,重庆出版社 2017 年版,第 1-23 页。

〔53〕 See Massimo Durante, What is the Model of Trust for Multi-agent Systems? Whether or Not E-Trust Applies to Autonomous Agents, 23 *Knowledge Technology & Policy* 347, 355 (2010); Warren J. von Eschenbach, Transparency and the Black Box Problem: Why We Do Not Trust AI, 34 *Philosophy & Technology* 1607, 1619 (2021).

第一,应保障算法用户的充分选择权。无选择,不信任。如果人们在没有选择余地的时候选择接受,可能只是单纯的依赖,而非信任。例如,我们乘坐交通工具、饮用自来水、去医院看病、呼吸空气等,就只是依赖,很难说是出于信任。〔54〕 由此,可以得出两个推论:其一,仅仅依据算法用户接受或使用算法的行为,不能证明他们是出于对算法的信任,因而,使用算法的数量增加与算法信任之间,也不存在相互推导关系;其二,为了维护良好的信任关系,算法应为其用户提供足够的选项,并保证算法用户可以便捷行使选择权。

第二,应以保障算法的规范性要素为目标指向。算法信任的规范性或构成性要素,包括算法能力、算法意图和算法认知。算法信任离不开算法赋权、算法义务和算法问责制度的保障,但制度保障只是算法信任的外在条件,从算法信任构成性要素的角度看,只有算法规范性要素的充分满足,才是实现算法信任的根本。就此而论,算法赋权、算法义务和算法问责制度只是算法信任的中间物,发挥工具性和过程性的作用,而非算法信任治理制度的最终目的,相反,它们都应该以保障算法规范性要素的实现为目标指向。

第三,应平衡算法权力与用户权利并向用户权利倾斜。算法权力与算法用户权利,存在天然冲突,需在两者之间取得精妙的平衡。一方面,算法信任离不开算法义务与算法问责,但是,我们需要避免陷入管理陷阱。课予算法过高与过多的义务和责任,容易导致算法违背应以积极负责的态度实现委托人规范性目标之要求。例如不当限制算法的自由决策空间,会导致以避免犯错与担责、而非以追求最佳效果为目标导向的消极态度等。〔55〕但是,另一方面,也必须看到,我们当下面对的现实是,人们之所以对算法缺乏信任,很大程度上并不是因为对算法技能的怀疑,而更多的是对算法利维坦与平台霸权、甚至是算法权力与传统权力合谋的忧虑与担心。在这种情形下,法律制度对算法权力的有力规制,并向算法用户权利的倾斜保护,便成为必要和紧急之事。以算法透明为例,为了防止算法滥用,维护算法安全与信任,应让算法提供者承担更多的算法透明义务与责任,而不是让算法用户承担;当试图将信息公开、实名认证、人脸识别等要求施加于算法用户时,需要充分的论证与审慎的权衡。

五 结 语

我们生活在一个算法依赖与算法风险并存的时代,如何保证算法可信,维护良好的算法信任,是时代的挑战。算法信任法律性质的理论探讨,只是其中必要的一环。算法信任是一项复杂的社会系统工程,也并非仅涉法律治理这一个方面。算法信任是社会信任的一部分,算法信任的有效建构,最终还是依赖于全社会的信任环境。

[本文为作者主持的 2022 年度司法部“法治建设与法学理论研究”项目“理性算法信任的法律规制模式研究”(22SFB3005)的研究成果。]

〔54〕 参见[英]奥妮尔著:《信任的力量》,闫欣译,重庆出版社 2017 年版,第 13-17 页。

〔55〕 参见[英]奥妮尔著:《信任的力量》,闫欣译,重庆出版社 2017 年版,第 47-54 页。

The Legal Nature of Algorithmic Trust

[**Abstract**] The discussion on the legal nature of algorithmic trust aims to answer the question of what we actually trust when we talk about algorithmic trust in a legal sense or what is actually guaranteed when we say that laws should provide guarantees for algorithmic trust. There are three main viewpoints in the academic community regarding this question: first, algorithmic trust refers to the sense of trust in algorithms; second, algorithmic trust should be understood as algorithmic trustworthiness; and third, algorithmic trust is the trust in algorithmic governance systems. Correspondingly, the law provides guarantees for trust in algorithms, ensuring the sense of trust in algorithms, the trustworthiness of algorithms, and the effective governance of algorithms. Unlike traditional views, this article will provide a defense for algorithmic trust as a responsive normative relationship in terms of its legal nature, and clarify its normative construction and specific connotations. This article points out that the discussion on the legal nature of algorithmic trust should include three dimensions: legal, social relationship, and conceptual. The theory of the sense of trust in algorithms refers to an emotional or psychological state of believing in algorithms. Therefore, legal regulation or adjustment of algorithmic trust aims to enhance people's sense of trust in algorithms. This approach ignores the special nature of legal regulation. Algorithmic trustworthiness theory holds that algorithmic trust is the trust in algorithmic trustworthiness, and legal regulation of algorithmic trust is to ensure the trustworthiness of algorithms. The trustworthiness theory of algorithmic governance system holds that algorithmic trust is the trust in the algorithmic governance system, and legal regulation of algorithmic trust is the self construction of the algorithmic governance system. Neither of these two views can reasonably explain the legal nature of algorithmic trust. This article provides a defense for algorithmic trust as a responsive normative relationship. As an object of legal regulation, algorithmic trust is a responsive normative relationship that arises in the interactive behavior between algorithm providers and algorithm users. Algorithmic trust is the trust that can be effectively formed in such relationships. In terms of normative construction, algorithmic trust consists of three constitutive elements: the reliable ability of algorithms to achieve the normative goals of algorithmic users, positive and responsible intentions, and the convenient and reasonable cognition of algorithmic capabilities and intentions by algorithmic users. This article argues that establishing and maintaining good algorithmic trust is a systematic project that requires multi-party collaboration. In terms of legal governance, algorithmic trust relies on legal protection. In the construction of algorithmic trust related systems, we need to follow the institutional rationality and regulatory logic of the law, and pay attention to shaping the resilience space of the system.

(责任编辑:支振锋)