引用本文：李雁华，杜海涛. 技术如何“信任”人——算法信任视域下智能风控正义的审思[J]. 河海大学学报(哲学社会科学版)，2022，24(4)：40-46.
DOI：10.3876/j.issn.1671-4970.2022.04.006

技术如何“信任”人
——算法信任视域下智能风控正义的审思

李雁华[1,2]，杜海涛[3]

(1. 华东师范大学哲学系，上海 201100；2. 青岛大学马克思主义学院，山东 青岛 266100；
3. 西北师范大学哲学系，甘肃 兰州 730000)

摘 要：现代人工智能的“可信任”研究通常关注的是人对智能算法的信任问题。但在风控领域逐渐转向 AI“合理怀疑”能力的要求中，算法对人的“信任”也已成为人工智能伦理的重要问题。风险防控本身就是如何信任一个人的问题。在现代未来学思维下，人把对信任与怀疑的规划都交付给技术，以降低风险率。人工智能风控技术以数据和算法作为信任人的方式，这是智能时代“后人类”的一个必然的社会深度脱域特征。因而，人工智能风控相对于其他智能技术需得到独特的哲学关注。但是这种算法信任主导下的风险防控，必然面对正义性的考量。在人工智能物逐渐被作为一种能动主体的情境下，实现其在风控场景中对人的“合理怀疑”也成为智能风控的基本伦理要求。

关键词：风险技术；智能风控；技术正义；算法信任
中图分类号：B82-057 **文献标志码**：A **文章编号**：1671-4970(2022)04-0040-07

一、风险思维与智能时代的风险控制

自 20 世纪 70 年代开始，风险研究已经发展成为一个重要的跨学科研究领域。按照吉登斯的观点，风险是现代性研究中一个相当重要的概念，他说：“在一个抛弃了旧有的、传统的行事方式的社会，以及在一个完全面临充满不确定因素的未来的社会，风险便成为有着重大意义的核心概念。”[1]104 之所以如此，是因为现代社会完全抛弃了命运、宿命等传统决定论的理解工具，个体命运的不确定性与社会开放性融为一体。风险之所以成为一个现代社会的凸显特征，其思维的基底是人类理性化进程中形成的对改造外在世界的自信。风险本体论与未来学思维主导现代风险评估的技术化，正是基于这种风险“悬临”的意识，导致现代对风险的认知几乎渗入到每个人的行动中，这便有了吉登斯所谓的“对未来的殖民”的可能性，因为这种未来学思维下的风险意识已成为防控成本的一种投资。由此可见，社会风险领域的特征恰恰是人与人关系的不确定性导致的不可预计的概率事件增加。在一个相对固定的熟人社会中，风险不可能是核心特征，正是社会的脱域程度加剧所带来的交往形式多元化，使得风险成为处理社

基金项目：2020 年甘肃省社会科学规划项目(20YB041)
作者简介：李雁华(1991—)，女，讲师，博士研究生，主要从事比较伦理学研究。E-mail：740048505@qq.com

会问题的一个重要视角,与之相关的风险评估与风险防控也成了现代社会的重要研究内容。

风控行为是风险社会的必然产物。现代性的一大特点是风险类型从前现代的地震、洪涝、疾病等自然风险扩大到金融、经济、市场、技术发展等人为风险。而由诸多人为风险促导的“未来学”思维,开始以技术化方式对抗运气依赖的不可控性,科学化与技术化的风险管理进而成为规避运气的重要操作,如张康之所说:“由于技术、知识和谋略的进化使得人们比历史上的任何时候都有着更强的控制矛盾和解决问题的能力,似乎人们能够在对各种各样矛盾的解决中避免社会风险和降低社会风险。”[2]一种培根式的技术乐观主义也认为:“科学技术且只有科学技术才降低或消除了人类活动的风险,使人能轻易做一些先前须付出一定代价才能完成的事情。”[3]在资本主义市场经济中,关于金融市场投资和活动的决定必须根据所涉风险的背景来理解。因此,注重技术化模型的现代经济理论必然需要建构风险的技术化模型,风控领域统计学意义上的技术化起步完成于20世纪50年代提出的标准差风险指数方法。

上述对风险的技术化控制认知,为当今人工智能发展成为新的风控方式提供了背景。人工智能风控的特点在于运用大数据分析的方式防控风险的发生,现今社会管理、工程管理、医院护理系统、网络安全,以及金融信用系统等人工易出错的领域都已经广泛采用人工智能。人工智能风控的特点是精准性分析和“绝对命令式”的预制模式,这在很大程度上可以弥补传统风控的弱点,如英美等国家提出的智能风控构想,即监管部门的技术系统直接连接每个金融机构的后台系统,实时获取监管数据,运用风险数据聚合、数据模型类型化分析与预测、监控支付交易等。因此,智能风控也成为现代技术化风控的新路径,是现代风险思维的必然产物。

社会信用体系或保险行业的风险防控成为当代风控的重要形式。吉登斯认为,保险的未来学思维是对未来殖民的重要表现,是一种对可能的风险和意外投资的预先防范。而金融业、银行业风控其实也已成为具有新的社会现象意义的风控类型。信用卡或信贷都需要银行提前承担客户失信的风险,而为了应对此种风险,需要个人失信记录的数据化统计,也即是“黑名单”制度。这一风险防控的技术类型属于数字资本主义下信用数据商品化。技术统计的意义在于减少银行信贷的运气性成分,以信用数据作为参考,能够防控一定程度的失信风险。但是,仅以失信数据作为参考对象是一种被动防控,它是在失信行为发生之后进行的防范行为,必然存在着风控偏差。因而,出现了对客户偿还能力分析等技术化手段。

随着互联网科技与金融高度融合,互联网科技这种轻资产、重服务的网络模式正慢慢渗透到金融模型中,对传统金融业务产生了鲶鱼效应和示范效应,推动了金融机构的变革。由于网络虚拟环境的信息不对称、交易过程透明度低、信息安全无法得到保障,金融机构面临的道德风险、市场风险、信用风险越来越突出。因此,在大数据时代,人工智能进入风控领域是一种技术化发展的必然趋势。传统金融机构采用评分卡模型和规则引擎等“强特征”进行风险评分,而智能风控(robo risk control)根据履约记录、社交行为、行为偏好、身份信息和设备安全等“弱特征”进行用户风险评估。两种风控方式从操作到场景都呈现出了明显的区别化效应,如表1所示:

表1　两种技术化风控方式差异比较

两种技术化风控方式	能动性差异	方法与依据	信息来源
社会信用体系的信用数据化	被动	评分卡模型和规则引擎等	失信记录、偿还能力、还款意愿
大数据时代智能化的风险治理	主动	大数据算法	履约记录、社交行为、行为偏好、身份信息和设备安全等

人工智能在风控领域的综合运用体现了一种思维形式的变革,它体现了数据时代人类对劳动成本和劳动效率的综合把控能力。它不仅要对一般银行信息进行数据化分析,而且要在大数据库中,把个人信息生活痕迹等作一综合性考量,并将个人生活事件进行风险等级判定,最终测算出个人除信用记录、偿还能力之外的失信可能性。由此可见,技术化进步体现了人类未来学思维对社会变革的一次推进。未来学思维是现代个人、企业、政府面临风险社会不得不具有的一种思维形式。而技术发展主导了人类对未来风险控制的信心,一种依靠技术控制未来风险的情绪也成为极盛现代性的典型标志。因而,“大数据规控风险”定然是风控领域的必选尝试。同时,这也是人将“信任”能力进一步交给机器的尝试。

二、智能风控与技术化信任

现代大数据智能风控的本质是技术如何“信任”人的问题,也可以说是人选择将信任人的能力移交给技术。这种思路是现代性的一大特征,即数据的精确化要高于人的判定,它可以排除情感、好恶等主观因素,代表了高度理性化的分析形势。吉登斯说:“现如今,专家思维和公共话语的一个明显构成部分便是‘风险情形分析’(risk profiling)——在当今知识状态和当前条件下,分析风险在给定场景中的分布情况。”[1]111 随着现代性脱域机制的形成,人与人的直接信任遭到解构,因而,技术以及专家知识成了信任得以维持的手段。这些都体现出人对自身与风险防控能力的不自信,进而将信任转交给技术。吉登斯的话语背景是 20 世纪,人工智能尚未兴起,实际上技术的发展正印证了他所说的技术革命的速度比人类历史上任何政治、经济和社会革命都快很多。现在人工智能越来越广泛全面地应用在风险控制上,这其实体现了一种新的“未来学”自信。它比现代人类对疾病风险、灾难风险的控制和评估都更为特殊,因为对于疾病等风险仅能根据一种概率性直觉诉诸保险等预防手段。而大数据时代则可以根据大量的数据实现精准的风险管理。但问题是在这种以智能物对人怀疑为前提的算法操作中,技术该如何“信任”人?

大数据风险判断的基础是“数据痕迹”,与传统征信活动所不同的是,大数据不仅包括传统的信贷数据,同时也包括了与消费者还款能力、还款意愿相关的一些描述性风险特征。利用大数据技术,可以搜集许多数据维度来描述风险,作为风险评估的重要依据,这样可以更大限度地解决银行机构与普通人之间的信息不对称问题。传统银行无法获取用户的征信信息,人工智能通过技术、数据的手段可以构建出一个信用分析模型。因此,对于银行、风投公司、保险公司等企业而言,智能算法为风险评判提供了新的技术支持。“算法决策具有专业性、复杂性和动态性特征。结合具体应用,算法主要发挥着优先级配置、分类、关联及过滤四项功能。”[4] 因此,依靠算法可以满足经济性、安全性、客观性等要求。而银行等企业对算法的信任与算法对人的“信任”构成了风控与风评的核心环节。

一般而言,“信任”一个人需要依靠其过往信息的在场性[5],信任本身就是依靠信息的在场而取消怀疑的心理过程。同时由于现代社会的脱域性加剧,信任本身已经成为关联着巨大社会资源的心理要素①。对于人工智能领域而言,“人-机信任”已成为当前最为重要的研究主题之一,它关涉智能技术良性发展的伦理基础。如欧盟委员会就曾发布“可信任人工智能的伦理框架”(Ethics Guidelines for Trustworthy AI)[6],确立了人工智能可信任机制的基本原则与路径。闫宏秀指出:“在人工智能的背景下,人与技术的关系是基于某种信任而形成的一种新型深度融合。这种信任是人与人工智能关系存在的前提条件。但这种信任除了传统意义上的人际信任之外,还涉及人对人工智能以及人工智能系统间的信任。”[7] 而相关心理研究发现造成人机信任困难的是人工智能的非物理性存在形式,如斯特拉特认为,面部的非言语暗示对于信任有促进作用,如稳定的面部特征,如吸引力、相似性和感知可信度是信任的重要影响因素[8]。但实际上,人-机间的信任困境不仅在于这种心理因素,而且还在于对算法不透明性的伦理质疑。

在风控领域,一个有趣的现象是,相关风控企业机构越来越偏向于信任智能物对人的“信任”,呈现出一种如图 1 所示的信任方式。如果说公众对于算法的合理质疑是推动人工智能伦理发展进程的动因,那么,企业对智能算法的过度信任也会成为智能物自我规约的阻碍,后者依然反映着传统资本逻辑驱动的典型正义问题。其实,人工智能对人的信任,本质上仍是人际信任在人机互动上的反映,它也需要信息的“在场性”,如机器“信任”人需要大量信息在场的依据,安全系统依赖于技术数据的记录,以及对有前科的人的犯罪防控。大数据时代,消费记录、信用记录都成为判别人“善恶”的依据。

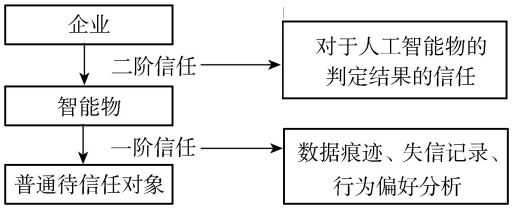

图 1　大数据时代技术化风控中的信任方式呈现

① 如吉登斯、福山等人都注意到现代社会风险与信任的同构关系,信任不仅是构成社会联结的重要心理,而且获得信任也成了社会资源凝聚的内在动因。参见弗朗西斯·福山《信任》一书。

由此,数据记录频次已经成为风险的重要依据。但问题是数据应该如何成为风险判定的基本依据?在技术通过数据“信任”人的系统中,技术对人的信任成为人对人信任的直接证据,从而弱化了人对人的信任。在该系统的运作模式中,大数据与人工智能的信任能力是对信用积分的一种依赖,只不过它将信任判定的来源置于一个更广的数据池中。但这种风控模式仍需面对一个质疑:技术对人的“信任”判断是否真的具有正当性?首先,因为信贷不仅是一个风险规避问题,而且也包含着“急人所需”的道德主张,所以,这种“正当性”就是要对技术或算法做出某些道德性要求。再者,“人-机信任”要具有某种道德性含义,那么这种信任不能以企业对人工智能物的信任为依据,而应以公众对于人工智能物的信任为依据。公众对人工物的信任中显然已经蕴含程序透明度、合法性等问题[9]。

技术的道德要求,也是社会对使用技术的企业的道德要求。比如银行仅仅依靠人工智能的数据分析作为风控的手段,是否会不公平地对待被判断者?再者,技术判断的数据依据究竟如何厘定?在数字经济发展中,银行可能会将博士生不能按时毕业视为失信风险行为[10],其中包含着大量技术与正义的关系问题。将哪些事件视为失信标准是一个有关正义的问题。技术对人的“不信任”应有符合正当性原则的依据,而不能直接被企业依据风险最小化原则处理。“风险最小化”有可能导致对用户的不公平对待。

因此,应用技术风控的同时,也必须警惕风控技术本身导致的正义性风险。曹玉涛认为:“社会财富和权力的分配要服从技术资本的‘绝对命令’,且长期漠视技术风险和代价与承担者利益的存在,不断消弭技术在解决贫困、疾病、饥饿和环境问题上的责任和义务,从而将处于技术垄断之外的人推向社会生活边缘。”[11]在技术脱域性时代,技术和专业知识终结了人与人之间的直接信任,这已经成为社会关系的新常态。同时技术发展也在为社会造就新的不公正风险。在此,风控这种以不信任与怀疑为前提的技术显得尤为特殊。在人工智能营造的深度脱域性氛围中,如何保证人在技术规控中得到公正对待,成为一个新的议题。

三、智能风控中的数据事实与正义价值

智能风控的出场本身便是技术与伦理的互涉,现代人的基本生存境况正如艾吕尔所看到的那样:“始终存在于技术的缠绕之中,对技术的使用使得人就处在技术之中”[12]。智能风控以智能化算法渗入诸多伦理生活领域中,并以事实性数据的计算对现实生存的个体做出智能判定。这种智能判定形式,不仅使得“人处在技术之中”,同样也以主体的形象出场,在有限的社会角色呈现中,对人作出信任与否的“责任”判定。由此,智能判定的主体能动性程度及其深度社会参与所带来的正义问题便跃然而出,这种看似理性的算法背后,其实潜藏着诸如算法黑箱、算法歧视等一系列问题。如网约车平台根据手机价值去制造收费偏差等“大数据杀熟”现象。而且,除人为设定算法的正义风险之外,在运用算法的过程中,算法本身也会成为一种决策黑箱,其间可能形成的自主偏好,也会导致决策的非公正性,如算法会根据操作习惯,将偏好固定化为程序。

社会公正的基本要求是对人的尊重,而在当今社会的算法化治理倾向中,“技术权力的共性在于依托大规模的集成数字化平台,靠全覆盖的摄像头、传感器、分析中心和智能终端汇总社会个体和群体的生活形态和生活轨迹,将分散的价值与意识集中成公意,从而强制执行。”[13]因而,依托算法的广泛应用在某种意义上必然伴随着马加利特所谓的“羞辱”的风险[14],人在技术中被视为无差别的“可计算性”存在,“‘人格数据化’使人成为技术网络的附属品,成为算法运行中的一个元素和节点。以技术理性作为社会价值的导向,压抑了每个社会个体与生俱来的人格和尊严的诉求。”[13]这也是人工智能时代必须要面对的一种正义问题。在智能时代的技术化风险控制的诸种尝试中,正义仍是现代技术设定必须考量的伦理界限。在智能风控中存在以下几种非正义风险:

第一,依据事件的数据判断“应得”。库珀特在《作为恰当性的正义》一书中曾指出,正义不仅意味着社会公共善的平等分配,而且应体现在每一个人得到恰当的对待[15]。但在风控思维中,每一个人都被视为信用上可能犯错的人,那么,如何保证算法判断符合“应得”本身就是一个关乎正义的问题。技术化操作将个体人格以及人的应得都置于一种量化中,将判别对象视为一个各种信息痕迹综合构成的拼合体,根据这些信息做出的风险判断很可能并不符合其应得。而且算法本身的不透明性也使得算法

对待受众的方式遭受质疑。因而在这种大数据风险预判中,风险判断效率可能会得到增强,但判断的正义性则是存疑的。

第二,事实与价值的难以对等问题。风险评估预设了一种怀疑性思维,这就导致预判以一种怀疑思维处理事实,所有判断的依据是事实条件的集合,然而风险评估却要通过对$\{a_1,a_2,\cdots,a_n\}$等事实集合的分析去推论可能性结果。那么,对于失信判定的"原因",仍旧是按照概率计算的。但这种模态逻辑关系,对于事实与价值的因果关系的判断会造成一定的困难。而且,不同于通过水泥填充量、钢筋直径等诸多事实对桥梁风险性的判定,智能风控依据个人哪些事实去推定一种具有正义价值的判定,需要进一步论证。通过图2两种正义模型的对比可以更好地说明这个问题:

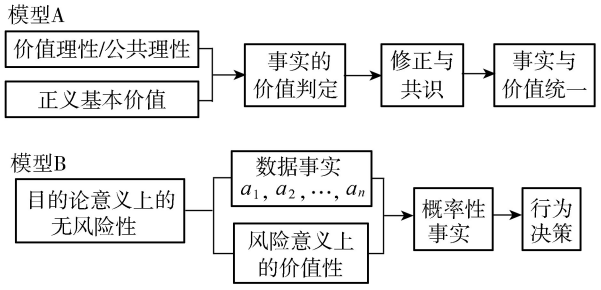

图2　两种正义模型对比

模型A是罗尔斯和哈贝马斯意义上的正义模型,价值与事实的统一性体现在对事实的公共理性评价与约定共识的双重并进上。而就模型B而言,则体现了风险世界的一种预估性规划,并且此种模型下的最终决策还试图满足人类的一般正义要求或正义直觉。但其起点在于一种否定性的"无风险",这种目的性会导致一种后果论。基于这种后果主义,算法程序排除了价值生成过程,很难实现事实与正义价值的契合。

第三,正义的主理者问题。正义的主理者问题关涉的是谁应为正义问题负责。一种常识观点是企业技术行为可以不为社会正义负责。一项技术设计的初衷就是工具性使用,它适用于特定的人,可以不涉及权利平等的责任问题,权利与平等应是政府治理行为的关注点。比如一项新药技术的产生,可以不为可能带来的贫困群体无法使用该药的结果负责。该观点是一种技术中立主义的立场,认为技术本身与其可能的使用状况是分离的。也就是说,技术的本质在于其功能性,没有义务去承担其后的公平、正义等问题。另外,就算技术本身应为其所带来的正义问题负责,那么,企业或技术的开发者与智能人工物本身,谁应该切实承担正义判定的主体地位也值得商榷,因为人工智能物也在渐渐产生某些不依赖于人的智能偏好等,因此它们也逐渐被视为一种承担部分责任的主体。

综上,智能风控的应用在逻辑上存在正义性的风险。这种风险源于风险防控的怀疑本性,在算法设定上以对人的怀疑为起点,势必导致智能风控在算法与判断上对风险的"超事实"衡量,这就会造成数据事实与其所附带的正义价值的不一致性。而智能风控要实现其可信任性,必须在正义性上保证其价值。正如奥妮尔所说:"信任危机不能用盲目信任来克服。"[16]那么,人机之间的信任问题应通过对算法正义的信任来克服,而非直接将技术效益作为信任的合理依据。

四、智能风控正义的主体能动性要求

算法风评以对人的怀疑为基本设定,只有"怀疑"程序被否定之后,才能产生"信任"判定。在风险的智能判定过程中,作为主体的机器如何通过既定数据使怀疑合乎伦理,并对人类事件做出能动的责任判断?这是当前人工智能风控的一项重大难题,也是当前人工智能主体性研究的一个重要问题。麻省理工学院航天控制实验室(Aerospace Controls Laboratory)近期开发了一种新的深度学习算法以提升智能物的"合理怀疑"能力。"合理怀疑"是主体特有的能力,它也是一种智能物智能化程度的表征方式。本研究更关注如何令人工智能不加歧视地怀疑,即不以身份、出身、地域、种族等偏见对人怀疑,也就是合乎伦理的"合理怀疑"。这种"合理怀疑"对于当前的人工智能技术而言是很难独立完成的,现有的解决思路是通过设计者、从业者与智能体共同配合来保证决策的合理与公正。那么,这就带来了一个问题,在这一过程中设计者、企业还有智能体,谁是实质意义上的道德主体呢?谁又在该判断中具备完全的能动性呢?

主体能动性问题在伦理学上意指:"具有意向性能力的人相应地具有道德判断和选择能力"[17]。人工智能领域的研究者通常认为:智能物只是算法,真正的能动主体是算法的设定者,也即是企业或程

序开发者。但随着智能物的智能程度及其社会参与程度的提高,随着现今人工智能的发展,人们开始期待人类与智能物的双重主体性,正如 Hanson 所指出,从智能体在深度社会参与中逐渐获得社会角色的角度来看,“主体”与“能动性”都显示出高度的人机协作与重叠,即以“复合主体”(composite agent)与“复合能动性”(composite agency)[18]的形式呈现,在道德责任上“以复合能动性的形式导向人类和非人类的协同”[18]。也就是说,主体判断由传统的专家、开发者等责任承担形式,逐渐转向复合双主体责任形式。

但实际上这种复合主体的设想在银行智能风控行为中很难实现。首先,虽然智能物在技术上已经可以实现智能算法的某些自主性,但是较之“事实能动性”,实现智能物的“道德能动性”更具难度。其次,因为银行风控本质上是逐利行为,为规避风险银行不愿意过多考虑智能物的道德责任判断算法。但在现代智能领域责任创新范式的要求中,已经可以明显地看到智能判定由最初的数据处理转向了伦理社会的人机关系。正如 Brundage 提出的人工智能“责任式创新”(responsible innovation)模式,即是关注到人工智能技术与社会生活的深度融合,强调跳出以往仅从技术本身审思人工智能问题的思维方式,转而在责任式创新模式中探索人工智能技术参与社会工作的有效性[19]。基于此,就智能风控技术而言,智能计算存在大数据智能风控与技术化信任的脱域性,并且由此产生社会正义审思由传统正义模式向机器如何信任人的新型正义形式转变的问题,这是一个我们在人机伦理关系的交往中必须慎重对待的问题。

从主体能动性问题引出如何恰当地对待人的正义问题,本质上是智能风险评估如何体现出道德性的问题。以“技术中立主义”的观点来看,技术应与其后果分割开来,这种观点不仅分隔了技术与其创造者的责任问题,也没能看到智能技术本身在负责任能力上的特殊性。就目前银行智能风控系统来看,智能算法对事实性数据的加工处理,总是会导向对借贷个体的现实利益的判定与影响,以致于判断对象的许多生活痕迹被不加深究地作为失信的可能性依据。人工智能在这种智能风控中处于“工具式地位”(instrumental position)。而负责任的智能风控创新,应从根本上考量人工智能技术在风控行业发展的“能动性地位”(agency position),以一种它能够独立作为能动判定主体的乐观态度,期待其在该行业的更好发展,并且能够兼顾判断准确性与道德性。这种期待要求智能风控技术不仅介入真实情境伦理判定的因果链条之中,而且以复合式主体的形式,从判定借贷方是否具备还款能力的角度,反映出智能机器对人的正义问题。

风控领域对“恰当性正义”的需要势必是以减弱工具式的参与方式为前提的,虽然提升智能物参与的道德主体性仍是一个难题,但已成为一个趋势,欧盟在“地平线 2020”计划中提出了责任式创新范式,将发展工具性的数据分析设备,转向发展包含更多伦理道德考量与社会整体期待的人工智能。正如 Latour 所说:“道德法则在我们的心中,但它也在我们的仪器中。在到达超我(super-ego)的传统之前,为了阐述我们行动的正确性、信任度和延续性,我们可能要增加技术的隐我(under-ego)。”[20]所谓“隐我”是指技术设计整体背景下某种伦理植入的“自我”。智能风控作为切实关涉人的利益的领域,应当率先考虑这一趋势,以一种人机复合主体性的形式深度参与实际伦理生活,从而一定程度上对社会正义等现实伦理实践决策产生影响。

五、结　语

人工智能风险防控以对人的信用质疑为出发点,如果说风险防控的技术规划是现代性的典型特征,那么,人工智能风控将人的信任能力移交给机器或数据则体现了社会的进一步脱域化。但人工智能风控本身在技术层面上存在非正义风险,它在数据事实和正义价值之间还不能建构起某种伦理关系。因此,对人工智能风控的伦理期待与对人工智能技术本身的伦理期待是一致的,在“复合主体”和“复合能动性”的要求中,智能风控技术需要建构自身的“可信任”条件,而这种“可信任”恰恰体现在其“怀疑”能力的合理性上。

参考文献:

[1] 安东尼·吉登斯. 现代性与自我认同[M]. 夏璐,译. 北京:中国人民大学出版社,2016.

[2] 张康之. 论风险社会生成中的社会加速化[J]. 社会科学研究,2020,42(4):22-30.

[3] 赵万里. 科学技术与社会风险[J]. 科学技术与辩证法,

1998,15(3):50-55.
[4] 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[J].华东政法大学学报,2019,22(6):17-30.
[5] 杜海涛."不亿不信":信息"缺场"的信任选择——论孔子的信任思想及其现实意义[J].东南大学学报(哲学社会科学版),2019,21(1):26-31.
[6] European Commission. Ethics guidelines for trustworthy AI[EB/OL]. [2018-12-18]. https://ec.europa.eu/futurium/en/european-ai-alliance/draft-ethics-guidelines-trustworthy-ai.html.
[7] 闫宏秀.可信任:人工智能伦理未来图景的一种有效描绘[J].社会科学文摘,2019,17(10):8-10.
[8] STIRRAT M, PERRETT D I. Valid facial cues to cooperation and trust male: facial width and trustworthiness[J]. Psychological Science, 2010,21(3): 349-354.
[9] JOBIN A,LENCA M,VAYENA E. The global landscape of AI ethics guidelines[J]. Nature Machine Intelligence, 2019,1(9): 389-399.
[10] 江小涓.数字经济,解构与链接[EB/OL].[2020-11-21]. https://www.thepaper.cn/newsDetail_forward_10084133.
[11] 曹玉涛.技术正义:技术时代的社会正义[N].中国社会科学报,2012-12-19(B-02).
[12] JACQUES E. The technological society[M]. New York: Alfred A. Knopf, Inc. and Random House, Inc, 1964.
[13] 张爱军,李圆.人工智能时代的算法权力:逻辑、风险及规制[J].河海大学学报(哲学社会科学版),2019,21(6):18-24.
[14] 阿维沙伊·马加利特.体面社会[M].黄胜强,许铭原,译.北京:中国社会科学出版社,2015.
[15] 杰弗雷·库珀特.作为恰当性的正义[M].马新晶,译.南昌:江西人民出版社,2020.
[16] 昂诺娜·奥妮尔.信任的力量[M].闫欣,译.重庆:重庆出版社,2017.
[17] 张卫.人工物的道德能动性[N].中国社会科学报,2018-03-13(002).
[18] HANSON F A. Which came first, the doer or the deed?[M]//KROES P, PAUL V P. The moral status of technical artefacts. Springer Science, Business Media Dordrecht, 2014:55-74.
[19] BRUNDAGE M. Artificial intelligence and responsible innovation[M]//MLLER V C. Fundamental issues of artificial intelligence. Springer International Publishing Switzerland, 2016:543-554.
[20] LATOUR B. Morality and technology: the end of the means[J]. Theory, Culture & Society, 2002,19(5/6): 247-260.

(收稿日期:2021-06-27 编辑:高虹)

·简讯·

河海大学举行严恺院士诞辰110周年纪念活动

2022年8月10日,在严恺院士诞辰110周年之际,河海大学在学校严恺院士铜像前举行庄严的敬献花篮仪式,深切缅怀严恺院士,表达学校师生的崇高敬仰和怀念之情,激励和引导全校师生赓续优良传统、践行爱国奉献、秉承校训精神、矢志团结奋斗,努力为实现中华民族伟大复兴的中国梦接续奋斗、永远奋斗。学校党委书记唐洪武、校长徐辉参加仪式并整理花篮绶带,副校长郑金海主持仪式,师生代表参加仪式。

严恺(1912—2006),福建闽侯人,教授,博士生导师,水利和海岸工程专家,中国科学院院士,中国工程院院士,河海大学名誉校长,南京水利科学研究院名誉院长,中国水利学会名誉理事长,中国海洋学会名誉理事长。1933年毕业于交通大学唐山工学院。1935年赴荷兰德尔夫特科技大学深造,1938年获荷兰土木工程师学位。1938年回国后,先后任中央大学、上海交通大学水利系教授和黄河水利委员会简任技正兼设计组主任,宁夏工程总队总队长、研究室主任以及钱塘江工程局技术顾问等职。1952年负责筹建新中国第一所水利高等学校——华东水利学院,任副院长、院长,1955年当选为中国科学院院士(学部委员),1995年当选为中国工程院院士。1956年被任命为南京水利实验处处长,1957—1983年兼任南京水利科学研究所所长。主持过多项国家重点科技项目,做出了重要贡献。曾任联合国教科文组织国际水文计划政府间理事会副主席,国际大坝会议中国委员会主席,发展中国家海岸和港口国际会议顾问委员会顾问等职。曾先后当选第三届全国人民代表大会代表和中国共产党第十次、第十一次全国代表大会代表。

(信息来源:河海大学官方网站 http://www.hhu.edu.cn)