

Doi:10.11835/j.issn.1008-5831.fx.2023.03.004

欢迎按以下格式引用:刘嘉铮.人工智能产品算法设计者的犯罪过失判断——以危惧感说的核心观点为立场[J].重庆大学学报(社会科学版),2024(3):228-241. Doi:10.11835/j.issn.1008-5831.fx.2023.03.004.



Citation Format: LIU Jiazheng. The criminal negligence of the designer of artificial intelligence products: Taking a stand on the central perspective of the sense of fearing theory[J]. Journal of Chongqing University (Social Science Edition), 2024(3): 228-241. Doi:10.11835/j.issn.1008-5831.fx.2023.03.004.

人工智能产品算法设计者的 犯罪过失判断 ——以危惧感说的核心观点为立场

刘嘉铮

(东南大学法学院,江苏南京 211189)

摘要:判断人工智能产品算法设计者的过失时,修正的旧过失论对导致危害结果发生的因果历程设置了具体预见可能性标准,这一标准与算法的黑箱属性以相关性而不是因果关系为基础的决策过程存在冲突。同时,此立场只重视结果忽视行为的逻辑会打击算法设计者的积极性,阻碍算法进步。新过失论虽然以结果避免义务作为犯罪过失的核心,但是其对结果预见可能性的标准缺乏具体设计,在判断预见可能性时往往束手无策。因此,两种立场都不是判断人工智能产品算法设计者犯罪过失的合理方案。相比之下,虽然危惧感说(超新过失论)认为结果预见可能性只需达到危惧感的观点受到了主流观点的批评,但这种批评值得商榷:其一,只看到了这种立场对结果预见可能性的低程度要求,却没有看到这种要求背后的核心观点对于判断犯罪过失的合理性;其二,将危惧感说提出者本人对个别案件的过失判断等同于危惧感说的全部,略显片面。与修正的旧过失论和新过失论相比,危惧感说的核心观点是:结果预见可能性与结果避免义务存在相互关联性,这是判断人工智能产品算法设计者犯罪过失的合理方案。以危惧感说的核心观点为思路,犯罪过失包括客观的结果预见可能性、客观的结果预见义务和客观的结果避免义务。人工智能产品算法设计者客观的结果预见可能性的标准是:一旦遭遇包含异常因素的特殊情况,算法有可能做出不利决策,进而引发消极后果。算法设计者客观的结果预见义务的内容是:其一,应当预见到其设计的算法不仅会被用于没有异常因素出现的正常情况,而且可能被用于伴随异常因素出现的特殊情况;其二,一旦其设计的算法面临特殊情况,该系统可能会做出不利决策。算法设计者客观的结果避免义务的内容是:应当避免在设计算法时植入为社会公众普遍反对或不赞同的价值理念;在设计时检验“投喂”给算法系统的数据质量,最大程度防止缺陷数据进入算法机器学习训练的“垃圾

基金项目:国家社会科学基金重大项目“智能技术赋能政法领域全面深化改革研究”(22ZDA074)

作者简介:刘嘉铮,东南大学法学院博士研究生,Email:2533982836@qq.com。

进”风险;及时告知产品生产者算法可能面对的异常情况。

关键词:算法;结果预见可能性;危惧感说;结果预见义务;结果避免义务

中图分类号:D914;TP18 **文献标志码:**A **文章编号:**1008-5831(2024)03-0228-14

一、人工智能产品致害的归责关键

“人工智能时代的科技革命已经从理论实验转向技术应用”^[1]。人们不断将人工智能产品应用于社会生活的多个领域,比如扫地机器人、医院看护机器人、无人机以及自动驾驶汽车等^[2]。人工智能产品在蓬勃发展的同时,也会带来一定风险,比如搭载了自动驾驶系统的机动车在道路上行进时可能引发交通事故。这些风险进而会产生刑法中的归责难题,包括谁应当对人工智能产品引发的危害结果承担刑事责任,以及刑事责任主体在什么情况下才应当承担刑事责任。判断刑事责任主体的关键是谁的决策引发了危害结果,或者谁的决策为结果的发生提供了条件。算法的逐渐成熟使决策主体正在发生改变。

(一) 智能产品算法对人力决策的取代

长期以来,人类选择是社会、经济、政治制度必不可少的部分,是人们私人生活和公共生活不可或缺的基础部分^[3]。人力决策一直是决策权力的核心,但人工智能产品的推广正在逐渐改变这种权力配置。当前的人工智能产品等于大数据加机器学习^[4],是一种搭载了智能系统的、能够凭借该系统进行自动化运转的机器产品,其核心是算法,比如自驾车是指机动车搭配人工智慧系统,使之得以自动操作车辆和运行^[5]¹⁹⁰。德国学者希尔根多夫教授指出:“自动化驾驶可以简述为自主系统辅助下的驾驶。所谓自主系统是指无需人类持续输入信号即可应对所面对之情境的系统。”^[6]³⁹⁶在自动驾驶等领域中,算法发挥了举足轻重的作用,它可以大幅提高决策效率,为消费者或用户提供更精准的服务^[7]。

算法逐渐脱离了纯粹的工具性角色^[8],有了自主性和认知特征,甚至具备了自我学习能力^[9],使其在服务于人类的同时从一种决策辅助工具进化成为一种新兴智能技术^[10]。作为机器学习新阶段的深度学习算法已经能够进行自主性再编程,以至于算法开发者也无法理解算法决策背后的逻辑^[11]。在搭载了算法系统的人工智能产品中,算法已经渐渐取代了人力决策,在未来可能成为实际上的决策主体。在明确人工智能产品装载的智能系统属于算法的情况下,人工智能产品引发危害结果时,过失判断的核心应当围绕算法决策进行。

(二) 智能产品致害的归责核心:算法决策的不确定性

人的意志是判断行为是否违反刑法规范的关键,我国台湾地区前高等法院大法官许玉秀教授指出:“人和他的行为是分不开的,行为所表现的是人的特质,人的态度、人对世界的诠释,而这才是作为这个世界的主体的我们所关心的、所在乎的,如果一个行为是一个梦游的人所做的,一个损害是被龙卷风刮起来的人所造成的,我们不会在意,也不能在意,因为没有人能对抗自己的梦游、对抗龙卷风。”^[12]因此不在自由意志支配之下的人的举止不属于刑法中的行为,而“人的自由意志支配之下的举止”这个行为定义可以被置换为另一种概念:人力决策举止。如果一种举止不是由人力决策而是由于梦游等外部原因导致的,该举止应属于外部强制。实际上,目前我国加速进入的“风险社会”中的“风险”含义并不是技术风险、事故风险、金融风险等具体类型的风险,而是社会发展不断

复杂化背景下人力决策的有限性带来的不确定性^[13]。

人工智能产品引发危害结果时,情况会发生根本性变化。人工智能产品依靠自身搭配的算法根据外部信息进行决策,进而对机器发出指令,这个过程并不包含自然人,因而人工智能产品致害的整个过程都没有人力决策的参与,危害结果发生过程实现了从人力决策到算法决策的角色转换。因此,在人工智能产品引发危害结果时,犯罪过失判断的核心应当是人工智能装载的算法决策的不确定性,而不再是人力决策的不确定性。

在人工智能产品致害过程已经完成从人力决策到算法决策的转换的情况下,对人工智能致害的归责判断应当围绕与算法决策有关的行为展开,谁的行为制造了支配算法决策引发危害结果的不被容许的风险,谁就应当对该危害结果承担刑事责任。“过失犯刑事责任成立的关键时刻,应该是在危险形成的时候,而不是等到危害结果出现”^[14]。犯罪事实的发生始终是先行为制造了不被容许的风险,然后现实化为危害结果,故过失犯罪危险形成的时刻应当是行为实施之时。智能产品算法的设计行为对于算法决策最终引发危害结果的整个过程起到了重要的作用,判断智能产品算法设计者在设计时是否存在过失是对人工智能产品引发的危害结果进行归责时必须解决的问题。

二、我国犯罪过失的立场之争

过失的立场是研究一切过失犯罪的理论前提,在讨论人工智能产品生产者的过失责任之前,刑法应当首先明确过失论的立场,进而进行过失犯罪构造的设计。目前为止,刑法理论关于过失的立场主要有修正的旧过失论和新过失论两种,结果无价值论者普遍支持修正的旧过失论,行为无价值论者普遍支持新过失论。

(一) 修正的旧过失论的判断立场

修正的旧过失论以旧过失论为基础,认为过失的核心是结果预见可能性。从这种立场出发,违法是一种纯粹客观的事实,与故意、过失无关^[15]¹⁶⁰,过失只是一种罪责要素,在构成要件合致性和违法性层面没有判断过失的余地,结果预见可能性的有无与程度是过失的唯一条件,只要行为因果性引起了危害结果并且行为人对该结果存在预见可能性,行为人就构成过失犯罪^[16]。为了防止以预见可能性作为判断过失的唯一标准过于扩张处罚范围,修正的旧过失论和旧过失论一样,都主张以判断过失犯罪成立的结果预见可能性即使不能达到分毫不差的程度,也必须达到较为具体的程度:对危害结果及引起结果发生的因果历程的基本部分存在预见可能^[17]。

(二) 新过失论的判断立场

新过失论认为过失的核心不在于结果不法,而在于行为不法。在行为人实施的行为与法律要求的举止相协调,即履行了交往中必要的注意义务时,该行为就是合法的^[18]。从新过失论的立场出发,人的行为与主观方面并非完全无关^[19]。故意犯罪和过失犯罪在构成要件该当性层面和违法性层面存在明显区别,故意行为的违法性明显高于过失行为的违法性,因此故意、过失首先是主观的违法要素,以及(作为违法类型的)主观的构成要件要素。“只要不承认作为构成要件要素的过失观念,就不可能论及过失犯的构成要件符合性。这样,过失也与故意并立,首先应被视为构成要件要素,可以称其为构成要件性过失”^[20]。

因为过失首先被定位为不法构成要件要素,所以过失判断的关键是行为人未尽一般社会交往中必要之注意^[21],一般社会交往中必要之注意是“以社会上实际之观点,追究在行为人所处之具体

情况下,具有思虑且谨慎之人应该采取如何之行为”^[22]。故,新过失论的过失构造是:违反客观的注意义务的作为或不作为^[23]。由于过失的这种双重属性,刑法必须分两个阶段对其进行审查。首先,刑法需要在构成要件层面审查过失,这涉及客观的结果预见可能性与客观的注意义务。其次,刑法需要在罪责阶段审查过失,这涉及行为人本人能否履行客观的注意义务并主观地预见危害结果的发生^[24]。由此可见,不同于修正的旧过失论,新过失论认为过失的本质并不是预见可能性,而在于行为人违反客观的注意义务中的结果避免义务,即偏离为刑法规范所期待的基准行为,“倘能遵守社会生活上为了回避结果所必要的一般注意即为适法”^[25]。

(三) 我国犯罪过失的立场争议

我国刑法对于犯罪过失的主要立场存在修正的旧过失论和新过失论的争论,支持修正的旧过失论的观点认为不法要件是中立、无色的,在进行不法判断时,只应考虑法益侵害或者法益侵害危险,因此故意、过失属于责任要素^{[26]328},只要行为人的行为对危害结果的发生具有紧迫危险,就是过失实行行为^{[26]378-382}。张明楷教授指出:“应当承认行为人对结果的预见可能性是责任要素。”^{[26]382}支持新过失论的观点认为以结果避免义务作为客观的行为基准来设定客观的注意义务的做法,能够限制过失犯的泛滥^{[15]163}。陈兴良教授指出:“只要采取了对一般人而言具有合理性的结果回避义务即基准行为,即便具有预见可能性,由此所出现的结果属于被允许的危险,并不具有违法性,由此限制了过失的处罚范围。”^[27]

三、主要过失立场对算法设计者犯罪过失判断的弊端

(一) 修正的旧过失论的缺陷

1. 具体的结果预见可能性要求与算法黑箱属性的冲突

修正的旧过失论对具体预见可能性的要求与算法的黑箱属性不相容,在判断人工智能机器算法设计者的刑事责任时,可能会导致处罚目的落空。

算法进步迅速,其计算方式已经由原本的监督学习改为非监督学习,无需人类知识和经验的指引^{[5]192}。监督学习和非监督学习都属于机器学习算法的一种,监督学习属于传统机器学习,非监督学习属于作为全新机器学习种类的深度学习^{[28]119}。人工智能产品依赖的机器学习算法能够在没有人类参与的情况下在后台自动运行^[29],其内部结构属于一种深度神经网络,深度神经网络从经验中学习,由成千上万个人造神经元组成,这些神经元对信息的判断不能简化为固定的顺序,单个或单独一组的神经元也不能单独决定对于决策来说哪个信息重要或哪个信息不重要,而是奉行一种交互式联系的理念,各个人造神经元进行一种分散式的通力合作,共同对决策所依赖的外部信息进行决定,进而做出决策。因此,机器学习算法的决策过程与人类的直觉相似^[30]。算法在此应用的,并不是单向的形式逻辑,而是大量运用或然率、臆测与自我纠正决策过程,使相关决策结果以相对高的或然率提供答案^{[31]218}。

由于在机器学习算法深度神经网络中,各个神经元共同判断信息并进行“直觉式”决策,如此决策所产生的结果,并没有一种百分之百的必然性,何种信息在算法的神经系统中受到何种评价,算法设计者无从得知^{[31]219}。绝大部分机器学习算法的输出结果,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,其动态的变化也难以预测。更重要的是,对于机器学习而言,输入数据的变化和累加,使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,它是一个无法实现透明的“黑箱”^[32]。这导致算法的缺陷性决策并不是出于设计者能够控制的“硬编码”,而可能出自于机器学习过程中的各种突发属性^[33]。算法设计者无法控制机器学习,从而也就无法预见算法究竟会在外部环境的刺激下做出什么样的决策。

由此可见,机器学习算法的复杂性、不透明性会导致因果追溯与归责更加复杂^[34],机器算法决策过程的黑箱属性使算法以相关性而不是因果关系为运转的基础,旨在发现不同对象之间的微妙关联而非理解其原因,导致算法的决策“行为”缺乏可理解性与可预见性^[35],进而使设计者不太可能对智能产品装载的算法决策所引发的危害结果历程存在具体预见可能性。因此,人工智能产品致害缺乏事前的预见可能性,是其最重要的特征^[36]。

然而,修正的旧过失论对具体预见可能性的要求是:行为人对危害结果的预见不需要达到严丝合缝的完美对应,而是对危害结果的基本部分存在预见。“危害结果的基本部分”不能孤立地考察结果本身,而应考察危害结果发生的整个因果关系历程的基本部分。过失犯罪危害结果涉及一个漫长的因果链,行为人实施了侵害行为,才有办法创造新的损害因果历程,损害才可能出现,通过主体→行为→损害之间串连,才有办法引发结果^[37],也就是说,预见可能性的对象并不是结果,而是不法构成要件的全部事实^[38]。因此,危害结果的基本部分指的是整个危害结果发生的因果关系历程的基本部分,具体是指去掉实际发生的因果历程中在刑法上并无重要意义的具体个别细节(比如时间、地点、行为人和被害人的姓名等)以后,剩余能够清楚说明行为人实施行为,进而引发结果的全过程核心部分。即:具体预见可能性要求行为人可能预见的不法构成要件事实必须能够在构成要件要素层面上进行抽象把握,即使行为人对于实际发生的全部事实没有预见,但只要对构成要件的基本事实存在预见可能就已足够^[39]²⁵⁰⁻²⁵¹。

具体到智能产品致害的因果历程中,具体的预见可能性要求是:虽然不要求算法设计者具体预见到算法的哪一个决策会导致哪一个时间、哪一个地点的哪个(群)人受到什么程度的损害,但是要求设计者预见在特定情况下算法的某一个决策的内容,以及这个决策会给人造成什么损害。但如前所述,算法决策的不透明性会导致法律判断谁在控制算法系统的风险或者识别哪一种代码、行为、输入值或数据最终引发了危害结果的代价非常高昂,甚至无法做到,这种因素会导致行为与危害结果之间的因果关系历程很难被判断^[40]。因此,“利用传统的人类认知论中的因果关系概念对依赖大数据相关性运作的算法决策及其侵害进行归责仿佛缘木求鱼”^[34]。

2. 对行为本身的忽视

以预见可能性作为过失犯罪核心标准的观点会落入无视行为只看结果的错误逻辑,导致算法设计者在研发时瞻前顾后,阻碍技术进步。

修正的旧过失论以是否存在预见可能性为核心标准,这样的主张将会导致刑法在判断过失犯罪时只重视结果而忽视导致结果发生的行为,只要行为在客观上引发了危害结果并且行为人主观上存在预见可能性,行为人就会被定罪。算法设计行为一方面能够将自然人从劳动中解放出来,促进社会进步,另一方面也可能会带来风险。如,人工智能机器可以调控室内温度,但可能把房间变成冰窖或火炉;可以自动开门,但可能把人们锁闭在屋内;可以自动驾驶,但可能引发事故;可以治病救人,但可能误诊杀人^[41]。当前社会发展日益复杂,风险日益增加,而风险的增加一般而言能够预见,如果利用过失犯原理追究对于损害具有预见可能性的行为人,不仅会因为风险的增加而扩张过失刑事责任的范围,更会导致人们为了避免承担过失刑事责任而被迫放弃能给生活带来便利,但

同时带来风险的新类型活动^[42]。

如果采取修正的旧过失论立场,认为只要算法设计者对于产品会给他人造成危害存在预见可能性,客观上产品又确实引发了危害时即构成过失犯罪,不仅与“现代多元化的风险社会中,人类必须放胆行事,不能老是在事前依照既定的规范或固定的自然概念,来确知他的行为是否正确”^[43]的正常行事逻辑相违背,也会导致刑法忽视设计者的行为本身,只根据危害结果进行定罪,落入结果责任的错误逻辑,打击科研人员开发人工智能产品算法的积极性。

(二) 新过失论的合理性与不足

新过失论以结果避免义务作为过失本质的观点能够在很大程度上减轻人工智能机器算法系统的开发者和产品生产者的心理负担,不采取结果避免措施才可能构成犯罪的观点使算法设计者能够大胆进行系统设计而无需在行为当时瞻前顾后。然而,新过失论在重视结果避免义务的同时,并没有对结果预见可能性的标准进行具体设计,虽然新过失论认为结果预见内容应更为抽象^[44],但是与修正的旧过失论对具体预见可能性的设计相比,新过失论并没有提出明确缓和理解预见可能性的标准。由于缺乏标准,新过失论在判断预见可能性时往往束手无策,最终被迫借助旧过失论的预见可能性要求,认为“新过失论在理论中同样强调行为人的具体预见可能性”^[45],或者并不直接支持具体的预见可能性,而是从结果避免可能性入手,提出行为人对不法构成要件的实现欠缺预见可能性就等于欠缺避免不法事实实现的可能性,进而认为避免可能性只是预见可能性的同义词^[46],从而将旧过失论对预见可能性的具体要求悄无声息地纳入新过失论的体系之中。支持这种概念置换的新过失论者进一步认为,修正的旧过失论主张的具体的结果预见可能性与新过失论主张的结果避免可能性的内容是一致的^[47]。

如果具体的结果预见可能性仍然在新过失论中被要求,那么新过失论同样会落入前述具体的预见可能性与算法的黑箱属性本质上不相容的难题。由于设计者无法具有具体预见可能性,最终无需承担结果避免义务而不成立犯罪,如此一来会产生非常荒诞的结论:人工智能机器的错误决策致人伤亡或造成他人损失,但最终却由受害人自己对自己的损害负责。因此,虽然新过失论比修正的旧过失论更加合理,但其仍然不是最佳方案。

四、危惧感说(超新过失论)判断算法设计者犯罪过失的合理性证成

在修正的旧过失论和新过失论以外,还存在一种对预见可能性作更加抽象理解的过失立场,被称为危惧感说(超新过失论)。与前述两种主要立场不同,危惧感说在提出以后受到了主流观点的猛烈批评,认为这种立场立足于行为无价值一元论,直接将结果不法逐出构成要件要素,不需要结果的预见可能性,只要没有采取措施消除不知道会发生什么的危惧感就成立过失^{[39]242}。我国刑法中的大多数观点也不赞同危惧感说,认为其对预见可能性的要求过低^{[26]371}。然而,主流观点对危惧感说的批判可能值得商榷,在判断人工智能产品算法设计者的犯罪过失时,危惧感说可能是合理的立场,这种合理性是由危惧感说的核心观点决定的。

(一) 危惧感说的抽象预见可能性要求

危惧感说是日本学者藤木英雄教授提出的立场,在继承了新过失论观点的基础上,对预见可能性进行进一步缓和理解,与新过失论并未提出明确的缓和理解预见可能性的标准不同,危惧感说提出了明确的预见可能性标准,认为危害结果的发生无需具体预见可能性,行为人只要对结果存在危

惧感和不安感就已足够。

藤木英雄教授就森永毒奶粉案^①的判决指出：“在科学技术无限发展和扩大的同时，未知的危险对我们人类的威胁越来越大，就这一点来说，要想利用科学技术，就必须负有社会性的责任。如果这种见解是正当的，那么，作为结果来说，即使是不能具体地预想到会有什么样的危险发生，但由于存在有不知道会对人的生命和健康发生何种破坏的有害结果的危惧感，在这种情况下，为能足以消除这种危惧感，防止结果发生起见，令其有合理的负担，应该说是理所当然了。”^[48]危惧感说对预见可能性的这种危惧感、不安感的理解可以被概括为一种抽象的结果预见可能性标准^[49]。

(二) 危惧感说预见可能性标准背后的核心观点

如前所述，主流观点对危惧感说进行了批评，然而主流观点对危惧感说“对预见可能性要求过低”的批判可能值得商榷。危惧感说确实在表面上对预见可能性只提出了危惧感、不安感的要求，但是这样的要求并不是凭空创造出来的，而是源自危惧感说的核心观点，危惧感说的核心观点并不是“只要行为人对危害结果的发生存在危惧感就够了”，而是“预见可能性的结果回避义务关联性”^[50]。在这种立场看来，预见可能性不能简单地以“有”或“没有”这种思路进行判断，而应当在与行为人采取的结果避免措施的关系上相对地被决定，低程度的结果预见可能性对应比较容易采取的结果避免措施；高程度的结果预见可能性则对应立即停止其行为的负担较大的结果避免措施^[50]。

从危惧感说的核心观点出发，结果预见可能性与结果避免义务之间存在相互关联性。一方面，行为人是否对危害结果存在预见可能性应当以法律要求其采取的结果避免措施的强度进行判断，当法律要求行为人承担高强度的结果避免义务时，行为人对于危害结果的发生需要具备高度的预见可能性；当法律只要求行为人承担低强度的结果避免义务时，行为人只需要对于危害结果的发生存在低度的预见可能性。即使危害结果发生的概率极低，如果法律只要求行为人承担最为基本的结果避免义务，采取最基本的措施，则认为行为人对于发生率极低的危害结果存在预见可能性也是合理的。另一方面，当法律没有明文规定行为人应当采取的结果避免措施时，行为人对于危害结果发生的预见可能性决定着行为人应当采取的结果避免措施的强度。当行为人对危害结果发生具备高度的预见可能性时，法律应当要求行为人采取负担较重的结果避免措施；当行为人对危害结果发生只具备低度的预见可能性时，法律应当要求行为人采取负担较轻的结果避免措施。行为人对危害结果的预见可能性取决于其注意能力，这种注意能力属于主观违法要素，因而支配着行为人应当采取的结果避免措施的强度，当行为人的注意能力高于别人，行为人应当承担的结果避免义务的负担也重于别人^[51]。因此，从危惧感说的立场出发，结果预见可能性的程度和结果避免义务的强度不是一成不变的，而应当根据二者之间的相互关联性，在每个案件中进行具体、个别的判断。

(三) 主流观点对危惧感说的些许片面理解

在森永毒奶粉案中，森永公司需要采取的结果避免措施非常基本：只需要检查一下购买的第二磷酸苏打是否掺杂了非食品物质，这种措施是每个食品生产者都必须采取的，即使奶粉中掺杂了砒霜导致婴幼儿死亡的结果发生概率极低，甚至是万分之一，但由于森永公司承担的结果避免义务非常简单，因此森永公司对上述结果存在预见可能性。藤木教授提出的危惧感说标准在该案中是合

^①日本高松高等裁判所认为：“把食品添加剂之外的、为供其他目的使用而制造的药品……添加到食品中时，应该会对使用上述药品感到一缕不安。正是这种不安感，表现出对本案中争议的危险具有预见。”最终森永公司的制造课长、工厂厂长成立业务过失致死伤罪。

适的,但是在其他案件中却未必合适,甚至是不合理的。所有过失案件中预见可能性和结果避免义务的判断应当考虑二者的相互关联性,而非将专属于森永毒奶粉案的危惧感、不安感套用于全部过失案件,主流观点对危惧感说的批评可能存在一些片面之处。

由此可见,修正的旧过失论要求的具体预见可能性标准与算法决策的黑箱属性不相容,新过失论没有对预见可能性的标准进行设计。与二者相比,虽然危惧感说对预见可能性的理解较为抽象,但这种抽象的要求只是个别案件的体现。危惧感说的核心观点不仅能够比较容易地判断预见可能性,而且能够根据预见可能性程度的高低决定结果避免义务负担的轻重,是判断人工智能产品算法设计者是否存在过失的合适立场。

五、危惧感说视野下人工智能算法设计者的犯罪过失构造

根据危惧感说的立场,过失的构造是:客观的结果预见可能性、客观的结果预见义务、客观的结果避免义务。在分析算法设计者是否存在犯罪过失时,应当依此构造进行判断,同时,应当从结果预见可能性与结果避免义务的相互关联性出发,判断结果避免措施的内容。

(一) 客观的结果预见可能性的判断

1. 客观的结果预见可能性程度:特殊情况下的“一旦式”标准

如前所述,人工智能产品装载的算法具有明显的黑箱属性,算法通过深度学习能够进行自主性再编程,搭配了算法的人工智能往往能够实施一些匪夷所思的行为^[11]，“作为外部观察者,我们无法确知人工智能产品内部的具体运算过程,而只能看到其最终的结果反馈”^[52]。算法黑箱让现实世界与算法之间产生了一道天然隔阂^[53],具备足够的知识、经验,能够理性思考和慎重决策的算法设计者似乎无法预见其所设计的算法系统到底会做出怎样的决策,进而会引发怎样的后果。

然而,如果因为算法设计者在大多数情况下无法预见算法的决策就否认客观的结果预见可能性存在是不合理的结论,因为程序设计者虽只是制造抽象的算法,但该程序设计却涉及未来的风险。故针对其程序设计可能所致的危害结果,有必要讨论刑事责任问题^{[5]206-207}。虽然算法设计者很多情况下无法预见算法决策所引发的危害结果,但仍然存在若干能够承认算法设计者预见可能性的情况,在这些情况中,即使人工智能机器的算法设计者无法预见算法到底会做出什么决策,以及究竟哪一个决策会引发哪种危害结果,他们也能预见算法系统可能会在某些特殊情况下做出“不好的”决策。

算法系统可能做出不利决策的场合,应当包含了异常因素的特殊情况。算法设计者只能信赖人工智能产品的算法在没有任何异常因素出现的正常情况下会做出正常决策,只能在人工智能产品只应用于没有异常因素出现的正常情况下主张自己对算法决策引发的损害不存在预见可能性。当算法不仅会被用于无异常因素的通常情况(比如自动驾驶系统只会被用于天气晴朗,路况良好的城市道路),还会被用于某些特殊情况(比如手术机器人不仅会被用于操作简单手术,还会被用于操作复杂手术)时,设计者就对算法可能做出不利于他人的决策进而引发危害结果存在预见可能性。此时的预见可能性,不要求理性一般的算法设计者必须对危害结果发生的因果关系历程的基本部分存在预见,即不要求理性一般的算法设计者必须预见到其设计的算法系统究竟会在具体哪一种情况下做出哪一种决策,以及该决策会引发哪种具体的危害结果。只需要算法设计者预见到一旦遭遇包括了异常因素的特殊情况,算法有可能做出不利决策,进而在外界引发消极后果这种程度就

已足够,这种程度的预见可能性是一种低程度的预见可能性。

2. 特殊情况的标准

上述包含异常因素发生的特殊情况应当采取一定的标准进行判断,由于现阶段的人工智能主要用于辅助人力决策,因此特殊情况的判断标准应当是:现阶段只有人类才能正确处理或者只有人类才最适合处理的异常情况,这些情况应当根据人工智能产品的类型分别判断。如,现阶段许多复杂手术在实施过程中会发生非常复杂的变化,面对这些复杂变化,只有人类医生才能根据情况随时变更手术步骤,才能在手术过程中根据不同情况选择不同的措施。与人类医生相比,目前的手术机器人还没有具备人类医生的经验和知识,无法应对特殊情况。又如,作为自动驾驶算法系统之一的车道维持系统目前只能根据车辆偏离主干道的情况将自动驾驶汽车拉回原位,但当驾驶员前方突然出现行人,或者疾病发作,必须将车辆急转弯到路边的土堆或者草丛中时,这套算法还是会判定车辆偏航进而将车拉回原位,阿沙芬堡案^②就是典型案例。由此可见,承认人工智能产品算法设计者对危害结果存在预见可能性的特殊情况标准应当是现阶段只有人类才能正确处理或者最适合处理的情况。

(二) 客观的注意义务的内容

1. 客观的结果预见义务

(1) 理性一般的人工智能产品算法设计者应当预见其所设计的算法系统不仅会被用于没有异常因素出现的正常情况,而且可能被用于伴随异常因素出现的特殊情况。理性一般的算法设计者在编写应用于现实需要的算法时,一般情况下都会考虑其所设计的算法可能会在将来某个时间遭遇异常情况,因此,算法设计者结果预见义务的第一个方面是设计者应当对这些特殊情况有所预料,这些包含异常因素的特殊情况应当根据人工智能机器的类型进行细化。比如,手术机器人不仅会被用于实施一些难度不大的手术(如阑尾炎手术),也可能被用于辅助医生实施一些难度较大的手术(如心脏搭桥);不仅会被用于给身体条件正常的患者实施手术,还可能被用于给一些特殊体质(如血友病)的患者实施手术。又如,自动驾驶系统不仅会被用于城市道路驾驶,而且可能被用于盘山路驾驶;不仅会被用于晴朗天气驾驶,而且可能用于沙尘、暴雨等天气驾驶。这些产品的诊疗算法设计者、自动驾驶算法设计者不仅应当预见该类算法可能被用于正常手术、驾驶的情况,而且应当预见该类算法可能被用于特殊手术、驾驶的情况。但必须指出:这些产品的算法设计者只需要预见算法可能被应用于特殊情况就已足够,至于这些特殊情况的详细内容(如血友病患者、沙尘暴天气)则不属于预见义务的内容。

(2) 理性一般的人工智能产品算法设计者应当预见一旦其所设计的算法系统面临特殊情况,该系统可能会做出对他人不利的决策,而这种决策会引发消极后果。算法设计者结果预见义务的第二方面是对算法系统可能做出的不利决策的预料。由于算法系统能够通过机器学习进行内部程序码的重写,因此设计者能够预见其设计的算法系统可能会做出偏离预期的决策,进而会对外界产生消极影响,而这种对他人不利的决策也应根据人工智能产品的类型进行细化。比如,突发行人冲上道路的异常情况,自动驾驶系统可能会在面临冲向行人从而维持正常道路行驶还是冲向路边草丛

^②该案发生于2012年,基本案情是:一辆装载了车道维持自动驾驶系统的汽车行驶在道路上时,驾驶员突然中风,在失去意识之前,该驾驶员将车辆向右急转弯,在正常情况下该车会停在路边草地上,然而车道维持系统此时“完美运作”,将已经右转的车辆再次拉回道路上以确保其仍在道路上行驶,最终该车撞上一对母子致其当场死亡。

从而保护行人的特殊情况时做不利于行人的决策,自动驾驶算法的设计者应当预见该类算法面对特殊情况可能做出不利决定,但是与结果预见义务的第一方面相同,算法设计者只需要预见一旦算法面对特殊情况可能做出不利决策就已足够,无论这种决策是对谁不利,至于特殊情况的详细内容,则非所问。

2. 客观的结果避免义务

根据危惧感说的核心观点,过失犯罪的结果预见可能性与结果避免义务相互关联,上述“一旦式”的结果预见可能性属于低程度的结果预见可能性,因此与这种程度的预见可能性相对应,算法设计者应当承担的结果避免义务也应当是负担较轻的义务。

(1) 理性一般的人工智能产品算法设计者应当避免在设计算法时添加含有偏见的价值理念。“算法的运行规则始终是由人设计并控制的”^[54],值得信任的算法应体现出人类的道德原则。虽然算法设计者无法准确预知自己设计的算法到底会做出怎样的决策,但是设计者至少可以决定用什么样的理念来设计算法。因此在算法设计研发的过程中,设计者应仔细考虑放入何种价值观,否则算法会成为一种自私自利、自我导向的怪物^[55]。由此可见,算法设计者植入算法的价值理念对于算法决策将起到重要作用,如果不注重对算法“价值观”的塑造,将一系列歪曲价值理念写入算法,算法在被歪曲理念影响下所做的决策结果很可能也是歪曲的,即使算法的初衷与法律的精神或宗旨相符,但如果在算法设计中没有完整地考虑法律价值及权利保障目标,也有可能引起此方面的风险^[56]。

因此,人工智能产品的算法设计者在设计装载于智能机器上的算法时,应当避免将一些不正常的,乃至明显歪曲的价值理念植入算法之中,这些不正常的价值理念应当满足一定标准:不正常的价值观是为我国社会公众普遍反对或者不赞成的价值观,而不能是只被一小部分甚至个别人反对的价值观。比如,驾驶汽车遭遇他人“别车”时,也要反“别”回去这一“人若别我,我必别人”的价值理念是一种为社会公众普遍反对的不正确驾驶理念,自动驾驶系统的设计者在设计时不能将这种价值理念写入算法。

(2) 理性一般的人工智能产品算法设计者应当在设计时检验“投喂”给算法系统的数据质量。机器学习算法本质上是向算法“投喂”大量数据,使算法系统能够依托数据集合进行模型训练,算法决策就是建立在海量数据的基础之上。然而,机器学习算法的模型训练存在一个很大的问题,也就是人工智能的悖论:“Garbage in, garbage out”(垃圾进,垃圾出),意思是不好或不对的数据会产生糟糕或是无用的结论^[57]。当作为训练样本的数据存在缺陷,人工智能产品的算法在面对外部情况时很有可能基于缺陷数据引发的不完整训练做出不利决策。为了防止算法系统做出不利决策,设计者应当在设计算法时对算法所依赖的数据集合进行事前检测。

目前,人工智能技术已经完全可以为算法设计者提供数据集合及训练过程检测工具(例如数据集合异常检测工具、数据完整性检测工具、训练样本评估工具等),主要用于检测训练算法的数据集合是否存在偏差或缺陷,还可以通过一定的算法检测在数据选取、数据标注、数据清洗以及其他预处理工作(如正则化、归一化等)过程中是否包含了偏离算法设计目标或足以导致结果发生显著偏差的操作^[56]。算法设计者应当通过这些检测工具对算法所依赖的数据集合进行检验,最大程度防止缺陷数据进入机器学习训练。根据危惧感说的核心观点,算法设计者的低程度预见可能性决定的结果避免措施应当是负担较轻的,这种事先对数据集合进行检验的应用场景应当是人工智能产品

面对的不包含异常因素的正常情况,即使算法设计者无法对所有情况进行预见,其也应当采取最为基本的、负担最轻的检验措施,即采取检验手段对会被应用于正常情况的算法的训练样本数据进行检测。比如,在自动驾驶汽车中,算法设计者应当对会被应用于没有异常因素的正常道路交通状况下自动驾驶系统的决策所依靠的数据集合进行事前检查,从而修正数据集合可能存在的缺陷,否则,算法设计者就违背了结果避免义务。典型案例是发生在邯郸市的特斯拉自动驾驶汽车车祸案^③。本案中,特斯拉的自动驾驶系统在正常驾驶情况下(天气晴朗)未能识别出前方有车,进而做出继续行驶不变道的错误决策,最终导致驾驶员死亡^[58],而自动驾驶系统未能识别出前方有车的键是作为该系统训练样本的数据存在缺陷,但设计者没有注意到该缺陷。因此,特斯拉自动驾驶系统设计者没有采取负担较轻的结果避免措施(只需要凭借数据检验工具在事前检查一下就已足够),违反了客观的结果避免义务。

(3) 理性一般的人工智能产品算法设计者应当及时告知人工智能产品生产者其设计的算法可能遭遇异常情况,从而使生产者能够针对潜在的特殊情况采取一定程度的警示和提示措施。如果可以通过提供合理的指示或警告来减少或避免产品造成的可预见危害,则没有进行警告就是有缺陷的。当风险可预见并且大量用户不会意识到风险时,警告或指示就是合理的^{[2]37}。智能产品装载的算法完全有可能产生一些设计者能够预见但很多人没有意识到的风险,这些风险指前文所述的包含异常因素的特殊情况。这种情况下,算法设计者理应告知人工智能产品生产者其设计的算法可能会被应用于包括异常因素的特殊情况,进而生产者能够对这些特殊情况的具体内容进行一定程度评估,从而制订若干警示和指示,使智能产品的使用者能够及时采取措施。以自动驾驶系统为例,自动驾驶系统的算法设计者在设计算法后应当及时告知该系统可能会遭遇的特殊情况,生产者进而可以将这些特殊情况具体化(比如沙尘暴、暴风雨等),从而制订人类驾驶员必须在某种情况下接管机动车,同时关闭自动驾驶系统的警示。

结语

人工智能产品不断被应用到人们的日常生活中,算法决策正在多个领域不断取代人类决策,探讨算法决策致害时的归责问题,明确谁应当对损害承担责任是一个不容忽视的问题。在判断算法致害的归责问题时,明确算法设计者是否存在过失及过失的构造和内容至关重要。本文在此基础上展开讨论,未来将继续深入研究这一问题,以为算法决策致害的归责问题提供更加合理的见解。

参考文献:

- [1] 付玉明. 自动驾驶汽车事故的刑事归责与教义展开[J]. 法学, 2020(9): 135-152.
- [2] 瑞恩·卡洛, 迈克尔·弗鲁姆金, 伊恩·克尔. 人工智能与法律的对话[M]. 陈吉栋, 董惠敏, 杭颖颖, 译. 上海: 上海人民出版社, 2018: 3.
- [3] GAL M S. Algorithmic challenges to autonomous choice[J]. Michigan Technology Law Review, 2018, 25(1): 59-104.

^③ 该案的基本案情是: 2016年1月, 一名男子驾驶自己的特斯拉轿车在京港澳高速河北邯郸段公路行驶, 未能及时躲避前方的道路清扫车而发生追尾, 事故导致该名车主身亡。据事故后行车记录仪中的视频分析, 事故发生时, 天气晴好, 特斯拉速度并不快, 但已处于“定速”的状态, 未能识别躲闪而撞上前车。

- [4] 李开复,王咏刚.人工智慧来了[M].台北:远见天下文化出版股份有限公司,2017:58.
- [5] 张丽卿.人工智慧时代的刑法挑战与对应:以自动驾驶车为例[M]//张丽卿.人工智慧与法律冲击.台北:元照出版有限公司,2020:190-212.
- [6] 埃里克·希尔根多夫.自动化驾驶的伦理及法律之挑战[M]//张丽卿.人工智慧与法律挑战.台北:元照出版有限公司,2020:396-417.
- [7] 丁晓东.论算法的法律规制[J].中国社会科学,2020(12):138-159,203.
- [8] 劳东燕.“人脸识别第一案”判决的法理分析[J].环球法律评论,2022(1):146-161.
- [9] 张凌寒.权力之治:人工智能时代的算法规制[M].上海:上海人民出版社,2021:28.
- [10] DIAKOPOULOS N. Algorithmic accountability: Journalistic investigation of computational power structures [J]. Digital Journalism, 2015, 3(3): 398-415.
- [11] YUR, ALİGS. What's inside the black box? AI challenges for lawyers and researchers [J]. Legal Information Management, 2019, 19(1): 2-13.
- [12] 许玉秀.主观与客观之间:主观理论与客观归责[M].北京:法律出版社,2008:155.
- [13] 劳东燕.风险社会中的刑法:社会转型与刑法理论的变迁[M].北京:北京大学出版社,2015:24-25.
- [14] 许迺曼.过失犯在现代工业社会的捉襟见肘[G]//许玉秀,陈志辉.不移不惑献身法与正义:许迺曼教授刑事法论文选辑.台北:新学林出版股份有限公司,2006:521.
- [15] 周光权.刑法总论[M].第4版.北京:中国人民大学出版社,2021:160-163.
- [16] 西田典之.刑法總論[M].第3版.橋爪隆,補訂.东京:弘文堂,2021:271-272.
- [17] 曹菲.管理监督过失研究:多角度的审视与重构[M].北京:法律出版社,2013:42.
- [18] 汉斯·韦尔策尔.目的行为论导论:刑法理论的新图景[M].增补第4版.陈璇,译.北京:中国人民大学出版社,2015:11-13.
- [19] 周光权.行为无价值论与主观违法要素[J].国家检察官学院学报,2015(1):78-92,174.
- [20] 大塚仁.刑法概说(总论)[M].冯军,译.北京:中国人民大学出版社,2003:199.
- [21] WELZEL H. Das neue bild des strafrechtssystems: Eine einfuehrung in die finale handlungslehre [M]. Verlag Otto Schwartz & Co. Göttingen, 1961: 30-32.
- [22] 陈子平.刑法总论[M].2008年增修版.北京:中国人民大学出版社,2009:149.
- [23] 大谷实.刑法讲义总论[M].黎宏,姚培培,译.北京:中国人民大学出版社,2023:185.
- [24] Wessel/Beulke/Satzger. Strafrecht Allgemeiner Teil: Die strafat und ihr aufbau [M]. C. F. Müller, 2023, § 18, 1107.
- [25] 井田良.关于日本过失犯论之现状[J].黄士轩,译.月旦法学杂志,2014(12):283-295.
- [26] 张明楷.刑法学[M].第6版.北京:法律出版社,2021:328-382.
- [27] 陈兴良.教义刑法学[M].第3版.北京:中国人民大学出版社,2017:511.
- [28] 王纪轩.人工智慧于司法实务的应用[M]//张丽卿.人工智慧与法律冲击.台北:元照出版有限公司,2020:112-142.
- [29] INTRONA L D. Algorithms, governance, and governmentality: On governing academic writing [J]. Science, Technology & Human Values, 2016, 41(1): 17-49.
- [30] BATHAEE Y. The artificial intelligence black box and the failure of intent and causation [J]. Harvard Journal of Law & Technology, 2018, 31(2): 889-938.
- [31] 葛祥林.数位化、大数据和人工智慧对刑事诉讼的冲击[M]//张丽卿.人工智慧与法律冲击.台北:元照出版有限公司,2020:216-236.
- [32] 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.
- [33] BODO B, HELBERGER N, IRION K, et al. Tackling the algorithmic control crisis - the technical, legal, and ethical challenges of research into algorithmic agents [J]. the Yale Journal of Law & Technology, 2017, 19: 133-181.

- [34] 王莹. 算法侵害责任框架刍议[J]. 中国法学, 2022(3):165-184.
- [35] SELBSTAD, BAROCAS S. The intuitive appeal of explainable machines[J]. Fordham Law Review, 2018, 87(3):1085-1139.
- [36] 江湖. 中国大陆自动驾驶汽车对法律的挑战[M]//张丽卿. 人工智能与法律冲击. 台北:元照出版有限公司, 2020:83.
- [37] 许恒达. 过失犯的预见可能性与回避可能性[J]. 台湾中研院法学期刊, 2020(27):55-139.
- [38] 蔡圣伟. 刑法案例解析方法论[M]. 第3版. 台北:元照出版有限公司, 2020:218.
- [39] 山口厚. 刑法总论[M]. 第3版. 付立庆, 译. 北京:中国人民大学出版社, 2018:242-251.
- [40] European Parliament. Civil liability regime for artificial intelligence, European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)). [EB/OL]. (2020-10-20) [2022-05-25]. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html.
- [41] LEHR D, OHM P. Playing with the data: What legal scholars should learn about machine learning[J]. Davis Law Review, 2017, 51(2):653-718.
- [42] SCHÜNEMANN B. Unzulänglichkeiten des fahrlässigkeitsdelikts in der modernen industriegesellschaft-eine bestandsaufnahme [M]. Berlin, Boston: DE GRUYTER, 2002:37-63.
- [43] 阿图尔·考夫曼. 法律哲学[M]. 第2版. 刘幸义, 译. 北京:法律出版社, 2011:319.
- [44] 蔡仙. 过失犯中的结果避免可能性研究[M]. 北京:法律出版社, 2020:21.
- [45] 姚万勤. 新过失论与人工智能过失刑事风险的规制[J]. 法治研究, 2019(4):98-107.
- [46] 黄荣坚. 基础刑法学(上)[M]. 台北:元照出版有限公司, 2012:391.
- [47] 刘艳红. 实质犯罪论[M]. 北京:中国人民大学出版社, 2014:195.
- [48] 藤木英雄. 公害犯罪[M]. 丛选功, 译. 北京:中国政法大学出版社, 1992:62.
- [49] 高桥则夫. 刑法总论[M]. 李世阳, 译. 北京:中国政法大学出版社, 2020:191.
- [50] 井田良. 講義刑法学・総論[M]. 第2版. 东京:有斐阁, 2018:217.
- [51] 井田良. 過失犯における「注意義務の標準」をめぐって[J]. 刑法雑誌, 2003, 42.
- [52] 黄陈辰. 论人工智能缺陷产品生产者的刑事责任[J]. 山东大学学报(哲学社会科学版), 2020(6):49-56.
- [53] 衣俊霖. 数字孪生时代的法律与问责:通过技术标准透视算法黑箱[J]. 东方法学, 2021(4):77-92.
- [54] 金梦. 立法伦理与算法正义:算法主体行为的法律规制[J]. 政法论坛, 2021(1):29-40.
- [55] 沈向洋, 施博德. 计算未来:人工智能及其社会角色[M]. 北京:北京大学出版社, 2018.
- [56] 苏宁. 算法规制的谱系[J]. 中国法学, 2020(3):165-184.
- [57] 王禄生. 司法大数据与人工智能开发的技术障碍[J]. 中国法律评论, 2018(2):46-53.
- [58] 观察者网. 邯郸特斯拉事故致死案:公司承认案发时处“自动驾驶”状态[EB/OL]. (2018-02-27) [2023-03-11]. https://www.guancha.cn/society/2018_02_27_448303.shtml.

The criminal negligence of the designer of artificial intelligence products: Taking a stand on the central perspective of the sense of fearing theory

LIU Jiazheng

(School of Law, Southeast University, Nanjing 211189, P. R. China)

Abstract: When judging the negligence of the designer of artificial intelligence product algorithm, the opinion of modified theory of old negligence is in conflict with the nature of black box of algorithm that depends on correlation rather than causation when making decisions; the logic of this standpoint that only valuing results while ignoring conduct may blow algorithm designers' enthusiasm, impeding algorithm's progressing; Although

the new negligence theory takes the obligation of result avoidance as the core standard of criminal negligence, it lacks specific design of the standard of possibility of foreseeing and is often at a loss when judging the possibility of foreseeing. Therefore, both standpoints are not reasonable schemes to judge the criminal negligence of intelligent products algorithm designers. In contrast, although the perspective of the sense of fearing theory (hyper new negligence theory) holds that the possibility of results foreseeing only requires the conductor to have a sense of fearing about the harmful result is enough, which is criticized by the mainstream view, this criticism is worth of discussion. Firstly, it only sees the fear requirement on the surface of this stance, but it does not see the core view behinds of the perspective of the theory of fearing: The correlation between the possibility of foreseeing and the obligation of avoiding results. Secondly, the opinion that the author's judgment on individual cases is equivalent to the theory of fearing itself is an overgeneralization. Compared with the modified old negligence theory and the new negligence theory, the core point of the theory of fearing that there is correlation between the possibility of the foreseeing of result and the obligation of result avoidance is a reasonable scheme to judge the criminal negligence of the algorithm designer of intelligent products. Based on the core view of the theory of fearing, criminal negligence includes objective possibility of the possibility of foreseeing results, objective obligation of foreseeing results and objective obligation of avoiding results. The criterion for the objective the possibility of foreseeing results of the algorithm designer of artificial intelligence products is that the algorithm system is likely to make an adverse decision once it encounters a special situation containing abnormal factors, which may lead to negative consequences. The objective result avoiding obligation of the algorithm designer is that the algorithm system should be foreseen not only for the normal situation without abnormal factors, but also for the special situation accompanied by abnormal factors. Once the designed algorithm system encounters special situations, the system may make adverse decisions. The content of objective result avoiding obligation of algorithm designer is as follows: it is necessary to avoid implanting values that are generally opposed or disapproved by the public when designing algorithms; The quality of the data fed to the algorithm system should be checked at designing time to prevent the risk of "garbage in" of defective data into the algorithm machine learning training to the greatest extent, Informing the product producer promptly that the algorithm may face abnormal conditions.

Key words: algorithm; possibility of results foreseeing; the sense of fearing theory; objective obligation of results foreseeing; objective obligation of results avoidance

(责任编辑 胡志平)