重庆大学学报(社会科学版) 2023年第29卷第5期
JOURNAL OF CHONGQING UNIVERSITY(Social Science Edition) Vol. 29 No. 5 2023

Doi:10.11835/j.issn.1008-5831.fx.2022.02.002

欢迎按以下格式引用:侯东德,张可法."人工智能黑客"的法律规制[J].重庆大学学报(社会科学版),2023(5):184-197. Doi:10.11835/j.issn.1008-5831.fx.2022.02.002.

Citation Format:HOU Dongde, ZHANG Kefa. Legal regulation of "artificial intelligence hacker"[J]. Journal of Chongqing University (Social Science Edition),2023(5):184-197. Doi:10.11835/ j.issn.1008-5831.fx.2022.02.002.

"人工智能黑客"的法律规制

侯东德,张可法

(西南政法大学 民商法学院,重庆 401120)

摘要:从历史上看,"黑客"迭代到"人工智能黑客",是伴随着计算机、互联网、大数据和人工智能等科学技术迅猛发展而产生的。现如今的"人工智能黑客"是人机交互体,既非人也非物,介于两者之间,它可以模仿人类、干扰人类认知,为达到设计者或决策者的目的对网络系统漏洞进行智能化侵入和破坏。"人工智能黑客"区别于传统"黑客"的主要特征在于其可以依靠智能算法自主学习、寻找网络系统代码漏洞和加强分布式攻击。部分学者将人工智能技术划分为弱人工智能、强人工智能、超人工智能三个阶段,甚至有学者建议从伦理上赋予强人工智能法律主体地位,赋权理由是强人工智能算法具有独立的"机器意思"表示能力,与人类有情感的联结。显然,这种赋权方式不仅违背"人本主义"原则的主体创新,而且现行法律主体包括自然人、法人、非法人组织,"人工智能黑客"不属于任何一类主体,突兀地将法律主体的理性意思表示与人工智能算法指令的"机器意思"相等同,容易形成"人工智能黑客"行为在算法正义法律评价和民事法律行为构造上的困境,干扰我们对"人工智能黑客"本质的判断。溯本清源,应当以法律上权利义务构造标准去判断"人工智能黑客"的法律属性。"人工智能黑客"本质上是自然人主体通过人工智能算法技术,利用网络媒介进行网络侵权或犯罪的行为。"人工智能黑客"的核心是通过计算机代码设置、大数据运算与机器自动化判断进行决策的一套机制。"人工智能黑客"在责任承担上不是适格的法律主体,只具有特殊的"人格性工具"法律属性。"人工智能黑客"的智能化攻击外在表现为算法程序的自动执行,但程序的设计和算法运行归属于现实经济生活中的人,也完全符合法律上间接侵权的调整范畴。对于"人工智能黑客"的侵权或犯罪行为,应当通过揭开"人工智能黑客"的"面纱",找到其背后隐藏的可规制法律主客体,利用"穿透"方式对"人工智能黑客"的非法行为进行伦理、技术和法律三个维度的有效规制。

关键词:人工智能黑客;算法;人格性工具;法律规制

中图分类号:D920.0;TP18 **文献标志码**:A **文章编号**:1008-5831(2023)05-0184-14

基金项目:教育部哲学社会科学重大攻关项目"人工智能发展中的重大风险防范体系研究"(20JZD026)

作者简介:侯东德,西南政法大学教授,博士研究生导师,Email:gxin_001@163.com;张可法,西南政法大学博士研究生,Email:12825027@qq.com。

20世纪50年代,逻辑学家艾伦·图灵在哲学期刊《心灵》上发表了一篇论文,题目是《计算机器与智能》,这篇文章使“人工智能”这一概念正式进入人们的视野①。随后计算机技术迅速迭代更新,“深蓝”机器人战胜了国际象棋棋王卡斯帕罗夫,成为了人工智能历史上的里程碑。在机器计算力成几何倍数增长的推动下,大数据互联网信息革命实现了从线上互联到移动互联,再到智能互联的飞跃发展,引起了包括生产方式在内的社会全方位重大变革,尤其是人工智能在持续学习、理解和解决问题方面得到普遍发展,并已经深深地嵌入到我们的社会结构中②。智能“小冰”机器人能在极短时间内创作诗篇③,“阿尔法狗”通过深度学习,利用深度神经网络技术能自主识别信息[1]。目前,在ChatGPT类技术的赋能下,人机交互在人工智能的发展中起着举足轻重的作用,人工智能由“分析式智能”向“生成式智能”进化[2]。因此,人工智能系统自身亦可成为“黑客”不再是危言耸听,它利用智能算法按照既定程序自主寻找目标网络漏洞,并以快速度、大规模方式攻击人类的社会经济生活④。该行为不仅是简单的网络攻击行为,还带来了社会治理所涉及的法理、制度规范和司法实践转型升级问题,亟需我们认真对待和积极回应。

一、“人工智能黑客”的概念厘清

(一)“黑客”的发展历史

“黑客”一词源自英文“Hacker”。《牛津英语词典》对“Hacker”作如下解释:“利用自己在计算机方面的技术,设法在未经授权的情况下访问计算机文件或网络的人。”“黑客”一般是指在计算机科学、编程和代码设计方面具有高度理解的人。19世纪的美国社会记录着黑客的技术生活,他们是最优秀的代码编写者,其中包括研究修改计算机产品的业余爱好者[3]。部分“黑客”被称为“电话飞客”,当时的“黑客”一词被用来称呼研究盗用电话系统的人士,这个特殊人群坚守协作创新,其法律行为主体是自然人,被称为“黑客”1.0级。

20世纪80年代,互联网技术飞速发展,个人计算机进入公众视野,个人电脑的广泛普及,成为“黑客”历史发展的分水岭。越来越多的人们开始使用计算机电脑,“黑客”主要攻击网络邮箱盗取有价值的信息。后来诞生了木马病毒,可以致使大批计算机瘫痪,“黑客”开始热衷于挑战各种高难度的病毒程序编写。随着计算机技术的发展,众多“黑客”工具与软件快速流行起来,通过自学成为一名“黑客”变成可能。“黑客”群体鱼目混杂且各自目的不尽相同,甚至有“激进黑客”组织盗取并发布机密文档,

---

①“图灵测试”检验方法的提出,标志着人工智能作为独立的学科正式诞生。“图灵测试”是指把一个人和一台计算机分别置于不同的封闭空间。人和计算机与外界交流的方式仅限于文字信息,人类裁判在另一个封闭空间,并不知道人和计算机所处的具体空间。人类裁判通过分别提问,并根据回答的情况判断哪个空间是人,哪个空间是计算机。如果作为裁判的人类不能做出正确的判断,那就可以认为,这台计算机具备了人的智能。

②《深圳经济特区人工智能产业促进条例(草案)》第2条规定:“本条例所称人工智能,是指运用人工的方法和技术,利用计算机或者其控制的设备,通过对收集的外部数据进行学习、分析,感知环境、获取知识、推导演绎,研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术以及应用的能力。”

③小冰是一套完整的、面向交互全程的人工智能交互主体基础框架,又叫小冰框架。少女小冰,是诗人、歌手、主持人、画家和设计师,与其他人工智能不同,小冰注重人工智能在拟合人类情商维度的发展,强调人工智能情商,而非任务完成,并不断学习优秀的人类创造者的能力,创造与相应人类创造者同等质量水准的作品。

④2021年7月5日,北京商报报道了新闻《黑客猖獗 一天内全球千家公司遇袭》。“黑客”将目标锁定在IT管理软件供应商Kaseya(美国资讯科技管理公司)上,通过袭击Kaseya公司一个名为VSA的工具,向使用该公司技术的管理服务提供商(MSP)进行勒索,同时加密这些供应商客户的文件。报道指出,现在“黑客”的攻击已经不仅仅是针对单个公司了,而是盯上了为成百上千公司提供服务的IT服务供应商。

揭露政府或组织秘密,充当网络侠客。此时的“黑客”属于“互联网+黑客”的组合,法律行为主体仍然是自然人,互联网只是“黑客”行为发生的场景和工具而已。同时,“黑客”技术发展的目的也有所变化,商业化的软件基金会鼓励“黑客”在推动优质软件和程序中发挥作用,并以高昂的奖金作为报酬。商业逐利性挟裹推动“黑客”逐利性行为,这种“黑客”可以称之为“黑客”2.0级。

随着21世纪大数据和人工智能技术的发展,脑科学和神经技术飞速发展并与计算机技术紧密结合,促进了类似于人脑智能计算机更高层次应用的突破。在大算力驱动强算法处理大数据支持下,算法模型的人机交互能力进一步提升,人工智能打破“人”和“物”的绝对界限,创造出了“非人非物”的新“物种”,特别是生成式人工智能具备自动捕获人类的偏好,编写恶意代码和注入攻击的能力,给人类未来存在的方式带来了巨大的不确定性。人工智能技术通过和现有互联网技术的结合,立足于物理世界之上,通过嵌入式智能算法,开辟了有别于传统物理设计的虚拟新空间。科学技术具有“中性”色彩,可以作为生产要素促进生产力的发展,亦可作为破坏经济社会发展的工具。“人工智能黑客”是人和智能机器的交互结合,生物人通过算法程序将自己的目的和意图嵌入智能程序中,该智能程序会依照算法模块自主查找漏洞和执行算法指令。“黑客”的行为分成了两部分,一部分是生物人对程序算法的事先设计;另一部分是算法程序模块通过智能技术寻找网络漏洞,模仿人类、干扰人类认知,为达到自己目的进行智能化侵入和破坏。此时的“黑客”显然成为了“人工智能+黑客”的人机组合,属性复杂,我们可以称之为“黑客”3.0级。

从历史路径看,“自然人黑客”发展到“人工智能黑客”阶段,“人工智能黑客”的人机交互特征明显区别于“自然人黑客”及自然人主导的工具性“黑客”,它既具有部分人的主体属性,能自主学习和简单思考,会自主触发设定条件下的程序运行;又具有部分物的属性,就如同拟制人的躯干一般,它是算法和程序代码的载体和运行组合。因此,“人工智能黑客”的属性定性事关承担责任的资格问题。如何依据法律解释学方法,进一步分析“人工智能黑客”行为的特征、法律地位及可调整性,是当下亟需解决的问题。

(二)“人工智能黑客”的特征

“人工智能黑客”是“黑客”主体的数据化和智能化,这种智慧化程度区别于黑客1.0和黑客2.0级别,具有明显的其他特征。“人工智能黑客”的攻击一般依赖于被建模或攻击的人类系统必须以计算机能够理解的方式模块化。“人工智能黑客”能够通过精确的目标函数算法,了解人类系统的发展、变化及漏洞所在,利用超强计算力优势,以人类无法预料的方式攻击目标网络系统漏洞。

1.具有学习能力

人工智能技术是把“双刃剑”,给网络安全带来了发展机遇和挑战。智能算法就像给操作系统安装了大脑,“人工智能黑客”操纵算法自主学习和运行,算法背后的决策者从而获得想要的结果。算法的自我深度学习和思考是人工智能科学的一个研究分支,核心在于通过对大数据信息分析和模仿人类大脑的认知,做到“数字化+智能化”的高度融合。“人工智能+黑客”人机交互,是一种类人化智能技术,它在信息技术、仿生学、脑神经学、社会学等大数据信息的支撑下,通过程序编辑模块模拟人的推理和解决问题的能力。算法赋予程序像人一样对存储的数据进行独立思考、分析、推理,这种逻辑一定程度上突破了机器原本的智慧框架,放大了机器对数据的自主学习判断,产生了一种机器的自我简单意识。目前,机器学习方面的深度学习已经能够自主建立模式识别模型。譬如,“人工智能黑客”利用算法分析大量被盗信息记录,识别潜在受害者特点并发送有效针对钓鱼类的电子邮件,获取非法利益。

智能化算法甚至通过分析人类植入的无序面部表情、呼吸、心率等数据来阅读人们的情绪变化,它的优势不仅在于拥有快速的运算能力,还能在某些过程中自主发现问题,为用户量身定做攻击诱饵,用超强的运算力催生出难以解释的算法逻辑⑤,并在科学技术迭代中实现升级。

2. 算法具有不可解释性

算法的可解释性可以从两方面去理解:一方面是算法模式是如何运行的?这是算法功能主义的视角;另一方面是算法到底能产生什么?这是从算法的事后解释视角而言的,其包含算法语言、算法模块的可视化工具等。显然,这些并没有触及可解释性的本质,要探究算法的可解释性,就必须理解技术原理及算法运行的实践,明确算法可解释的路径及解释程度。人工智能是脑科学、神经学、心理学、生物学、机械学、电子学和信息学等学科的有机整体,它增强了系统的行为、感知和认知能力。迄今为止,算法的可解释性在学界和实务界既未达成共识,也未探索出一种相对主观的角度进行描述。此处的“解释”是指人类与算法决策者的一种交互行为,双向构成一个面,就是说解释行为本身是算法决策者的精准代理,且该解释又能为人类所理解。如果达不到这种双向目标,就不具备可解释性,即智能算法具有不可解释性。人工智能运算是函数关系的逻辑表达,我们目前能做的仅仅是算法模型的设计符合法律要求。而强人工智能通过深度学习后产生的不可解释,给“人工智能黑客”的攻击留下了足够的生存空间。显然,人类也是积极通过履行算法透明义务类似的方式,通过可视化、示例或元素重新排列等不完全精准的方式应对算法的不可解释性。

3. 主动寻找代码漏洞

目前很多公共系统都使用了开源机器学习库,攻击者可以随意查看其中的代码,从而寻找可供利用的漏洞,譬如,2022 年 4 月,西北工业大学遭美国国家安全局“黑客”攻击,窃取该校关键网络设备配置等核心技术数据。“黑客”只需要掌握很少的系统信息,便可制作出有效的对抗性样本,尤其“人工智能黑客”可以利用技术侵入的手段,通过预先设计的算法逻辑,锁定攻击目标(一般是金融机构或 AI 企业),用算法程序自动去寻找、检测被攻击对象的系统漏洞,进行目标追踪、攻击样本的篡改、私自加密重要文件或者模拟人的行为发出勒索要求等,这些行为都是在算法程序中进行了事前行为预设,预设的可能性方案几乎达到了穷尽。因此,从发生概率的角度而言,只要有漏洞存在,就会被预先设定的算法技术捕获,“人工智能黑客”寻找漏洞的能力大大超越传统“黑客”。

4. 加强分布式拒绝服务攻击

分布式拒绝服务攻击是指大量受黑客控制的僵尸主机对目标主机发起攻击行动的行为[4],其特点类似于军事上的饱和性攻击,“人工智能黑客”进一步加强了这种攻击的能力。智能化“黑客”可以在不同位置潜伏下来,多个攻击者同时向一个或数个远程目标发动攻击,或者一个“人工智能黑客”攻击者控制位于不同位置的多台机器,利用程序代码控制这些已受侵入的机器对漏洞目标同时实施攻击。这是目前最难防御的网络攻击行为。“人工智能黑客”不像人一样具有伦理道德的价值判断能力,也不会存在犯意中止的觉醒,它会以人类无法预料的方式入侵系统,这种攻击速度快、规模大、危险性强。具体表现为:一方面,“人工智能黑客”利用大数据和智能算法快的技术优势,病毒程序传播速度惊人,它能在最短的时间甚至于几秒就可以完成一个侵入网络系统的过程;另一方面,“人工智能黑

⑤2016 年,人工智能程序“阿尔法狗”在与世界顶级围棋棋手李世石的对弈中获胜,其最著名的一手是第 2 局的第 37 手,这手棋是高水平人类棋手绝不会选择的一手,人工智能程序“阿尔法狗”的选择很难加以解释。

客”在程序设计之初的目标明确,自动按照程序代码寻找网络系统漏洞,它会潜伏等待触发时机,也会通过欺骗技术获得网络使用人的信任,以极快的速度发现可以利用的漏洞,将以人们无法预测的规模进行攻击,破坏力巨大,危险性更强。

(三)“人工智能黑客”的法律属性

纵观民事主体发展“人可非人”到“非人可人”的历史脉络,阐释了法律主体只是满足社会需要的法律形式。从古罗马法到现代民法,主体范围在不断扩大[5]。譬如,《德国民法典》在确立法人人格时,创造性提出“权利能力”概念并不具备伦理学上“人”的所有特性[6]。主体资格的认定尽管要考虑伦理性,但科技的进步使该因素被弱化,主体资格的确认还应当关切财产和责任主体的独立性。法人人格的确认,是出于一种功利主义视角,这种主体资格认定的异化不会产生负外部性。“人工智能黑客”虽然逐渐在脱离法律客体“物”之属性,但是异化后的主体人格有限,其本质仍是一种“人格性工具”。因此,对于“人工智能+黑客”的人机交互行为,赋予“人工智能黑客”有限人格,还是确定为“人格性工具”,这是建立“人工智能黑客”监管整体框架的首要任务。

法律主体说认为人工智能具备人格属性,法律应当明确承认人工智能的法律关系主体资格。理论基础是人工智能在超级阶段存在感情表达,具备意识能力,与人类有情感上的联结⑥。法律客观说认为人工智能不具有自然人的自然属性,也不同于立法技术拟制下产生的独立法律人格,缺失主体适格的法理基础,人工智能就是编程技术代码、算法的综合体,不具有法律主体地位,具有民事权利上的“工具性人格”。

赋予法律主体资格的目的是保障社会有序发展,“人本主义”是任何制度的基本原则遵循,人之主体地位的获得,依赖于人在进化发展过程中养成的自由意志[7],人工智能法理也不例外。人工智能主体说基于假象的人工智能可能存在的某种能力,在未区别算法系统展示的机器“自我意识”和“人本主义”下人类“独立的意思表示”的巨大区别,不加分析和澄清就赋予人工智能主体资格实属荒谬。同样,作为“人工智能黑客”的主体认定也就成了无稽之谈。“人格性工具”的定位,有利于人工智能法律治理的实现。譬如,“人工智能黑客”冒充财务负责人给财务人员发送钓鱼电子邮件或语音,指示其完成一笔特定转账。客观说能从法律关系的构造上越过人工智能的主体因素,直接寻找真正的责任主体,而非虚拟化的“人工智能黑客”主体。

笔者认为,“人工智能黑客”建立的“算法—学习—思考—行动”的自主性逻辑机构,并不能合乎法理地得出“人工智能黑客”法律人格的结论,况且“自主性”行为的判断主要依据意思自治和决策能力,自我学习和简单思考只是人工智能发展的方向。“工具性人格”判断的本质不是理性逻辑结构,构建法律上的权利和义务关系框架才是其认定标准。

在“人工智能黑客”法律关系的认定中,责任承担的主体是“人工智能黑客”的创造者、设计者、所有者和管理者等,权利义务关系的调整无须另建一套存在瑕疵的人工智能主体法律归责体系。只需在现有法律框架下,充分考虑“人工智能黑客”的特殊性,尤其是智能化程序和算法区别于一般客体的特点,在法律关系的构造以及逻辑关系上进行梳理,其方法是在现有相关规范基础之上,运用解释工具进行调整或在制度供给上适当增加。

---

⑥我们将人工智能按照发展程度分为弱人工智能、强人工智能和超级人工智能三个阶段,有些学者认为,超级人工智能具有独立的自我意识,具备意识能力。

二、“人工智能黑客”带来的法律挑战

“黑客”是非法侵入计算机系统的行为人,人工智能“人格性工具”身份的表达通过算法或算法系统实现。“人工智能+黑客”人机交互结合,会产生如何区分是人的故意的行为还是算法载体的问题,加上算法的不易解释性和不透明性,人们很难理解它的逻辑或其决策机制,使这个问题更加复杂化。“人工智能黑客”由他人编写运行程序,利用大数据技术支撑,在算法设计的范围框架下智能系统自动运行的结果,这种网络侵入或攻击并不是漫无目的的,通常是从利益者本位出发。美国学者蒂姆·林报道了一项深入研究,声称网络犯罪分子使用深度造假和其他人工智能技术试图破解银行、企业和政府使用的人脸和语音识别安全[8]。“黑客”攻击行为影响了人工智能软件及数据算法决策控制者的利益⑦。而且这种“人格性工具”否定性行为后果还面临回应算法程序介入法律价值、法律关系以及法律行为带来的价值判断问题。

(一)算法正义性法律价值的评价问题

人工智能技术是新型生产力,该种技术力量会引起社会生活方式和社会价值观念的深刻变化,随之评价该行为的法律制度也会面临变革,尤其数据和算法作为重要的新型生产要素,它在运行中是否具有“正义观”的价值评价会产生争议,就此得出法律评价的结论颇具困难。理由在于,“人工智能黑客”具有自动寻找、检测系统漏洞并主动攻击的能力,在评价该攻击行为时,算法命令的始作俑者是出于技术本身的缺陷,或是在商业利益和不良动机的诱惑下,通过算法对互联网系统嵌入内核套件、流氓软件、勒索软件、病毒软件等“恶意代码”并进行人为控制,背后真相似乎没有那么简单。现实中设计算法和代码的程序员被赋予绝对的决策权,且对算法审查有限[9]。这些程序员像是互联网空间的规则制定者,建成了一个虚拟秩序下的算法“黑箱”,它挑战人类决策的知情权。在算法场景中,不透明的算法,造成非人本身成为了决策主体,人的自主性面临考验。“人工智能黑客”是例外情形下的一种执行恶意算法规则的工具,在特定算法中寻找有价值的网络漏洞,盗取个人隐私信息或商业信息牟利,危及社会诚信与公平公正,甚至会通过金融手段违法犯罪。如何规制算法、防范算法风险,是关乎算法正义法律价值的重要内容。

(二)算法介入法律关系的构建

信息技术革命打造了现实与虚拟相互交融的同构生态,传统的生产方式在物理空间上得到突破,创设了新型法律关系主体和客体,尤其是人工智能技术开启了人机融合的新时代,智能化的生产方式,加入了非人类的机器主体,不断冲击着传统世界的确定性,当机器变得足够复杂的时候,机器将不再是人类的朋友,也不是人类的主人,而是人类的伙伴[10]。智能机器人人格属性问题日渐突出。主体作出行为,客体承受行为,一个具有人格性的客体会被如何认定?这种不确定的区分将会带来法律上的难解之题[11]134。智能行为介入生产,甚至于发生侵权行为,譬如自动驾驶汽车,“黑客”发起的某一次攻击行为,都会对传统法律行为主体、客体范畴的定义、内涵、外延、法律属性等带来挑战。传统意义上的

⑦新华每日电讯报道:2021年5月10日美国联邦调查局发布声明称,美国主要成品油管道运营商之一科洛尼尔管道运输公司遭到黑客攻击,旗下承载着美国东海岸近45%供油量的输油干线7日起被迫关闭,至今尚未恢复工作,导致美国多个州和地区燃油供应面临危机。这起切断美国供油“大动脉”的黑客攻击来自哪里?如何发生?为何对美国能源基础设施运营带来如此巨大影响?据介绍,“黑暗面”勒索软件采取逐渐成为主要趋势的“双重勒索”策略,首先窃取存储在受害者系统内的敏感信息,然后对这些敏感数据加密并发出换取密钥的赎金要求。

权责利关系正面临重塑,该示例所涉及主体包括汽车制造商、软件开发商、汽车所有人、程序开发人员和交通事故受害者。算法赋值自我检测漏洞能力下智能化攻击,涉及长链条的众多利益相关者,其软件编写程序员众多,算法和代码迭代周期短,易造成“人工智能黑客”的身份难以捕捉,如何归责不易判断,诸如种种问题就会变得异常复杂和模糊,其结果只能是链条上的责任相关者都会成为被告,出现“法不责众”的尴尬结局[11]136。人工智能本来是技术要素,“黑客”智能程序自主寻找漏洞并侵入的行为,显然将一方的权利扩大,譬如某些富人想利用人工智能来逃税,“人工智能黑客”就会通过算法黑箱稳固有利于特定群体的权力结构。

(三)对传统法律行为界定的干扰

法律行为属于法律事实的一种,是能引起法律关系产生、变更和消灭的活动(行为)。同法律事件不同,它以人的意志为转移,是人们有意识的活动。“人工智能黑客”在虚拟化和匿名化特性的遮掩下,使算法动机、目的和因果关系复杂难辨。在人机混合模式下,人工智能获得深度学习的算法训练,智能算法的运行是设计者的意思表示还是机器本身纯粹的“算法意识”?“人工智能黑客”是决策者的意志还是机器人自主学习的意志?机器人的深度学习开发,会出现连程序员都无法解释的高度不确定性和难以控制性,给法律行为的认定带来重大挑战。智能化发展使人机协同更加紧密,计算机智能程序、互联网体系从一种工具变成了“代理人”,对人的道德评价很难在机器上实现,通过机器行为和责任脱钩将会成为普遍现象,越来越多的人将责任转嫁到机器身上[12]。在自动驾驶的交通事故,无人机错判误杀平民,“人工智能黑客”攻击金融网络等问题上,“谁来为机器的行为负责”,是技术设计缺陷或其他原因?计算机系统中的错误或不当行为的来源可以原则上溯及至软件设计作出的决定人,无论合理与否[13]。恰恰追责难度在于,程序不像一个产品,有出厂的地址,程序的设计可能由多人开发,程序的产生可能无法追踪到某个具体的个人或组织[14]。为了回应“人工智能黑客”带来的法律难题,我们必须去伪存真,找到“人工智能黑客”的真实面貌。如前面论述,“人工智能黑客”具有“人格性工具”属性,是自然人主体通过算法技术利用网络媒介进行网络犯罪的行为,“人工智能黑客”的核心在于程序算法,也就是说,对“人工智能黑客”规制的核心在于对智能算法的规制。

三、“人工智能黑客”的可规制解构

可规制性探讨的是某一行为主体是否具有可归责的资格。“人工智能黑客”的到来,似乎把算法程序设计者推向了幕后,智能算法自动执行的“人格性工具”成为假象,在算法的设计者与算法作用的对象之间产生了阻隔,形成“算法程序开发的人”—“人工智能黑客”(算法实现)—“被攻击的对象”这样的逻辑路径。但是,智能算法的本质是人机交互决策,其逻辑结构是且、或、非。即使是自主决策系统,也只是在且、或、非三种逻辑结构中程式化、模块化地作出决定[15]。现有算法不是完全确定和完全随机的,受程序化边界的限制,是为了快速实现某个目标对一组数据进行处理的逻辑步骤,形式上算法系统自主作出法律行为,但实质上此种智能化的意思表示仍然是设计者意思表示的延伸。因此,在算法与法律行为的关系上,二者不能作同类型化对待。

算法可作狭义、广义和中义的定义,狭义的算法源于数学与计算科学,是用于表述解决数学与计算科学难题的一系列规则,狭义的算法被视为纯粹的科学或技术。例如数据结构算法、数论与代数算法、计算几何算法等。广义的算法是指算法不仅应用于数学计算科学领域,还被广泛应用于其他社会科学领域,例如广泛使用各领域为减少决策步骤和程序的运算,进行自动化决策,证券自动交易系统就属于

这种范畴。中义算法介于狭义和广义之间,界定为人类和机器交互的决策,即人类通过代码设置、数据运算与机器自动化判断进行决策的一套机制[16]。人工智能算法可以通过大数据技术形成一定的学习能力,甚至达到深度学习的程度。但是,智能算法的学习区别于人类大脑的学习,它并不是具有情感和认知层面的学习。举世瞩目的沙特机器人公民“索菲亚”,外形与生物人几乎无差别,但其本质脱离不了算法和数据处理器。“索菲亚”的行为完全靠智能算法主导,按照工作逻辑框架,将人的语言和对外界的感知在控制系统储存的众多数据中进行加工和处理,进而输出语言和特定的指令动作。

“人工智能黑客”采用的算法是自动化决策,自主寻找网络漏洞或者专门针对特定攻击目标的恶意算法,这种算法既不是数学或计算机科学意义上的算法,也不是纯粹关于人类行为的决策算法[17]。“人工智能黑客”的算法既有人的决策,又有机器的自动判断,属于中义上的算法。这种算法除了纯粹的技术因素,还通过场景化的方式把算法自动执行的指令深入介入生活的方方面面。例如为达到特定的商业目的,利用算法实现最大推送量、获得最高点击率、追求利益最大化[18]。攻击者还可以通过设计故意触发缺陷的程序实现某种获利。现实中有一种叫作“双层爱尔兰夹荷兰三明治”的避税方法,美国高科技企业利用智能算法在多国之间寻找税收法律系统漏洞,通过其设在爱尔兰和荷兰的子公司,将企业利润转移至低税或无税管辖区,从而逃避其应当在美国缴纳的税款⑧。

总而言之,机器的决策具有客观性,是主体实现目标的媒介。人的目的性决策取决于程序的编制和算法的构建,是借助技术手段把主体的主观善恶植入算法中,算法程序设计者通过客体的媒介,利用算法技术,为达到不可告人的特定目的。媒介恰恰成了幕后决策者的“面纱”,有这层“面纱”的隔离,就会阻碍算法正义和法律行为构造上嵌入价值判断,如果不揭去“面纱”,对价值与伦理问题视而不见,就很可能忽视算法对人类价值伦理所带来的挑战。“人工智能黑客”的算法并非价值的中立,在“人本主义”理念下,算法系统隐含着幕后算法程序开发者的意思表示也并非例外。

四、“人工智能黑客”的规制措施

人工智能带来的担忧不是空穴来风,究其根本,是人们来自内心深处对人工智能不确定性的恐慌[19]。人工智能系统越智能化,它就越不透明和不可解释。“人工智能黑客”恰恰利用了算法的不透明和不可解释性,以及发现漏洞的突出能力,给金融、国家安全等重点领域带来巨大威胁,甚至在被视为“黑客”天堂的域外拉美各地出现了利用当前社会危机获取人们个人信息的情形[20]。因此,社会需要建立能够快速有效应对攻击的弹性治理架构。

(一)设定人工智能科技伦理软规则

1. 构建人工智能算法的伦理规范

智能算法时代,形成了物理(现实)/电子(虚拟)的双重空间,数据挖掘和智能算法,编织出全新的智能生态。在虚拟空间算法为王,掌握算法决策权就意味着掌握了虚拟空间话语权[21]。也就是说,算法成为一种特殊的经济资源,在一定条件下深刻影响着社会秩序的建构。

由于人工智能被嵌入人类的知识、程序和算法,具有了一定的自主学习能力,并模仿人的智慧性甚至感情,因此“人工智能黑客”会大大超越以前没有灵性的机器,人工智能在给人们带来欣喜的

⑧双层爱尔兰夹荷兰三明治国际避税模式[EB/OL]. [2021-12-21]. https://wenku.baidu.com/view/057050e5a31614791711cc7931b765ce04087a14.ht.

同时挑战着法律、伦理和秩序。科技的发展离不开自由和创新精神,一般情况下,人工智能设计者不会理会技术引发的伦理问题。设计算法技术时对道德观重视不足,代码和算法的秘密操控技术也许会掌控在社会责任感冷漠的人手中,由此来看,除了法律规制,需要构建可靠的科技伦理规范。2021年发布的《深圳经济特区人工智能产业促进条例(草案)》第68条规定设立人工智能伦理委员会⑨,构造人工智能的伦理框架。一方面,需要将道德算法嵌入算法体系当中,程序设计者有道德上的信义义务编写程序,程序的运行不能嵌入非道德意识指令,更不能人为创设算法"黑箱",融入个人或少部分人的利益因素;另一方面,在使用者和人工智能互动场景中加强道德建构[22],加强对技术利用的指导,让使用智能算法的人在道德高地不能利用算法技术牟取私利,甚至利用算法漏洞进行违法犯罪行为。要用道德义务抑制不公平、不道德的因素渗透到代码和算法的应用场景,从而减少"人工智能黑客"利用漏洞带来的社会风险。

2. 寻求利益相关者的共同支持

为实现对人工智能技术风险的有效治理,实现人工智能"善治",我们需按照"以人为本"的伦理发展路径,让算法相关决策者、使用者在管理过程中适当让利益相关者参与其中,在算法技术风险问题上进行多方合作,并且对利益相关者之间的利益进行协调平衡,建立有效沟通机制,使人工智能算法在设计上受到技术道德约束,使用者受到法律制度约束,决策管理者受到信义义务约束。只有充分考虑人工智能产品伦理的可规制性、社会运用的平稳赞许性和发展的"人本主义"可持续性,才能建立智能算法设计者、人工智能产业界和市场末端利益一体化,对"人工智能黑客"治理目标一体化,从而达到对超越技术设置框架的"黑客"行为的规制。

(二)智能技术对智能"黑客"的中性监管选择

代码和算法是人工智能的基础和关键,也是"人工智能黑客"的真实面貌,它决定着人工智能行为的识别判断、逻辑能力和行为方式。"人工智能黑客"的行为在代码和算法技术框架下进行,应对"黑客"风险的存在,离不开设计开发者的技术自律规制。美国颁行《为人工智能的未来做好准备》规则,第1条明确指出了自我审视⑩。同时,面对"人工智能黑客"隐蔽性强、技术程序复杂的现状,加强应对"人工智能黑客"的人工智能技术是最佳的选择。

1. 发展人工智能监管"黑客"的技术

法律对人工智能安全风险防范的评价是一种价值判断,技术自身的功能价值由技术客观属性决定。同理,"人工智能黑客"产生的社会后果归属于法律去评价,但是人工智能技术被"黑客"利用,或者该项技术就是一种"黑客"的智能算法程序,那么,对于技术上的风险防范通过技术渠道更为科学。一是设置风险防控智能决策程序。对于每一项算法程序运行的框架边界,首先要进行科学的评估,在此基础上,用更先进的人工智能技术监督算法程序的运行情况,如果算法运行有逃逸

⑨《深圳经济特区人工智能产业促进条例(草案)》第68条:市人民政府应当按照国家人工智能治理相关规定,设立人工智能伦理委员会,履行下列职责:(一)研究制定人工智能领域的伦理规范;(二)建立健全人工智能伦理安全标准管理制度,引导和规范人工智能伦理安全标准的制订和实施;(三)对数据垄断、算法歧视、智能滥用、深度造假、数据投毒、隐私保护、伦理道德、不平等智能操作以及对社会结构的影响等重点领域开展监测与研判;(四)发布人工智能伦理安全实践指南、人工智能伦理安全白皮书以及人工智能企业伦理安全治理优秀案例集等,引导不同类型的人工智能企业建立完善伦理安全治理制度;(五)评估、监督本市人工智能企业的伦理规范执行情况;(六)其他应当开展的活动。

⑩《为人工智能的未来做好准备》第1条建议:鼓励私人和公共机构自我审视,判断自身是否能够,并通过何种方式负责任地以造福社会的方式利用人工智能和机器学习。

设置框架边界的风险,就会启动智能决策程序,对该算法程序运行的风险进行评估和跟踪,对使用者单位或个人进行预警或提醒,这种技术能对技术的监督产生“安全阀”的效果。二是建立人工智能技术开发审慎原则。在技术开发上加强前瞻性预防和约束性引导,建立安全、可靠和可控底线思维模式,坚决制止将价值判断的决策权利交给算法,避免算法在超强计算力的支撑下,形成“制定标准、编写代码的人拥有的控制力量”[23]。三是优化算法可解释性及透明度。人们对人工智能计算的担忧在于其具有不确定性,普通人根本不会知道算法运行的原理和背后的目的,算法可以利用人类的弱点欺骗人类。可解释性义务本质是针对算法设计的,实质在于保证算法模型的设计符合法律要求,使算法能够被监管者知悉是否在正当的框架内运行。算法的透明度在于确保使用者在一定程度了解算法运行的原理,并对涉及权益的算法决策建立合理预期。四是设计人工智能监督软件,训练其辨别网络正常与可疑行为之间差别的能力,形成“防火墙”,当发现潜在威胁时,做到及时预警,跟踪、锁定目标,并切断网络侵入⑪。

2. 加强国际间人工智能技术的合作

2017 年 1 月,美国计算机协会下属的美国公共政策委员会发布《算法透明性和可问责性声明》文件,其规定了七条规则:其一,相关主体应当意识到算法中可能存在的偏见和潜在危害;其二,监管机构应当构建一定的机制,以便因算法决策而受到不利影响的个人和组织能够提出质疑并得到救济;其三,算法做出决策的后果应当由使用它的机构承担;其四,鼓励相关组织和机构对算法的具体决策进行解释;其五,算法的设计者有义务对训练数据加以说明;其六,模型算法、数据和决策应当被记录以便审查;其七,算法构建者应当定期做测试以评估该算法是否会产生损害性结果,并鼓励将结果公之于众⑫。2018 年 1 月,美国纽约州通过《算法问责法案》,其对人工智能的算法可解释性、可问责性等提出了要求,以应对算法黑箱和算法错误等问题。2019 年 4 月,欧盟发布《人工智能伦理准则》。2020 年 2 月,罗马发布《罗马人工智能伦理宣言》等。我国也应当积极参与和应对,秉承开放合作的态度,与各国携手探索人工智能的科学前沿,进一步完善人工智能及网络安全战略的不足,逐步提升我国人工智能安全的国际话语权[24],共同推动人工智能的创新应用。

(三)加强“穿透问责”的规则约束

“人工智能黑客”在法律人格属性上表现出浓烈的技术性和替换性。“人工智能黑客”主体的不适格,并不代表智能“黑客”攻击行为造成的损失无法追究。主体不适格是从主体类型化的角度,给予“人工智能黑客”主体资格否定性评价。从“人工智能黑客”功能主义视角,“人工智能黑客”作为“人格性工具”适格,物的功能发挥主要是主体利用客体产生一定的法律关系。“人工智能黑客”的产生源自人工智能设计开发者,智能“黑客”的行为就是设计者程序编制的意思表示外在延伸,是一种责任主体和特殊客体的利用与工具关系,因此,采用“穿透”问责的方式能更好理清法律关系。

1. 决策者的刑事责任

人工智能技术的非法应用会产生两种危害:一种是人工智能体被滥用或误用造成的社会危害;另一种是人工智能体借助“深度学习”技术做出“自主决策”和行为导致的社会危害,“人工智能黑客”就属于第二种情况。针对利用“人工智能黑客”攻击人类的行为,从“人格性工具”归责出发,强

⑪中国科学院成都文献情报中心信息科技战略情报团队. CERN 科学家用人工智能新手段防御黑客攻击[J]. 中国教育网络,2017(8):43.

⑫ACM US Public Policy Council. Statement on algorithmic transparency and accountability[EB/OL]. (2017-01-12)[2021-12-21]. https://www.acm.org/articles/bulletins/2017/january/usacm-statement-algorithmic-accountability.

调“人工智能黑客”是否具有刑事责任主体资格。从目前主流观点的法律地位否定说而言,一方面“人工智能黑客”没有独立于人类的权利和价值,只有利益附属性;另一方面“人工智能黑客”没有人类的思想观念和伦理感知。因此,“人工智能黑客”不能被拟制为犯罪主体,不具备刑法预防和惩治犯罪的功能承担。否则,决策者会利用“人工智能黑客”的主体地位逃避法律责任承担。

从实证角度出发,“人工智能黑客”程序的运行始发于设计编程,设计者是人工智能的实际控制人,哪怕是“黑客”行为,也没有脱离程序的控制。“人工智能黑客”不具备承担责任的自主控制能力,因为人的自主控制能力带有内心意思表示,“人工智能黑客”即使是自主触发指令,那也是在设计者预先设计好的框架下触发的,触发的攻击不是机器思考的结果,是设计者事前预设的结果,这种设计者的事先预设行为就具有主观上的故意性。“人工智能黑客”引起的社会危害性构成犯罪的,必然要承担刑事责任。

因此,为维持网络空间的安全与秩序,打击利用计算机系统与网络信息技术实施犯罪的行为人,《中华人民共和国刑法》规定了一系列关于破坏计算机信息系统的犯罪,以保护计算机信息系统安全。《中华人民共和国刑法修正案(九)》(以下简称《刑法修正案(九)》)又补充规定了一系列网络犯罪罪名⑬,以保障网络空间安全和网络服务业的健康、有序发展。因此,故意利用人工智能进行犯罪,如故意杀人、盗取金融财产、偷税漏税等,此时的人工智能程序算法应当被认定为犯罪工具,依照故意犯罪追究算法程序的设计者、编制者责任。譬如,当“黑客”入侵非计算机场景系统盗窃数据信息时,可参照“非法侵入计算机系统罪”惩戒犯罪⑭。倘若“黑客”入侵并控制了汽车的智能系统涉及危害社会公共安全时,犯罪手段和结果形成了牵连关系,为保护更大法益,则应当按照侵害公共安全性故意犯罪论处[25]。如果是因为设计者的过失导致,在刑事责任评价上,应当预见人工智能程序在运行过程中会产生“黑客”程序的客观后果,由于技术上的疏忽或轻信发生概率极低而忽略,程序设计者承担过失犯罪的刑事责任。

假如设计者既无故意,也无过失,即人工智能脱离了设计编程实现设计框架,应当认定为脱离了人的控制,不再是人的工具,符合人工智能产生时因技术局限而无法意识到的缺陷的抗辩事由,属于意外事件。另外,刑法作为维护社会秩序的最后一道屏障,在发挥预防和惩戒犯罪功能时,为促进人工智能技术的发展,应当保持适当的谦抑性。

2. 决策者的行政责任

行政处罚作为法律制裁的一种方式,能够在一定程度上剥夺或限制违法行为人的人身和财产权利,具有惩戒性,是网络安全行政法保护的重要保障。网络空间安全的行政处罚对象是违反网络信息安全秩序的行政相对人。我国实行网络违法“一案双查”制度,即在对网络违法犯罪案件开展侦查调查工作时,同步启动对涉案网络服务提供者法定网络安全义务履行情况的监督检查。在“人工智能黑客”非法行为的行政责任承担上,“人工智能黑客”不具备行政相对人的资格,当其违法行为发生且不构成犯罪时,它对社会秩序造成的破坏由相适应的法律规范加以评价和约束。“人工智

⑬最高人民法院、最高人民检察院发布7起侵犯公民个人信息犯罪典型案例之五:“肖凡、周浩等侵犯公民个人信息案”——利用黑客手段窃取公民个人信息出售牟利,构成侵犯公民个人信息罪。

⑭“肖开海非法获取计算机信息系统数据案”——以黑客手段窃取手机ID密码的刑事责任,(2017)粤20刑终258号。该案中犯罪嫌疑人以黑客手段,侵入苹果手机系统,通过有关软件获取该苹果手机ID密码,再高价予以出售。苹果手机ID密码可以在任何计算机终端使用,用于确认用户在计算机信息系统上操作权限的数据,是身份认证信息,属于计算机信息系统数据。犯罪嫌疑人的上述窃取行为符合非法获取计算机信息系统数据罪犯罪特征。

能黑客”的实际控制人是适格的行政相对人主体,有义务承担相应的法律后果。

《中华人民共和国网络安全法》(以下简称《网络安全法》)列举了实施危害网络安全的具体行为,责任承担的处罚种类涵盖警告、罚款、信用惩戒等;《中华人民共和国计算机信息系统安全保护条例》规定了危害计算机系统安全的行政处罚条款,其中危害行为包括故意输入计算机病毒以及其他有害数据危害计算机信息系统安全,未经许可销售出售计算机信息系统安全专用产品。《计算机信息网络国际联网安全保护管理办法》专设法律责任一章规定了危害网络安全的违法行为。《互联网信息服务管理办法》对互联网平台的责任作出规定,条款主要针对非法网络行为。“人工智能黑客”对信息系统的侵入及智能算法程序非法传播就属于上述规定所调整的危害行为。

同时,当“人工智能黑客”在网络空间的违法行为发生刑事和行政法处罚竞合的情况时,譬如对实际控制人的从业禁止如何适用,《刑法修正案(九)》中采用准用性规则:其他法律、行政法规对其从事相关职业另有禁止或限制性规定的,从其规定。此规范在实践中应当理解为可适用于竞合行为。考虑到刑法打击违法行为的谦抑性,行政法对于从业者的从业禁止处罚适用优先于刑法规定。

3. 决策者的民事责任

民法是调整平等主体之间人身关系和财产关系的法律规范,民事责任的承担在于法律的规定和当事人之间的约定,一般归责原则下,无权利侵害就不必承担法律责任。要确定“人工智能黑客”在民事领域的责任承担,就面临确认侵权责任的主体、“人工智能黑客”的侵权行为追责对象、归责原则与风险分配如何设置等问题。在侵权行为构成要件中关于过错认定和因果关系方面,“人工智能黑客”对其产生了冲击,一方面“人工智能黑客”致损的发生机制复杂,其中包括设计缺陷、他人干扰和决策者不当利用等原因;另一方面,“人工智能黑客”的自主运行和学习能力加剧了侵权行为的不透明和不可预见性,多种可能性并存必然超越侵权责任构成要件的有效射程。因此,刺破“人工智能黑客”的“面纱”,向“人工智能黑客”的实际控制人追究民事责任,能高效厘清法律关系,及时定分止争,惩罚利用人工智能技术侵权的行为人。

实践中,“人工智能黑客”攻击目标由决策者类型化方式确定,并通过算法的自主性运行实现。“黑客”会按照程序算法设定的参数具体化锁定漏洞并实施侵入。比如,“黑客”可以进入别人的网站,更改网站内容,或者将重要内容进行删除、修改,以非法手段搜集个人信息中的私密信息。《中华人民共和国民法典》第四编“人格权”第 1035 条规定了个人信息搜集和处理的原则及具体规则;《网络安全法》第 42 条确立了使用个人信息的禁止性规则和采取技术保护措施的义务。因此,一方面,根据民法上人与物二分性基本原则[26],“人工智能黑客”是特殊的“人格性工具”,即使在超人工智能阶段,也无法摆脱工具的客观属性。上述侵权行为可以分为行为人的直接侵权和“工具归属于人的间接侵权”。“人工智能黑客”的攻击外在表现为算法程序的自动执行,但程序的设计和运行算法归属于现实的人,完全符合民法上的间接侵权调整范畴。另一方面,设立民事连带责任。基于“人工智能黑客”行为会影响传统意义上的民事责任转移,使多方利益相关者可能承担连带或单独的赔偿责任[27],具体可以设定设计者、决策者接受算法审查义务,以算法是否接受审查为标准设置不同的责任形式。如接受主管部门审查通过的,设计者、决策者承担有限责任;未经审查进行商用的,设计者、决策者承担无限连带责任。

五、结语

“黑客”技术随着科技发展实现自我迭代。现如今,大数据时代来临,智能算法对人类经济生活

的影响越来越深刻,借助生成式人工智能的代码自动生成和理解能力,可以支持网络攻击,使网络安全形势更加严峻[28],“人工智能黑客”是技术发展对法治治理提出的挑战。在面对“人工智能黑客”法律属性问题上,本文梳理了现有观点,强调人工智能的发展必须遵循“人本主义”这一根本,它发展的每个阶段都是在生物人的主导之下,法律关系的主体类型不能随意类型化出“机器人”或“人工智能黑客”主体。“人工智能黑客”是人机交互,具有“人格性工具”属性,只有明确了这一法律地位后,我们才能解决人工智能给法律行为、法律关系带来的挑战,才可以对智能算法的设计者、决策者提出算法伦理义务、算法设计道德要求,用人工智能技术对付“人工智能黑客”。总之,智能算法会越来越多地内嵌于不同技术场景,无论智能“黑客”是否超越3.0阶段,智能算法的本质不会变,我们应当揭开智能算法的“面纱”与价值中立性,对算法进行规制,甚至“穿透”算法去寻找幕后的算法决策者。唯此,智能算法才能避免异化并预防“人工智能黑客”,算法科技才会给社会带来福祉。

参考文献:

[1]徐献军.人工智能的极限与未来[J].自然辩证法通讯,2018(1):27-32.

[2]王禄生.ChatGPT类技术:法律人工智能的改进者还是颠覆者?[J].政法论坛,2023(4):49-62.

[3]RITTENHOUSED. Slantwise moves:Games,literature and social invention in nineteenth-century America respawn:Gamers,hachers,and technogenic life[J]. American Historical Review,2021,93(2:)336-337.

[4]FRANCESCO P. Energy-oriented denial of service attacks: An emerging menace for large cloud infrastructures[J]. The Journal of Supercomputing,2015,71(5):1620-1641.

[5]许中缘.论智能机器人的工具性人格[J].法学评论,2018(5):153-164.

[6]卡尔·拉伦茨.德国民法通论[M].王晓晔,邵建东,程建英,等,译.北京:法律出版社,2004:100-101.

[7]刘练军.人工智能法律主体论的法理反思[J].现代法学,2021(4):73-88.

[8]RING T. Europol: The AI hacker threat to biometrics[J]. Biometric Technology Today,2021,2: 9-11.

[9]卢克·多梅尔.算法时代:新经济的新引擎[M].胡小锐,钟毅,译.北京:中信出版社,2016:214.

[10]约翰·马尔科夫.人工智能简史[M].郭雪,译.杭州:浙江人民出版社,2017:208.

[11]克里斯多夫·库克里克.微粒社会:数字化时代的社会模式[M].黄昆,夏柯,译.北京:中信出版社,2018:134-136.

[12]皮埃罗·斯加鲁菲.智能的本质:人工智能与机器人领域的64个大问题[M].任莉,张建宇,译.北京:人民邮电出版社,2017:170-171.

[13]尤瑞恩·范登·霍文.信息技术与道德哲学[M].赵迎欢,宋吉鑫,张勤,译.北京:科学出版社,2014:223.

[14]王利明.人工智能时代对民法学的新挑战[J].东方法学,2018(3):4-9.

[15]刘颖.论算法与法律行为的关系:制度影响与法律回应[J/OL].重庆大学学报(社会科学版),2021. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CAPJ&dbname=CAPJLAST&filename=CDSK20211202001&uniplatform=NZKPT&v=tq_pSMDZi_922eEYwQDfNN4rwXRh5GCJTv2NaydB8pIBNdsV-uY6dLethwUonI8c.

[16]丁晓东.论算法的法律规制[J].中国社会科学,2020(12):138-159,203.

[17]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.

[18]宣言.不能让算法决定内容[N].人民日报,2017-10-05(4).

[19]马锋,张军锐.当高新技术风险遭遇媒介:不确定性的终结与恐慌的生产[J].陕西师范大学学报(哲学社会科学版),2015(3):172-176.

[20]STEPHEN W. Hackers paradise: Hackers across Latin America are taking advantage of the current crisis to access People's personal data. If not protected it could spell disaster[J]. Index on Censorship,2020,49(2):40-42.

[21]张康之.数据治理:认识与建构的向度[J].电子政务,2018(1):2-13.

[22]田海平.让“算法”遵循“善法”[N].光明日报,2017-09-04(15).

[23]詹姆斯·柯兰,娜塔莉·芬顿,德斯·弗里德曼.互联网的误读[M].何道宽,译.北京:中国人民大学出版社,2014:122.

[24]侯东德,姚万勤.美国网络安全战略及其对我国的启示:兼论我国《网络安全法》的规定及未来的完善[J].人工智能

法学研究,2019(1):77-89.
[25]张依楠.黑客攻击自动驾驶汽车:犯罪风险及刑法规制[J].智能网联汽车,2021(2):54-60.
[26]吴汉东.人工智能生成发明的专利法之问[J].当代法学,2019(4):24-38.
[27]LIU H C. National regulations and local rules-A hybrid regulatory model of intelligent connected vehicles in China[J]. Advances in Social Sciences Research Journal,2021,8(2):85-101.
[28]张弛,翁方宸,张玉清.Chat GPT在网络安全领域的应用、现状与趋势[J].信息安全研究,2023(6):500-509.

Legal regulation of “artificial intelligence hacker”

HOU Dongde, ZHANG Kefa
(*College of Civil and Commercial Law, Southwest University of Political Science and Law, Chongqing 401120, P. R. China*)

Abstract: Historically, the iteration of “hacker” to “artificial intelligence hacker” was accompanied by the rapid development of science and technology, such as computer, internet, big data and artificial intelligence. Nowadays, “artificial intelligence hacker” is a human-computer interaction, which is neither a person nor a thing. It can imitate human beings, interfere with human cognition, and intelligently invade and destroy network system vulnerabilities for the purpose of designers or decision makers. The main feature of “artificial intelligence hacker” is that it can rely on intelligent algorithms to learn autonomously, find network system code vulnerabilities and strengthen distributed attacks. Some scholars divide artificial intelligence technology into three stages: weak artificial intelligence, strong artificial intelligence and super artificial intelligence. Some scholars even suggest that strong artificial intelligence be given the legal subject status ethically, because strong artificial intelligence algorithm has independent “machine meaning” expression ability and has emotional connection with human beings. Obviously, this way of empowerment violates the principle of “humanism”. Also, the current legal subject includes natural person, legal person, unincorporated organization, “artificial intelligence hacker” does not belong to any kind of subject, abruptly equating the rational expression of legal subjects with the “machine meaning” of artificial intelligence algorithm instructions can easily lead to difficulties in algorithmic justice legal evaluation and civil legal behavior construction of “artificial intelligence hacker” behavior, interfering with our judgment on the nature of “artificial intelligence hacker”. The legal attribute of “artificial intelligence hacker” should be judged based on legal rights and obligations. In essence, “artificial intelligence hacker” is a natural person who uses network media to commit network infringement or crime through artificial intelligence algorithm technology. The core of “artificial intelligence hacker” is a set of mechanism for making decisions through computer code setting, big data operation and automatic judgment of machines. “Artificial intelligence hacker” is not a proper legal subject in terms of responsibility, but has a special legal attribute of “personality tool”. The intelligent attack of “artificial intelligence hacker” is manifested by the automatic execution of the algorithm program, but the design and operation of the program belong to people in real economic life, which also fully conforms to the adjustment scope of indirect infringement in law. For criminal acts or infringement of “artificial intelligence hackers”, the hidden subject and object can be regulated by law should be found through piercing the “veil” of “artificial intelligence hacker”, and the “penetration” method should be used for the effective regulation of illegal behaviors of “artificial intelligence hackers” from the dimensions of ethics, technology and law.

Key words: artificial intelligence hacker; algorithm; personality tool; legal regulation

(责任编辑 胡志平)