

人工智能时代刑法归责的走向

——以过失的归责间隙为中心的讨论

储陈城*

内容摘要:随着全面的人工智能时代即将到来,刑法作为回顾性法律,也需要对未来可能产生的风险进行提前性展望。由于人工智能技术的复杂性、裁量的独立性和行为自控性等特点,使得现实中会出现人工智能因过失导致的法益侵害,存在无法归责的空白地带。对此归责上的间隙,目前理论上形成了“机器人刑法”与“传统刑法修正”两套应对方案。机器人刑法虽然在一定程度上会防止技术发展的萎缩,但是在刑法的主体、刑罚正当化和自由意志等方面备受质疑。而传统刑法修正,虽然会保留刑法的本来面貌,但是会导致出现刑法介入过失犯的范围过宽的局面。应对人工智能时代的风险,社会应当在保障技术发展的长远目标的基础上,限缩过失犯的成立,保持刑法的最后手段性。

关键词:人工智能 过失犯 归责间隙 机器人刑法

一、问题的提出——人工智能领域过失犯中归责的间隙

机械随着人类文明的进程不断地演进。因为电子计算机技术的迅猛发展,任何信息都可以数据化,全球化的互网络已经形成,在这种技术的发展中,需要有法律的存在。作为机械相关的初期的法律规定,中世纪的矿业法,17、18世纪的制粉法等,工业革命时期,蒸汽锅炉相关的立法也开始出现,比如铁路法的制定,其也成为现今责任法的基础。^[1]目前,机器人技术的进步已经日新月异。不受人的直接操作,独自地根据周围的环境,进行目的指向性行动的机器(自控型或者部分自控型机器)正在走向人类的生活。未来随着机器人技术的进步,机器人将会和人类完全一样自主地进行活动。“自控型”是指机器不受来自人类的指示,有判断的裁量性,并且具有学习和适应能力,使人无法预测其具体的行动。^[2]

当经过了机器人作为人类的道具被使用的阶段之后,接近奇点理论的时候,人工智能或者机器人开始脱离人类的控制,^[3]将会引发什么样的问题?如果人类是将人工智能作为支配性工具而辅助性地使用的话,出现归责问题时,基本上只要通过现行法及扩张解释就可以应对。因为,此时人类是作

*安徽大学法学院副教授,早稻田大学法务研究科外国人研究员。

[1]参见エリック・ヒルゲンドルフ:《法と自律型機械——問題概説》,富川雅満 訳,《千葉大学法学論集》2016年第2号。

[2]See Mark McKamey, Legal Technology: Artificial Intelligence and the Future of Law Practice, 22 Appeal: Rev. Current L. & L. Reform 45, 58 (2017).

[3]See Legrand, Pierre, On the Singularity of Law, 47 Harv. Int'l L.J. 517 (2006).

为主体行动,只不过是人工智能作为工具而使用,最终的法律效果(权利和义务)都归属于人类,人类也应该承担法律责任。而与此相对,当人工智能发展到近乎奇点理论的时候,人工智能或者机器人脱离了人类控制,现行的法律规则是否能够全面应对呢?

当然,无论是依据美国生物化学家Isaac Asimov曾经提出的机器人三原则,^[4]抑或是日本学者新保史生教授所提出的机器人使用的八原则,^[5]都绝对禁止故意生产、设计或者使用能够对人类产生法益侵害的机器。比如远程操作型的机器就会产生这样的一般性问题:即因为该机器是远程操作,即便是遭到反击或者逮捕,也不会受伤,因此这种机器可能会被人类恶意使用作为抢劫银行的工具。即便没有恶意使用到那种程度,也可能通过远程操纵而破坏他人财物。还可能会因为网络线路的故障,导致指示传达的滞后,并出现射杀和意图的对象不一致的其他人的结果。^[6]对于设计、生产和使用这种机器人的主体,因为机器只是人类实施犯罪行为的工具,即便该机器人是自控型的,也可以利用和解释间接正犯理论来对具体人进行归责。然而,在人工智能时代,刑法适用中成为问题的是过失犯领域,往往会出现归责的间隙。

以德国刑法学者约尔登教授曾探讨的莱茵河案为例:在浓雾当中,导航向正在与莱茵河流向呈平行线行驶的驾驶员发出指示,让其向右转弯。已经50岁左右的驾驶员听从该指示向右转弯,使得自己和汽车一齐飞入到莱茵河当中,致使驾驶员和同乘的乘客溺水死亡。导航大概是想让驾驶员更容易地改变行驶路线,而在此处并没有横跨莱茵河的大桥存在,要想到达对岸必须要通过船运才可以实现,认识到这一点实际上并不困难。但是因为浓雾,驾驶员没有注意到,甚至都没有踩刹车就跌入到莱茵河当中。^[7]关于此案,当该导航只是普通的电子产品的时候,对于驾驶员和乘客的死亡结果,可以考虑追究导航制造者的刑事责任。其原因是,制造者可以预见到发生这种死亡事故的发生。另外,制造者应该能够预见到,导航在指示道路变更时,在有浓雾等例外状态中,会给驾驶员错误的指示的可能性,这样或许可以通过程序设计来回避错误指示的出现。也可以通过设计程序,让驾驶员在航道前停车。但是,如果该导航具备人工智能的属性,通过设计者和生产者的初始设计,设定了行动目标后,该导航系统摆脱了人类的控制,根据自己的判断进行学习、适应。此时,该智能型导航因为错误导致驾驶人员死亡的案件,又该如何处理呢?^[8]

二、可能的解决路径——德国刑法学界的争议

关于刑法应该如何回应上述归责上的间隙问题,对此,德国学者分成两派观点。第一种是对刑法中已有的概念进行修改、完善,开辟出刑法对机器人适用的可能性。这一派观点的代表性学者是希尔根多夫教授。第二种则是直截了当地否定刑法对机器人适用的可能性,强调刑法的人类中心主义,试图通过对现行刑法理论的解释、修正予以应对。这一派的主流学者是贝克教授和约尔登教授。

[4]第一条:机器人不能对人类产生危害,且不能因为忽略这种危险,而导致危及人类安全;第二条:机器人必须要服从人类所发出的命令,但是,所发出的命令如果违反第一条的话,不在此限;第三条:机器人只要在不违反前两条规定的前提下,必须保护自己。See Pittman, Spencer C., Mwafuilirwa, Mbilike M. Autonomous Vehicles and the Trolley Problem: An Ethical and Liability Conundrum, 88 Okla. B.J. 1719 (2017).

[5]这被称之为“新保试案”:第一,人类优先原则;第二,服从命令原则;第三,保护秘密原则;第四,限制利用原则;第五,安全保护原则;第六,公开透明性原则;第七,个人参加原则;第八,责任原则。参见新保史生:《ロボット法新8原則》,《時の法令》2017年第2·3号。

[6]参见スザンネ・ベック:《グーグル・カー、ソフトウェアエージェント、自律的武器システム—刑法にとっての新たな挑戦?》,根津洗希 訳《千葉大学法学論集》2017年第3·4号。

[7]参见ヤン・C・イエレン:《ロボット工学の刑法的諸観点》,今井康介 訳,《千葉大学法学論集》2016年第2号。

[8]这样的案件已经不再是个案,曾经有过多起机器人错将附近的人类误认为是安全威胁,而通过程序处理,导致人类死亡的案件。See Yueh-Hsuan Weng, Chien-Hsun Chen & Chuen-Tsai Sun, Toward the Human-Robot Co-Existence Society: On Safety Intelligence for Next Generation Robots, 1 INT. J. Soc. ROBOT 267, 273 (2009).

(一)机器人刑法之倡导

行为、归责、责任这样的概念,一直以来都是不可侵入且不可变化的,被认为是无法适用于机器的。即便机器是自主地进行活动,但是因为该活动不属于刑法当中的行为,则对机器进行责任归属,或者认定责任都不具有可行性。^[9]希尔根多夫教授认为,这样的立场是将机器人所惹起的损害,让机器人自身承担责任的可能性,先验地予以完全排除。但是,认定人类和机器人之间的差异是否存在何种理由呢?对于人类所适用的概念用于机器人时,犹豫不前,难道不是人类固执地认为“我们人类是独一无二的存在”所导致的吗。^[10]

一般的语法当中,记号的意义并不是使用者单方所规定的,重要的是概念的接受者该如何理解。换言之,语言的共同体和解释的同一性就是记号的意义所在。^[11]但是,即便是在同一的语言共同体中,也会存在同样的表达出现不同的意思的情形,语言用法不明确的情形并不少见。因此,明确概念的方法还必须要进行摸索。赋予概念明确意思的第一种方法就是注释。所谓注释,是对某个概念的意思进行具体化,对该概念的各个标准进行分析。比如刑法当中将“夺取”这一概念,注释为“侵害他人的占有以及确立新的占有”。^[12]与此相对,第二种方法是定义。定义,不管是在使用不明确的语言的时候,还是在通过定义来赋予语言新的概念的场合,都会使用到。通过定义可以设定占位词,将冗长的表达予以简略化(省略功能)。比如:德国《刑法》第11条第1项第1号当中所规定的内容:“本法律的意义中,1.所谓亲属是指:(a)直系血亲、直系姻亲、配偶、订婚者、兄弟姐妹、兄弟姐妹的配偶以及配偶的兄弟姐妹;婚姻已经不复存在的场合,或者血亲关系、姻亲关系消灭的时候,也是亲属。(b)监护人和被监护之子女。”^[13]

定义,并不是表述意思,而是通过规定来确定语言新的意思。因此,定义没有正确和错误之分,只有相当性(合目的性)程度的差异而已。这样的话,语言的意思就不是不可变化的。也即,语言是在历史的过程当中不断变化的,可以通过定义来改变其意义,可以赋予语言全新的内涵。而且,正是在法律领域,通过新的解释,使得概念适应需求的变化,并使该变化不断地延续。

行为、责任这样的基本概念并非只限于人类。比如民法和刑法上,因为存在法人的概念,因此作为责任主体的人的概念,并非没有扩张的可能性。^[14]既然可以将行为、归责、责任等概念,扩张到法人领域,自然将其也用于机器人的可能性就无法完全被否定掉。现在的语法绝非不可侵的,而是可以变更的。一般的语法在历史的流程中不断地变化,而在专门领域当中的语法,可以通过应用专门语法之目的而创设出来的定义进行修正。

第一,关于机器的行为,希尔根多夫教授认为,行为被理解为通过意思而被控制的举止,至少,机器人在外观上看起来像是自己进行的举动。但是,确实机器人会四肢动作、发出语音,对周边环境施加影响等,但是还不能明确地说这些举止是基于意思进行的控制。这其中的“意思”更可能是对机器进行程序的人或者使用的人的意思,这样的话,就会将机器的举止归属于机器背后的人,而不是机器本身。要提出“机器行为”这一概念,就必须要在机器自身当中找到和人类意思完全相当的意思,这里往往会想到控制机器的程序。最为简单的程序是,按照“遇到什么场合→做什么动作”这样的组合来进行设计,机器人在A状态中,就会做出A动作,在B状态中,就会做出B动作。这样可能会妨碍机器基

[9]国内相关论述,参见马克昌:《刑法中行为论比较研究》,《武汉大学学报》(社会科学版)2001年第2期;杨开湘、蒋凌申:《自动柜员机刑法意义新探——基于对最高人民法院司法解释的分析》,《东方法学》2009年第4期;陈忠林、梅锦:《论人格在定罪中的运用》,《现代法学》2012年第6期等。

[10]参见エリック・ヒルゲンドルフ:《ロボットは有責に行為することができるか?—規範的な基本語彙の機械への転用可能性について—》,伊藤嘉亮 訳,《千葉大学法学論集》2016年第2号。

[11]参见陈波:《语言和社会意义的社会建构论》,《中国社会科学》2014年第10期;黄涛:《走向语言的共同体——对于当前法哲学的语言学转向的一点思考》,《语言战略研究》2017年第5期等等。

[12]参见[日]山中敬一:《刑法各论》,日本成文堂2015年版,第296页。

[13]参见前引[10],エリック・ヒルゲンドルフ文。

[14]参见夏井高人:《アシモフの原則の終焉——ロボット法の可能性》,《法律論叢》2017年第4·5 合併号。

于意思而控制举止的可能性了。鉴于人类原则上也是按照一定的规则,或者根据该规则的指引,来选择行为举止的,这样的话,人类和机器进行对比,将机器基于程序而进行动作,表述为“意思控制”也具有可能性。

但是,这一类比并没有考虑“自由意志的问题”。自由意志当中,以“人类不能(完全地)决定自己的举止,和能够自由决定”为前提,这一命题在刑法当中尤为重要。当然,在实务当中,对于意思自由的结论是保留的。法官只关注否定意思控制的要件,比如无法抵抗的强制。这样的话,机器人的行为也可以作同样的评价。也即,机器也有行为能力,只需要在无法抵抗的强制这样的因素存在时,否定行为即可。^[15]

第二,关于对机器的归责。希尔根多夫教授质问,只有自然人才能成为责任主体吗?民法上的企业能够承担责任已经被予以认可了,另外,考虑到很多国家甚至已经认可企业可以承担刑事责任,^[16]那么自然人以外的主体承担责任是可能的。原理上,肯定人类以外的非人类主体承担责任是法律的自由范畴。因此,至少在概念上,肯定机器人的责任并没有什么问题。另一方面,承认责任的法律效果会怎么样,则成为问题。机器人因为没有财产,所以不能判决机器人承担损害赔偿。但是,这一问题可以通过如下途径解决,比如对于机器人的使用者,为了使得机器人能够附着损害赔偿的义务,需要强制性购买保险这样的法律义务。或者法律规定,所使用的每个机器人的时候,必须向银行账户中转入一定的存款,万一发生机器人损害赔偿的时候,可以从中支取。^[17]

第三,关于机器的责任能力,希尔根多夫教授认为,要承认刑法上的追责,不仅需要行为能力,还必须以责任能力为前提。责任一般被解释为“非难可能性”,1952年德国联邦最高法院的判例当中就意思自由作出如下判解:“通过责任这样的无价值评价,行为者可以采取遵从法律的举止,尽管能够选择适法的决定,却没有采取合法的行为,转而决定选择不法,就能够被非难。责任非难的内在根据是,人只要道德已经成熟,只要这种自由且道德地决定自己行为的能力,没有因为病理原因暂时性或者长时间地受阻甚至丧失,那么就视为具有自由、有责且道德地进行决定的能力,因此可以选择适法行为,抵抗不法行为,使得自己的举止作为法律所允许的合规范行为,回避法律所禁止的行为。”^[18]

因此,德国联邦最高法院就意思的自由持如下立场:“人只要没有被病理所困扰,就可以在合法与非法、善与恶之间,自由地抉择。”这种意思自由是责任刑法的本质性要素,在现今德国刑法领域广受支持。基于这样的背景,初见之下,似乎无法肯定机器人的责任。但是,21世纪的脑科学研究,对意思的自由产生了质疑,法学又恢复到将意思的自由视为是一种拟制的立场。^[19]如果意思自由只是一种拟制的话,那么这种拟制不就有扩张适用到机器人身上的可能性了吗。拟制不是反映现实,有的时候甚至是和现实相反。拟制是基于合目的性这样的观点而被提出的,所以将机器人的自由意志作为一种拟制予以引入的话,也只有满足实际生活的必要性时,才具有可能。但是在目前,这种必要性还无法获得认同。和人类的处罚相类比,考虑对机器人进行“处罚”,比起法学来说,更像是科幻的主题。要想变更机器人的举止,虽然最简单的方法是更改程序,但是,必须因为有自行学习的系统,而使得程序更改无法奏效的时候,就需要讨论能够使得机器举止发生变化的“危险补充”。但是,为了应对这种“危险补充”,并不需要多重意义的刑罚。

然而在法学上,实际上意思自由并没有那么重要的意义。行为者的责任,只要满足:1.责任能力;2.特别责任要素(比如德国《刑法》第315条c所规定的道路交通危险罪中的鲁莽要素);^[20]3.责任过失、

[15]参见前引[10],エリック・ヒルゲンドルフ文。

[16]参见樋口亮介:《法人処罰と刑法理論》東京大学出版会2009年版,第1—137頁。

[17]参见前引[10],エリック・ヒルゲンドルフ文。

[18]Vgl. Andreas Matthias, Automaten als Träger von Rechten. Plädoyer für eine Gesetzänderung, 2008.

[19]国内相关研究,参见谢望原、邹兵:《论期待可能性之判断》,《法学家》2008年第3期;[德]汉斯-海因里希·耶舍克:《德国与奥地利刑法中责任概念的流变》,陈金林译,《刑事法评论》2011年第1辑。

[20]参见李圣杰等译:《德国刑法典》,台湾元照出版公司2017年版,第398頁。