

故意;4.不法的意识;5.免责事由不存在,就可以予以肯定。这其中完全没有涉及到意思的自由问题。这可以说是德国刑法中的实际主义(或者即物主义),将意思自由这种形而上学的问题置于考虑的范围之外。因此,只要机器具备上述5点要件的话,就可能认定机器人的责任。这样来看的话,可能会有人批判认为,机器实际上并不具有理解所为不法的能力,不可能进行实际的无谋行为。但是,对于这种批判,即便是人类,这种所为的“实际上”到底是什么意义,并不明确。另外,适用上述词汇的时候,只是被单纯的拟制,那是基于实务上的理由而被引入的,并不需要实际地阐明内在的、心理的状态。

总而言之,希尔根多夫教授认为行为、责任等概念可以用于到自控型的机器身上。这些概念只要有必要运用到机器人领域,就可以在法律的框架内,基于扩张解释来对这些概念进行必要的修正。虽然会出现形而上学的难题,但是意思自由这样的问题,在实务上几乎无需有所顾虑,而且人类的自由,对于国家来说也只是必要的拟制而已。基于上述的原因,对于机器人可以认定行为、归责和责任。

(二)传统刑法修正对策

约尔登教授是反对机器人刑法的典型代表,他认为关于用何种方法来判断对机器人的决定是否应承担责任的问题,可以通过承认法律中的自由行为起到何种作用这一观点来解释。这一讨论虽然也在脑科学和哲学中被提出,但是重要的是,要区别从法律的角度来看的情形、从生活实践中所见的情形、以及将来所见的情形。其认为,自由行为为法律所认定,是被告人从国家那里获得作为主体的资格。因此,行为尚且不自由的情形下,比如法律视野下的儿童那样,只能作为国家的客体予以对待。约尔登认为人类的尊严、责任、有责性等概念如果没有自由的概念的话,就没有任何意义。人类是基于自由的决定而遵守规则的。而机器人的自由观念和人的自由是无法相容的。^[21]

那么通过传统的过失犯论,应对自控型的机器人所产生的归责问题,会产生很多难题。因为要想成立过失犯,必须要有对注意义务的违反和对危险的预见可能性。

而现在,机器人工程学当中,注意义务划定的参考标准几乎不存在。界定注意义务的大致标准是作为工业标准的ISO标准,或者是DIN标准。在使用这一标准的时候,必须要注意一些事项。^[22]

第一,这些标准都是非国家组织所制定的民间标准,未必和刑法上的注意义务标准相一致。因为刑法也保护部分的社会道德,和民间标准所要保护的利益并不完全一致。因此,即便是遵循民间标准,也有可能构成过失犯。刑法和民间标准的保护目的之不同的一个例子是,刑法中所谓的“伤害”概念。按照DIN标准,和机器人一起作业的人受伤的话,如果该受伤没有出现出血的情况,那么应该忍受身体侵害的危险,如果出现出血的话,则不需忍受该危险,也即违反了该标准。但是,刑法中的伤害的概念不以出血为必要,只以身体是否有重大伤害作为区分标准。

第二,因为民间标准是非国家组织所制定的,所以其制定的标准缺乏民主主义的正当性。这是因为,在制定标准的时候,参与制定的委员的选任方法、标准的起草以及最终采纳的方法等,都不甚明确;而且,甚至和该标准相关的领域专家都完全没有参与。

尽管有上述的疑问之处,但是实务界仍然对民间标准有着很大的依赖性。正是由于民间标准有着这种官方的印象,很多学者和制造商认为,只要遵守民间标准,就不会有法律上的问题。如果,开发者所设计的机器人对他人造成伤害,该开发者确信“只要遵从民间标准的就没有问题”的话,虽然该确信对于认定注意义务的违反没有影响,但是有可能会否定他们的违法性意识。

过失犯成立的第二个要件,需要要求危险的预见可能性。在利用自控型的机器人的时候,总觉得该机器人可能会造成某种损害,这很容易想象得到。比如,自动驾驶的汽车^[23]、军事机器人^[24]等都是比

[21] Jan C. Joerden, *Strafrechtliche Perspektiven der Robotik*, in: Eric Hilgendorf/Jan-Philipp Günther (Hrsg.), *Robotik und Gesetzgebung (Robotik und Recht Bd.2)*, 2013, S.195ff.

[22] 参见スザンネ・ベック:《過失解釈における民間規格の意義》,谷井悟司 訳,《比較法雑誌》2016年第2号。

[23] 参见小林正啓:《自動運転車の実現に向けた法制度上の課題》,《情報管理》2017年第4号。

[24] 参见岩本誠吾:《致死性自律型ロボット(LARs)の国際法規制をめぐる新動向》,《産大法学》2014年第3・4号。

较好的例证。但是,对具体的结果预见、因果经过的基本部分等的预见,几乎不可能。^[25]这是因为,自控型机器人有学习和适应能力,使人无法预测机器人的具体行动。刑法上所要求的预见可能性,需要对该具体事实以及和因果关系相关的具体的预见可能性,那么这就无法对机器人的设计者等肯定其预见可能性。另一方面,如果将预见可能性理解为“大概可能会对人造成危险”这样抽象的预见可能性的话,显然会使得过失犯的范围变得过于宽广了。

通过传统的过失犯理论来解决有上述困难之外,自控型机器人领域的刑事案件的问题之复杂性,还远不止此。这是因为,机器人大多数都是在和人类打交道。也即,机器人的过失行为的原因,是因为设定机器人学习和适应的程序设计者的行为导致的吗?还是因为制造和贩卖的人导致的呢?抑或是进行训练和调整的利用者的行为呢?这调查起来非常困难,认定谁拥有多大程度的过失,可以说是几乎不可能完成的任务。

作为这种情况下的责任分配的方案,能够考虑的大概有以下四种方法:1.相关主体当中的一人,比如利用者独立承担责任。2.相关主体当中,只由能够通过证据证明存在过失的人承担责任。3.和机器人相关的所有的人,理解为一种新的法律人格,由这种新的法律人格来承担责任。4.可以将机器人视为是对社会来说重要的存在,因此由机器人所产生的危险性、责任性应由社会全体来分担。这些解决方案的背景是,因机器人的利用而受益者,对于机器人的过失而应该被课以金钱或者刑法上的制裁,机器的行为应该归责于受益者。即使这种机器的行为超过了正常人的预想。但是,在机器人的体系当中,对过失行为、损害赔偿进行追责不是问题的所有。刑法本是只有回顾性地对犯罪进行非难的机能,所以并不具有对这种现代性风险进行概括性控制的导向。

针对上述复杂问题,能够想到的解决对策之一,是对过失概念的修正。以下是这种修正过失犯论的核心。

这种修正的过失犯论当中,最早危险的抽象预见可能性,不能成为刑法中责任的标准。这是因为,汽车、核能发电等,虽然统计学显示具有危险性,但是广泛地被社会所接受。^[26]因此,要侧重于设定注意义务的基准,这种注意义务的基准不应该用民间标准或者刑法以外的规定。如前所述,民间标准和刑法以外的规定,衡量比较某种行为的优缺点,这样的规范未必和刑法规范的保护目的不一致。因此,这种利益衡量无法直接用于划定注意义务的标准。机器人的有用性和机器人所拥有的危险性衡量以后的结果,尚能够为社会所接受的话,就应该在“可容许的危险”的范围内进行。

在过失责任分配的时候,重要的要素是“信赖原则”。根据信赖原则的规定,所有的人原则上都可以信赖别人进行合法行为。^[27]在机器人的体系当中,这就意味着,比如自控型机器人侵害人类法益的时候,只要设计者、贩卖者、利用者相互之间可以期待适法行为的话,任何人都不承担责任,只有没有实施被期待的合法行为的人,承担侵害结果的责任。这样来看的话,责任分配的结果要么是谁都免不了责任,抑或是谁也无法被证明有责任,那么谁都无法被处罚。

但是显然这又会产生新的归责空白,为了填补上述责任上的空白地带,也有观点指出让利用者完全承担责任。此时,除了程序设计、制造时存在明显的构造上的缺陷,否则我们让机器人代为实施的行为造成的损害,均由利用者来承担责任。但是,这样的归责结果很难被接受。因为我们将判断转移给自控型的机器人,让机器人代为行使,就是让人类从判断的负担中解放出来。也即,是为了减轻人类的负担。^[28]如果即便将判断转移,而机器人的行为必须由利用者来承担责任的话,那么结果是人类的负担并没有予以实质地减轻。比如长距离的运输当中,利用机器人来进行自动驾驶的时候,虽然是将路线和速

[25]参见小田直树:《過失の問い方について》,《神戸法学雑誌》2013年第2号。

[26]参见古川伸彦:《刑事過失論序説——過失犯における注意義務の内容》,成文堂2007年版,第231—234页。

[27]参见西原春夫:《交通事故と信頼の原則》,成文堂1969年版,第13—14页。

[28]See Scherer, Matthew U. Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies, 29 Harv. J. L. & Tech. 353 (2015—2016); 参见金岡京子:《自動運転と民事責任:自動運転と民事責任をめぐるドイツの状況》,《ジュリスト》2016第1501号。

度调整等都交给机器人代为决定,但是如果因机器人判断失误而导致出现事故,事故责任由利用者来承担。这样一来,利用者必须不休息,一直注视着机器人的驾驶,这显然需要利用者付出和自己驾驶同等程度的注意义务。那么,这种归责方式显然和使用自控型机器人的意义本末倒置了。这样的话,与其为了避免责任上的空白地带而采取上述不当归责方法,还不如让社会容忍这种归责空白地带。^[29]

三、人工智能归责间隙的刑法应对

(一)对机器人刑法的回应

机器在运转的时候发生的侵害法益之行为,能够对机器追究刑法上的责任吗?或者说,机器应该承担责任吗?当计算机出现崩盘的时候,导致数据丧失,计算机的所有者基于愤怒,可能会对计算机施加私人的“处罚”,但是国家能够基于刑事程序,对计算机施以国家层面的刑罚吗?希尔根多夫教授将刑法中的行为、归责、责任等概念进行分析,认为概念存在的重要性,需要考虑语言共同体当中的合目的性,得出也可将时至今日只能适用于人类的刑法概念体系,扩张到人类以外的实体当中的结论,并进而创造出机器人刑法的概念。确实,在不久的将来,机器极有可能在没有人介入的情况下,进行决断实施某种行为。此时,根据机器所遵守的原则,可能并不是在制造过程当中所确立的,而是在其实际运转过程中发生变化的原则,并且机器会根据自身需要进行变更所遵守的原则。也即,人类开始对机器失去控制时,机器在运转过程中所产生的法益侵害结果,到底由谁来承担,就成了刑法上的空白地带,那么机器人刑法似乎成了必然结论。^[30]如果不让人类来为机器的行为买单,机器人真的有可能为自己行为承担责任的可能吗?

根据希尔根多夫教授的观点,创设机器人刑法,可以通过对现行概念进行修正,扩张适用到机器人领域。通过对自然人设定经济义务,让机器人在被追究刑事责任时有财产可供(罚金)执行。这种观点并非没有任何意义。这一观点,一方面使得遭受机器人过失侵害的被害者有获得救济的途径,能够缓解因机器人过失侵害无法受偿所导致的社会恐慌和不满。另一方面,通过对机器人进行处罚,虽然实际上是由自然人实际负担经济责任,但是却实现了罪“刑”分离。也即虽然设计、生产、销售和使用机器人的自然人可能会因为机器人的过失侵害,实质背负罚金刑,但是因为不是犯罪主体,所以不会因为机器人的过失行为,而被认定为犯罪。这会极大减轻自然人设计、生产、销售和使用机器人的刑法成本,在过失犯领域,将自然人本应承担的刑事责任实质转化成为了只承担民事赔偿责任。考虑到刑事责任成本过高,会导致对技术创新造成萎缩效果,故这一观点能够保障人工智能技术的发展。

尽管如此,批判机器人刑法观点的声音仍不绝于耳。创设机器人刑法的概念,第一个反对的声音来自于刑法是只以人作为规制对象的传统刑法观念。这一观念也被称之为“刑法人类中心主义”。^[31]基于这一观念,国家发动刑罚的对象则必须是人类。确实,刑法典当中对此并没有明确规定,但是大部分的规定都内含着“实施……的人”这样的用语,从中可以看出刑法所针对的对象也是人类。

诚然,纵观人类刑法的历史,直至启蒙时期,在刑法秩序当中确实存在着很多非人类的实体,也即刑法曾经经历过并非完全以人类为中心进行建构的时期。所谓的非人类的实体,包括动物或者物体。比如曾经有刑法规定可以对作为犯罪发生场所的建筑物,或者行为人所住过的房屋进行销毁,即所谓的废墟化。另外,还存在一种口袋之刑,即将罪犯和动物一起装入袋子,沉入河中溺死的刑罚。^[32]甚

[29]スザンネ・ベック:《インテリジェント・エージェントと刑法—過失、答責分配、電子の人格》,根津洗希 訳,《千葉大学法学論集》2017年第3・4号。

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

[31]参见姜敏:《系统论视角下刑法修正案犯罪化限制及其根据》,《比较法研究》2017年第3期。

[32]Eric Hilgendorf, Können Roboter schuldhaft handeln, in: Susanne Beck (Hrsg.), Jenseits von Mensch und Maschine, Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs, 2012, S. 119ff., S. 125.

至,还存在对物或者动物进行裁判的刑事制度。但是,这种非完全以人类为中心主义的刑法,已经在启蒙时期到来之后基本消失。其后,刑法迎来了专门以人类为中心的时代。这里所言的人类,是指作为拥有自由地进行行为的能力,且有理性的先天能力的个体的人类。换言之,即作为被课以需要遵守法律义务的国家市民的人类。这样的话,动物或者其他物体已经不再能够成为刑法规制的对象,这是因为,动物或者其他物体不具有自由和理性的能力,而且,动物也不是国家中的市民组成。刑法进入人类为中心的时代,标志着刑法文明的巨大进步。

针对机器人刑法的第二个质疑是,目的能够将所有的手段都正当化吗?即对机器施以刑罚制裁,无法获得承认。即便现代刑法又返回非完全以人类为中心的时代,刑法可以规制机器人的行为,那么对于机器该如何施以刑罚呢?

对此,有的学者认为,可以对机器人进行再程序化或者重新设计,即认为机器人和人类不同,机器人的制造者可以将机器人的程序信息全部抹消,将其恢复到和最初出厂时完全一样的状态,然后,机器人就可以重新按照最初的程序来决定和定义自己的行为。因此提议将机器予以再程序化融入到刑法的刑罚制度体系中去。^[33]这种刑罚的设置和当前的刑罚体系格格不入。以德国刑罚体系为例,其主刑包括自由刑和罚金刑,而再程序化的处罚和这些刑罚无法相适应。通过对机器人予以再程序化,显然无法达到自由刑那样的限制行为自由的效果;同样,机器人没有财产权,通过再程序化也无法强制机器人支付罚金。

那么是否可以将再程序化单独列为一种独立的刑罚呢?作为再程序化的对象是机器的“身体”“生命”或者“人格性”。如果将机器人作为和人类等同的人格主体的话,则针对这三者的处罚,都是为现代刑法所排除的刑罚手段。针对机器人的再程序化这样的刑罚设置建议,实际上相当于针对人类的生命刑及身体刑。换言之,机器人刑法是以将机器人视为与人类拥有同等程度的自控性人格为前提。对机器人实施再程序化,就相当于对人类实施脑白质切除术,以强行改变人格的不可逆性一样,对被承认具有人格的自控型机器人,也就不能实施强制变更其内置程序。再程序化应该和人类死刑一样被禁止。

反对机器人刑法设立的第三个核心论点来自机器人刑法所主张的自由意志。诚然,自然人的自由意志也不过是一种虚拟,这种观点为很多学者所认同并主张。比如有学者提出“自由意志和他行为的可能性,完全无法证明”,因此认为自由意志理解为规范的要求。^[34]井田良教授认为,作为确定责任的有无和程度的基准的自由和可能性,并不是经验性事实,而是规范的要求或者假设。^[35]高桥则夫教授则主张:“责任并不是作为实体的责任,甚至可以说一种社会的虚构。”“没有这种社会的虚构的话,社会就无法存续,结果、责任和刑罚,如果没有将自由意思假设作为前提的话,也会成为无法存续的制度。”^[36]但是,即便自然人的自由意志是基于规范的要求而作的法律之虚拟,也不能直接肯定机器人的意思之自由也是法律之虚拟。正如佐伯千仞教授所指出的那样,刑法上的责任和意思自由即便是一种拟制,“那也是在社会生活的现实中有基础的,绝不是被架空的”。^[37]自然人由于存在一定的社会基础,才能够允许将自由意志进行假定的话,那么机器人的责任和自由意志的假定,也必须要求有同样的社会基础。

但是,到底容许自然人自由意志的社会基础是什么呢?增田丰教授作出这样的批判:“通过作为国家拟制的自由,对行为人归属责任的非难,对行为人课以刑罚的构想,大概不会得到我们相互之间主观的承认,也无法获得正当化。”认定对行为人进行责任归属,必须要具备“为追究责任的相互主观的、交流的条件”。也即,“作为责任非难基础的自由,即便是习惯性的存在,社会的、文化的构

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

[34] 参见内藤谦:《刑法讲义总论(下) I》,有斐阁1991年版,第786页。

[35] 参见井田良:《讲义刑法学·总论》,有斐阁2008年版,第358页。

[36] 高桥则夫:《刑法总论(第3版)》,成文堂2016年版,第334页。

[37] 佐伯千仞:《刑法における期待可能性思想(下卷)》,有斐阁1949年版,第614页。