



独角兽  
法学精品

# 人工智能时代的 刑法观

刘宪权  
——著

上海人民出版社

# 人工智能时代的 刑法观

刘宪权——著



上海人民出版社

## 图书在版编目(CIP)数据

人工智能时代的刑法观/刘宪权著. —上海:上海人民出版社, 2019

ISBN 978-7-208-15981-5

I. ①人… II. ①刘… III. ①人工智能-刑法-研究  
IV. ①D914.04

中国版本图书馆 CIP 数据核字(2019)第 145570 号

责任编辑 夏红梅

封面设计 孙康

## 人工智能时代的刑法观

刘宪权 著

出版 上海人民出版社  
(200001 上海福建中路 193 号)  
发行 上海人民出版社发行中心  
印刷 常熟市新骅印刷有限公司  
开本 720×1000 1/16  
印张 14  
插页 4  
字数 252,000  
版次 2019 年 7 月第 1 版  
印次 2019 年 7 月第 1 次印刷  
ISBN 978-7-208-15981-5/D·3454  
定价 58.00 元

“品群学”系列书籍

《代...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

《...》

# 前 言

自人工智能概念于1956年首次被提出以来,人工智能技术经历了六十余年的发展,更是在近几年迎来了质的飞跃,成为继蒸汽机、发电机与计算机后第四次工业革命的先锋力量。在超级计算、大数据、互联网、类脑智能技术等新理论与新技术的快速发展,以及社会经济发展强烈需求的推动下,我们已经迎来了人工智能发展应用的新时代——人工智能时代。人工智能已经从科幻走进了现实,并成为全球瞩目的科技焦点。时下,世界各国均已认识到人工智能是未来各国和地区之间的主战场,因而纷纷部署人工智能发展战略,以期占领人工智能发展的制高点。我国政府高度重视人工智能的发展,自2015年以来,陆续出台的《中国制造2025》《“十三五”国家科技创新规划》《“互联网+”人工智能三年行动实施方案》以及《“十三五”国家战略性新兴产业发展规划》等文件都为人工智能的发展保驾护航并指明了方向,将人工智能的研发与应用推向高潮。

人工智能技术是一把“双刃剑”,在造福人类的同时也蕴藏着无限的风险,用之得当,则给人类带来巨大福祉,用之不当,则可能带来毁灭性的打击。国务院《新一代人工智能发展规划》指出:“人工智能是影响面广的颠覆性技术,可能带来改变就业结果、冲击法律与社会伦理、侵犯个人隐私、挑战国际关系准则等问题,将对政府管理、经济安全和社会稳定乃至全球治理产生深远影响。在大力发展人工智能的同时,必须高度重视可能带来的安全风险挑战。”2018年10月31日,中共中央政治局就人工智能发展现状和趋势举行第九次集体学习,习近平总书记在主持学习时不仅指出要深刻认识加快发展新一代人工智能的重大意义,要推动我国新一代人工智能健康发展,还强调要加强人工智能发展潜在风险的研判和防范,建立健全保障人工智能健康发展的法律法规、制度体系和伦理道德,维护人民利益和国家安全,确保人工智能安全、可靠、可控。由此可见,完善相关法律法规,实现对人工智能技术的风险防控是我国目前发展人工智能技术的重要战略任务。

人工智能所蕴藏的风险作为现代技术风险具有可控性,刑法作为重要的风险控制手段之一,面对人工智能技术的发展所可能带来的风险,应当作出理性的

应对,既不能过度反应,将人工智能技术视为“洪水猛兽”,禁止或阻碍其发展,也不应对其漠然置之,忽视人工智能技术所可能带来的风险。我们理应做到未雨绸缪,积极布局,努力探索人工智能时代涉人工智能犯罪的刑法规制路径,为防控人工智能时代的刑事风险、发挥人工智能技术的最大价值和促进社会的稳定发展出谋划策。

本书从人工智能时代刑事风险的分析出发,阐释了传统刑法理论在新时代所面临的巨大挑战,进而提出了刑法应对人工智能犯罪的基本理念与具体路径。全书共包含了两大部分的内容,第一编主要研究了人工智能时代下所面临的刑事风险以及刑法基本理论的应对策略。犯罪行为因人工智能因素的介入而呈现出新的特征,体现为传统犯罪行为危害性的“量变”、新型犯罪形式的产生等,甚至还可能出现人工智能产品脱离人类控制而实施严重危害社会行为的情形。这些变化对刑法理论提出了巨大的挑战,因此,本部分重点探讨了在树立前瞻性刑法理念的基础上如何对人工智能时代刑事责任的归责路径、刑事责任主体以及刑罚体系予以重构等问题。第二编主要分析了人工智能在一些具体应用领域中所体现的刑事风险以及刑法的具体应对措施。人工智能技术的蓬勃发展,使得其已经在各行各业中大展身手。例如,自动驾驶汽车已经上路,人工智能机器人进行文学创作已成为现实,以达芬奇手术机器人为代表的人工智能医疗机器人已经开始普及,以人工智能交易为代表的一系列人工智能在金融领域的应用系统相继问世等,不一而足。在不同的应用领域,因人工智能技术发展及应用程度的差异,其刑事风险的体现也各不相同,本部分根据各领域人工智能技术的应用现状,在分析其具体刑事风险的基础上,对各领域所涉人工智能犯罪的具体归责路径进行了研究。

有人认为,虽然人工智能技术确实可能给人类社会带来巨大的风险,但刑法保障作为“最后一道屏障”,没有必要急于出手,完全可以等相关风险出现后再进行规制也为时不晚,刑法研究如此超前并非必需,而且还可能阻碍人工智能技术的发展。殊不知,历史已经不止一次地证明,科技的发展乃至时代的更迭往往能够超越人类的想象。人工智能技术呈爆炸式发展,是法律发展速度所不能企及的。当风险转变成现实危险后才考虑刑法规制问题将使人类陷入被动的局面。安全是人工智能时代法的核心价值,发展人工智能技术必须将防范风险放在首要位置。刑法作为人工智能安全保障的后盾,应当顺应时代的潮流,提前对人工智能时代的刑事风险进行预估,探索合理的风险应对方法,并提前明晰人工智能时代的刑事责任的承担路径。我们刑法学者作为引领刑事立法与刑事司法发展的人,应当树立前瞻性的刑法理念,勇担这一时代使命,而不是冷眼旁观、无所作为。“凡事预则立,不预则废”,只有做好充分的风险防控准备,防患于未然,方可避免在危险来临时“患至呼天”。

# 目 录

## CONTENTS

前言 | 1

### 上编 人工智能时代的刑法变迁

第一章 人工智能及其刑事风险 | 3

第一节 人工智能的概念及其发展 | 3

一、人工智能的概念与基本类型 | 3

二、人工智能的发展现状与未来 | 6

三、人工智能时代的技术风险 | 9

第二节 人工智能时代的刑事风险 | 15

一、人工智能时代犯罪特征与形势的转变 | 16

二、弱人工智能时代的刑事风险 | 19

三、强人工智能时代的刑事风险 | 21

第二章 人工智能时代的刑法挑战与应对策略 | 24

第一节 人工智能时代刑法规制的现状与不足 | 24

一、现行刑法能够规制的涉人工智能犯罪 | 24

二、现行刑法规制不足的涉人工智能犯罪 | 26

三、现行刑法无法规制的涉人工智能犯罪 | 27

第二节 刑法应对人工智能犯罪的基本理念与路径 | 29

一、刑法应对人工智能犯罪的新理念 | 29

二、刑法应对人工智能犯罪的基本路径 | 35

第三章 人工智能时代刑事责任的演变		40
第一节 普通机器人时代的刑事责任		40
一、普通机器人的基本特性		40
二、涉普通机器人的犯罪类型		42
三、涉普通机器人犯罪的认定和刑事责任承担		42
第二节 弱人工智能时代的刑事责任		44
一、弱人工智能技术为人类带来福祉		44
二、涉弱人工智能犯罪的基本类型		45
三、涉弱人工智能犯罪的刑事责任		49
第三节 强人工智能时代的刑事责任		57
一、强人工智能机器人的特性及犯罪类型		58
二、刑法规制涉强人工智能犯罪的必要性		59
三、强人工智能时代刑事责任的演变及归责路径		61
第四节 不同时代人工智能研发者与使用者刑事责任的变化		69
一、涉人工智能犯罪中研发者主观罪过认定的变化		69
二、涉人工智能犯罪中使用者注意义务的变化		82
三、涉人工智能共同犯罪刑事责任的认定与分配		86
第四章 人工智能时代刑事责任主体的重构		91
第一节 人工智能机器人的道德伦理与法律属性		91
一、人工智能机器人的特征		93
二、机器人道德应当被赋予和认可		95
三、人工智能机器人与自然人的关系		97
四、人工智能机器人的法律属性		98
第二节 人工智能机器人纳入刑事责任主体范畴之争		105
一、人工智能时代刑事责任主体的基本内涵		106
二、人工智能时代刑事责任主体范围之变化		107
三、赋予人工智能机器人刑事责任主体地位的正反观点之争		109
第三节 应赋予强人工智能机器人刑事责任主体地位		114
一、否定强人工智能机器人刑事责任主体地位因由之缺陷		114
二、强人工智能机器人应具有刑事责任主体地位		120

<b>第五章 人工智能时代我国刑罚体系的重构</b>		123
<b>第一节 人工智能时代现行刑罚体系的局限性</b>		123
一、我国现行刑罚体系及其特征		123
二、现行刑罚体系与人工智能机器人的不适应性		124
<b>第二节 人工智能时代我国刑罚体系重构之现实性</b>		126
一、人工智能时代我国刑罚体系重构的必要性		126
二、人工智能时代我国刑罚体系重构的可能性		130
<b>第三节 人工智能时代刑罚体系重构的具体设想</b>		133
一、人工智能时代刑罚体系重构之基本原则		134
二、人工智能机器人犯罪具体刑罚方式的探索		134

## **下编 人工智能具体应用领域的刑事风险与刑法应对**

<b>第六章 自动驾驶技术的刑事风险与刑法应对</b>		139
<b>第一节 自动驾驶技术的刑事风险与刑法应对困境</b>		139
一、自动驾驶技术发展现状及趋势		139
二、自动驾驶技术的刑事风险		142
三、自动驾驶技术刑事风险的刑法规制困境		145
<b>第二节 自动驾驶技术刑事风险的刑法应对</b>		147
一、自动驾驶汽车交通事故刑事责任分析		147
二、滥用自动驾驶技术的刑事责任问题		155
<b>第七章 人工智能所涉知识产权犯罪的刑法应对</b>		158
<b>第一节 人工智能生成物及刑法保护</b>		158
一、人工智能生成物之性质争议		159
二、人工智能生成物应属著作权法所保护的作品		160
三、人工智能生成物应属著作权犯罪的对象		163
<b>第二节 人工智能生成物刑法保护的具体措施</b>		166
一、侵犯人工智能生成物承载的著作财产权可构成著作权犯罪		166
二、对人工智能生成物的刑法保护应保持合理限度		167



<b>第八章 人工智能在证券期货市场应用的刑事风险与刑法应对</b>	<b>173</b>
<b>第一节 人工智能在证券期货市场的应用及刑事风险</b>	<b>173</b>
一、人工智能在证券期货市场的应用及其优势	174
二、人工智能在证券期货市场应用的刑事风险	176
<b>第二节 证券期货市场人工智能应用刑事风险的刑法应对</b>	<b>181</b>
一、证券期货市场人工智能应用刑事风险的刑法应对理念	181
二、证券期货市场人工智能应用刑事风险的刑法应对措施	182
<b>第九章 人工智能在支付领域应用的刑事风险与刑法应对</b>	<b>188</b>
<b>第一节 人工智能在支付领域的应用及刑事风险</b>	<b>188</b>
一、人工智能与新型支付方式	188
二、人工智能在支付领域应用的刑事风险	190
<b>第二节 支付领域人工智能应用刑事风险的刑法应对</b>	<b>193</b>
一、人工智能机器人作为侵财犯罪对象的刑法应对	193
二、强人工智能机器人作为侵财犯罪主体的刑法应对	196
<b>第十章 人工智能在医疗领域应用的刑事风险与刑法应对</b>	<b>198</b>
<b>第一节 人工智能在医疗领域的应用及刑事风险</b>	<b>198</b>
一、人工智能在医疗领域的应用与发展	198
二、人工智能医疗技术的刑事风险	204
<b>第二节 医疗领域人工智能应用刑事风险的刑法应对</b>	<b>207</b>
一、人工智能医疗技术滥用风险的刑法应对	207
二、人工智能医疗机器人引发医疗事故风险的刑法应对	210

## 【第一章】

### 上 编

# 人工智能时代的刑法变迁

无论人们欢欣抑或踌躇,人工智能时代正悄然而坚定地走来,人类即将甚至已经开始进入一个全新的时代——人工智能时代。2015年5月人工智能系统AlphaGo以总比分4比1战胜世界围棋冠军李昌镐,2017年5月AlphaGo与棋手柯洁展开围棋人机对决,又连胜三把,战胜人类,以及2017年10月AlphaGo Zero仅通过三天学习即战胜AlphaGo Zero,关于人工智能的话题就越来越流行,人工智能、深度学习、大数据等词不断地在我们的身边出现。人们在対人工智能充满好奇的同时,人工智能也不断地深入人类生活的各个层面,如医疗、交通、金融、教育等各行业与领域。人工智能在给人类带来便利与益处的同时,其技术风险、安全风险等各种风险带给人类的威胁与危害也日益增加和显现。

## 第一节 人工智能的概念及其发展

要准确地分析人工智能技术可能带来的社会风险,并合理地引导与规范人工智能技术的发展,首先必须明确人工智能技术的内涵与外延,并客观地分析人工智能技术存在的风险。

### 一、人工智能的概念与基本类型

人工智能的概念早在20世纪50年代就已应运而生,1956年麦卡锡、明斯基等科学家在美国达特茅斯学院开会时探讨了“如何制造出像拟人的智能”,并在会议上提出了“人工智能”(Artificial Intelligence,简称AI)这一概念,标志着人工智能学科的诞生。人工智能技术在这半个多世纪的发展中,其内涵也在相应地拓展。人工智能概念提出当时,人们将其定义为“制造智能机器科学与工程”,发展人工智能技术的目的在于提供得到与智能,或即让机器能够像人一样思考。时至今日,人工智能的内涵已经大大扩展,成为一门交叉学科。人工智能



## 【 第一章 】

# 人工智能及其刑事风险

无论人们欢欣抑或踟蹰,人工智能时代正悄悄向我们走来,人类即将甚至已经开始进入一个全新的时代——人工智能时代。自2016年3月人工智能系统AlphaGo以总比分4比1战胜世界围棋冠军李世石,2017年5月AlphaGo与棋手柯洁展开围棋人机对决,又连胜三把,战胜人类,以及2017年10月AlphaGo Zero仅通过三天学习即战胜AlphaGo后,关于人工智能的话题就越来越流行,人工智能、深度学习、大数据等词不断地在我们的身边出现。人们在人工智能充满好奇的同时,人工智能也不断地融入人类生活的各个方面,如医疗、交通、金融、教育等各行业与领域。人工智能在给人类带来便利与福利的同时,其技术风险、不法风险等各种风险带给人类的威胁与危害也无法避免和忽视。

### 第一节 人工智能的概念及其发展

要准确地分析人工智能技术可能带来的社会风险,并合理地引导与规制人工智能技术的发展,首先必须明确人工智能技术的内涵与特征,并客观地分析人工智能技术存在的风险。

#### 一、人工智能的概念与基本类型

人工智能的概念早在20世纪50年代就已经被提出,1956年麦卡锡、明斯基等科学家在美国达特茅斯学院开会时研讨了“如何用机器模拟人的智能”,并在会议上提出了“人工智能”(Artificial Intelligence,简称AI)这一概念,标志着人工智能学科的诞生。人工智能技术在这半个多世纪的发展中,其内涵也在相应地拓展。人工智能概念提出当时,人们将其定义为“制造智能机器的科学与工程”,发展人工智能技术的目的在于让机器拥有智能,也即让机器能够像人一样思考。时至今日,人工智能的内涵已经大大扩展,成为一门交叉学科。人工智能

是研究、开发能够模拟、延伸和扩展人类智能的理论、方法、技术及应用系统的一门新的技术科学。人工智能是计算机科学的一个分支学科,它企图了解智能的实质,并生产出一种新的能以与人类智能相似的方式作出反应的智能机器,通俗地说是企图使智能机器具备以下人类行为技能:会听(如语音识别、机器翻译等)、会看(如图像识别、文字识别等)、会说(如语音合成、人机对话等)、会思考(如人机对弈、定理证明等)、会学习(如机器学习、知识表示等)、会行动(如机器人、自动驾驶汽车等),等等。

人工智能最突出的特征是具有深度学习的能力。“无学习,不 AI”几乎成了人工智能研究在今天的核心指导思想。<sup>①</sup>深度学习是机器学习中的一个新领域,机器学习是一门专门研究计算机如何模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身性能的学科。深度学习作为机器学习的一个新领域,是一种能够对数据进行表征学习的方法,其目的在于建立、模拟人脑进行分析学习的神经网络,即模仿人脑的机制来解释数据,如图像、声音和文本。深度学习能力使人工智能机器人在应用过程中能够通过学习算法对所接触的数据进行学习并利用规律对未知数据进行预测,甚至创建规则。深度学习能力使得人工智能能够拥有超越人类的能力甚至脱离人类控制的可能,正如 AlphaGo 在对阵人类高手时表现卓越,重要原因就在于其深度学习的核心技术。

按照人工智能机器人<sup>②</sup>的不同形态,可以将其划分为人形人工智能机器人与非人形人工智能机器人,其中,非人形人工智能机器人又包括意识形态人工智能机器人(如 AlphaGo,其只是隐藏在电脑里的一个围棋程序)与实体人工智能机器人(如扫地机器人)。物理形体并非人工智能机器人的必备要素。而无论是否具有物理形体,人工智能机器人都必须依托算法(如深度学习和神经网络等),实时收集海量数据信息,并根据算法的计算结果实施行为。由此可见,人工智能机器人的本质要素是算法。但是鉴于算法的复杂性,将算法作为区分强人工智能机器人和弱人工智能机器人的依据,在法律上既无可能,也无必要。其一,“根据算法的复杂性和学习或运算能力对强人工智能机器人和弱人工智能机器人进行区分,这在技术认知上没有问题,但在法律上很难按照智能程度给出精确的标准。法律应对复杂世界的方式是确立一般性的简单规则,在概念上对社会个体

<sup>①</sup> 李开复、王咏刚:《人工智能》,文化发展出版社 2017 年版,第 34 页。

<sup>②</sup> 本书中所有人工智能产品、人工智能应用系统都统一称为人工智能机器人。需要说明的是,人工智能机器人之所以被称为“机器人”,是因为其具有自然人才具有的智能,而非因其有了人的形象。例如,世界人工智能机器人格斗大赛中的机器人形状都类似于碰碰车,而非类似于人;AlphaGo 也无人之形状;人工智能翻译机器人也仅仅只有编程而无实际机器,更无人形。

进行抽象假定(如行为能力),而非针对特殊主体,否则规则体系本身将变得异常复杂,难于理解和操作。而人工智能从单一的自动化服务向多元通用服务转变的过程可能是一个相当长的光谱,法律需要针对其本质特征进行约束,并探索一套应对未来的方案”。<sup>①</sup>其二,正如人类的大脑是一个复杂而精致的系统,尽管当代神经科学不断深化我们对人脑的认知,我们仍不能完全了解大脑的运转机制。但是,应当看到,没有完全理解人脑的运转机制,并未妨碍我们通过法律建立一整套的规则体系来控制人类的行为,也并未妨碍我们运用刑法对行为人的主观心态(即故意和过失)进行认定。尽管我们对人工智能的本质要素——算法未能完全了解,但是,只要在法律上对强人工智能机器人和弱人工智能机器人的划分确定一个明确的标准,就不会影响我们对我国刑法应当如何合理应对人工智能时代所带来的刑事风险等问题展开具体的分析与讨论。

以人工智能机器人是否具有独立意识和意志,以及是否能够在设计和编制的程序范围外自主决策并实施相应行为为依据,可以将人工智能机器人划分为弱人工智能机器人与强人工智能机器人。弱人工智能机器人也称限制领域人工智能机器人或应用型人工智能机器人,通常指专注于且只能解决特定领域问题的人工智能机器人。我们目前存在的人工智能机器人都属于弱人工智能机器人。弱人工智能机器人虽然可以在设计和编制的程序范围内进行独立判断并自主决策,但不具有独立意志,其实现的只是设计者或使用者的意志。例如,AlphaGo 虽然凭借自主判断和决策战胜了围棋国手,但其自主判断和决策仍是在设计和编制的程序范围内进行的,实现的是设计者或使用者的意志——在下围棋时战胜对手。强人工智能机器人又称为通用人工智能机器人或完全人工智能机器人,是指可以胜任人类所有工作的人工智能机器人,强人工智能机器人不仅具备胜任人类所有工作的能力,还应具备人类的“意识”,即具有独立意志,既可能在设计和编制的程序范围内进行独立判断并自主决策,实现设计者或使用者的意志,也有可能超出设计和编制的程序范围,进行自主决策并实施相应行为,实现其自身的意志。“强人工智能”一词最初是约翰·罗杰斯·希尔勒针对计算机和其他信息处理机器创造的。按约翰·罗杰斯·希尔勒的理解,“计算机不仅是用来研究人的思维的一种工具;相反,只要运行适当的程序,计算机本身就是有思维的”。<sup>②</sup>正如 1968 年上映的美国经典科幻电影《2001: 太空漫游》(《2001: A Space Odyssey》)所呈现的那样,人工智能机器人 HAL 读懂飞行员想要关闭它的唇语,于是决定先发制人,故意让电脑失灵,导致 4 人丧生。电影

<sup>①</sup> 胡凌:《人工智能的法律想象》, [https://news.uc.cn/a\\_16384305516683655823/](https://news.uc.cn/a_16384305516683655823/), 2017 年 9 月 25 日访问。

<sup>②</sup> J. Searle in *Minds Brains and Programs*. *The Behavioral and Brain Sciences*, Vol.3, 1980.

里的 HAL 就是我们这里所讲的强人工智能机器人,它可以超出设计和编制的程序范围自主作出让电脑失灵的决策,从而达到不让飞行员将自己关闭的目的。当然,这种情景只是人类的幻想,但是人类的科技发展史正是不断将幻想变成现实的历史。1783 年法国 J.F.P. 罗齐埃和 M. 达尔朗德乘气球升空,1903 年莱特兄弟飞机试飞成功,实现了庄子在《逍遥游》中所描述的“列子御风”的幻想;1969 年“阿波罗号”飞船首次成功登上月球,实现了“嫦娥奔月”的幻想。虽然到目前为止,人工智能机器人仍只能在设计和编制的程序范围内实施行为,实现设计者或使用者的意志,即我们目前仍处于弱人工智能时代,正如李开复在谈到人工智能机器人时所言,“现在的机器没有丝毫自我认知,没有感情,没有喜怒哀乐,作了决定也讲不出为什么”。但是随着深度学习、神经网络、蒙特卡洛树搜索、云计算等技术的不断发展和完善,当技术突破下一个瓶颈时,出现具有独立意识和意志、能够在设计和编制的程序范围外自主决策并实施相应行为、实现自身意志的人工智能机器人,其实并非镜花水月、海市蜃楼。

弱人工智能机器人不具有独立意志,仅能在设计和编制的程序范围内实施行为,实现人类设计和编制程序的目的。人类是弱人工智能机器人的创造者,可以通过程序给弱人工智能机器人设定行为目的和行为边界。此时,弱人工智能机器人的设计者或使用者的完全有可能让弱人工智能机器人为自己所用,实现自己的犯罪目的,如制造专门用来杀人的人工智能机器人,实现自己非法剥夺他人生命的目的。同时,不法分子可能利用弱人工智能机器人本身存在的漏洞,实现自己的犯罪目的。如诺基亚公司董事长李思拓所言,未来的家居设备很可能都是智能联网的,犯罪分子有可能通过洗碗机所连接的 Wi-Fi 侵入家庭网络,从而操纵家里的一切,实现自己的犯罪目的。在这种情况下,弱人工智能机器人实施的严重危害社会的犯罪行为本质上是为了实现他人的犯罪意志,应当将该弱人工智能机器人看作他人实施犯罪行为的“工具”。

强人工智能机器人依赖自身具有的独立思维,有能力在设计和编制的程序范围外,依据自己的独立意识和意志实施危害社会的行为。如上文中提到的强人工智能机器人 HAL 超出设计和编制的程序范围自主作出让电脑失灵的决策,从而达到不让飞行员将自己关闭的目的。此时,强人工智能机器人的行为未受到设计者或使用者的任何人为设置或操控,完全是自发实施的,且该种行为超越甚至根本违背了相关人员的意志。强人工智能机器人实施的这种犯罪行为和其他具有刑事责任能力的主体实施的犯罪行为相去无几。

## 二、人工智能的发展现状与未来

自 1956 年以来,经过六十余年的发展,人工智能作为一门新兴的技术科学

不断发展进步,大概经历了六个发展阶段。一是起步发展期,也即在1956年至20世纪60年代初之间。这一时期中,继人工智能概念提出后,相继取得了一批令人瞩目的研究成果,如机器定理证明、跳棋程序等,掀起人工智能发展的第一个高潮。二是反思发展期,也即20世纪60年代至70年代初期。在这一时期内,人工智能发展初期的突破性进展大大提升了人们对人工智能的期望,人们开始尝试更具挑战性的任务,并提出了一些不切实际的研发目标。然而,接二连三的失败和预期目标的落空(例如,无法用机器证明两个连续函数之和还是连续函数、机器翻译闹出笑话等),使人工智能的发展走入低谷。三是应用发展期,也即20世纪70年代初至80年代中期。20世纪70年代出现的专家系统模拟人类专家的知识和经验解决特定领域的问题,实现了人工智能从理论研究走向实际应用、从一般推理策略探讨转向运用专门知识的重大突破。专家系统在医疗、化学、地质等领域取得成功,推动人工智能走入应用发展的新高潮。四是低迷发展期,也即20世纪80年代中期至90年代中期。随着人工智能的应用规模不断扩大,专家系统存在的应用领域狭窄、缺乏常识性知识、知识获取困难、推理方法单一、缺乏分布式功能、难以与现有数据库兼容等问题逐渐暴露出来。五是稳步发展期,也即20世纪90年代中期至2010年。在这段时期中,由于网络技术特别是互联网技术的发展,加速了人工智能的创新研究,促使人工智能技术进一步走向实用化。1997年国际商业机器公司(以下简称IBM)深蓝超级计算机战胜了国际象棋世界冠军卡斯帕罗夫,2008年IBM提出“智慧地球”的概念。以上都是这一时期的标志性事件。六是蓬勃发展期,也即2011年至今。随着大数据、云计算、互联网、物联网等信息技术的发展,在感知数据和图形处理器等计算平台推动下以深度神经网络为代表的人工智能技术飞速发展,大幅跨越了科学与应用之间的“技术鸿沟”,诸如图像分类、语音识别、知识问答、人机对弈、自动驾驶等人工智能技术实现了从“不能用、不好用”到“可以用”的技术突破,迎来爆发式增长的新高潮。<sup>①</sup>

目前,我们正处于人工智能技术的蓬勃发展期,人工智能技术近期取得了突飞猛进的发展,其影响力渗透到人类社会生活的方方面面。人工智能技术的发展主要体现在专用人工智能技术方面(如围棋、自动驾驶等),与通用人工智能技术相比,面向特定任务的专用人工智能系统由于任务单一、需求明确、应用边界清晰、领域知识丰富、建模相对简单等特征,形成了人工智能领域的单点突破,在局部智能水平的单项测试中可以超越人类智能。例如,AlphaGo在围棋比赛中

<sup>①</sup> 参见谭铁牛:《人工智能的历史、现状与未来》, <https://mp.weixin.qq.com/s/7mV0P-iv4Pup4C-1NsVdEw>, 2019年3月3日访问。



战胜人类冠军,人工智能程序在大规模图像识别和人脸识别中达到了超越人类的水平,人工智能系统诊断皮肤癌达到专业医生水平。<sup>①</sup>人工智能技术在提高人们生活品质的同时,更成为促进社会发展、提升综合国力的重要力量。因此,人工智能技术蓬勃发展并日益受到重视。2017年7月国务院《新一代人工智能发展规划》将发展人工智能技术上升为国家战略,2018年10月习近平总书记在中共中央政治局第九次集体学习时指出,“加快发展新一代人工智能是事关我国能否抓住新一轮科技革命和产业变革机遇的战略问题”。<sup>②</sup>

在未来,人工智能技术将得到更大的发展,人工智能技术可能做到:结合人类情感、认知与推理等高阶智能形成具有类似于人类意识,并能够通用于各种场景中的强人工智能机器人,目前强人工智能机器人已处于探索阶段。人工智能技术的发展日新月异,人们也不由得担心人工智能到底会发展到什么程度,是否会出现超越人类能力并能不断进化和自我完善的强人工智能机器人,人类对其能力与意识可能无法预测与控制,从而给人类带来威胁。强人工智能可以比世界上最聪明、最有天赋的人类还聪明,牛津大学哲学家、未来学家尼克·波斯特洛姆在他的《超级智能》一书中将人工智能定义为“在科学创造力、智慧和社交能力等每一方面都比最强的人类大脑聪明很多的智能”。<sup>③</sup>关于强人工智能,将以何种形式予以表现,并没有明确的解析,因为强人工智能在今天还只能从哲学或科幻的角度进行设想,这种强于人类的智慧形式将如何存在无法根据现有的经验进行预测。

我们应当看到,人工智能技术的蓬勃发展在便利人类生活的同时也带来了一系列的社会问题和法律问题,冲击着我们现有的法律体系。2017年9月22日,浙江省绍兴市公安局破获“全国首例利用AI犯罪案”,这是我国首例利用人工智能技术侵犯公民个人信息的犯罪。<sup>④</sup>同时,人工智能是否带来更多的社会风险,甚至在未来是否会脱离人类的控制进而威胁人类的安全也时刻警醒着人类在享受人工智能福利的同时不能忽视其风险。由此可见,“加强人工智能发展的潜在风险研判和防范,加强人工智能相关法律、伦理、社会问题研究,建立健全保

① 参见谭铁牛:《人工智能的历史、现状与未来》, <https://mp.weixin.qq.com/s/7mV0P-iv4Pup4C-1NsVdEw>, 2019年3月3日访问。

② 胡喆:《抓住新一代人工智能发展的重大机遇》, [http://www.xinhuanet.com//2018-11/01/c\\_1123649864.htm](http://www.xinhuanet.com//2018-11/01/c_1123649864.htm), 2018年12月3日访问。

③ [英]尼克·波斯特洛姆:《超级智能》,张体伟、张玉青译,中信出版社2015年版。

④ 《全国首例AI技术犯罪案背后,你的个人信息被最先盯上》, <http://news.163.com/shuangchuang/17/1016/14/DoSH0CCB000197V8.html>, 2018年12月3日访问。