·科技建议·

# 人工智能应用中的安全、隐私和伦理挑战及应对思考

在1956年美国达特茅斯学院召开的学术研讨会上,人工智能(AI)首次成为一个正式学术领域,发展至今已有60多年的历史。随着人工智能技术的不断成熟,会有越来越多的机器人或智能程序充当人类助手,帮助人们完成复重性、危险性的任务,然而,随着人工智能技术的成熟和大范围应用的展开,人们可能会面临越来越多的安全、隐私和伦理等方面的挑战。

## 1　安全性挑战

人工智能技术应用引发的安全问题源于不同的原因,主要有以下几点。

### 1.1　技术滥用引发的安全威胁

人工智能对人类的作用很大程度上取决于人们如何使用与管理。如果人工智能技术被犯罪分子利用,就会带来安全问题,例如,黑客可能通过智能方法发起网络攻击,智能化的网络攻击软件能自我学习,模仿系统中用户的行为,并不断改变方法,以期尽可能长时间地停留在计算机系统中;黑客还可利用人工智能技术非法窃取私人信息;通过定制化不同用户阅读到的网络内容,人工智能技术甚至会被用来左右和控制公众的认知和判断。

### 1.2　技术或管理缺陷导致的安全问题

作为一项发展中的新兴技术,人工智能系统当前还不够成熟。某些技术缺陷导致工作异常,会使人工智能系统出现安全隐患,比如深度学习采用的黑箱模式会使模型可解释性不强,机器人、无人智能系统的设计、生产不当会导致运行异常等。另外,如果安全防护技术或措施不完善,无人驾驶汽车、机器人和其他人工智能装置可能受到非法入侵和控制,这些人工智能系统就有可能按照犯罪分子的指令,做出对人类有害的事情。

### 1.3　未来的超级智能引发的安全担忧

远期的人工智能安全风险是指,假设人工智能发展到超级智能阶段,这时机器人或其他人工智能系统能够自我演化,并可能发展出类人的自我意识,从而对人类的主导性甚至存续造成威胁。比尔·盖茨、斯蒂芬·霍金、埃隆·马斯克、雷·库兹韦尔等人都在担忧,对人工智能技术不加约束的开发,会让机器获得超越人类智力水平的智能,并引发一些难以控制的安全隐患[1-2]。一些研究团队正在研究高层次的认知智能,如机器情感和机器意识等。尽管人们还不清楚超级智能或“奇点”是否会到来,但如果在还没有完全做好应对措施之前出现技术突破,安全威胁就有可能爆发,人们应提前考虑到可能的风险。

## 2　隐私保护挑战

大数据驱动模式主导了近年来人工智能的发展,成为新一轮人工智能发展的重要特征。隐私问题是数据资源开发利用中的主要威胁之一,因此,在人工智能应用中必然也存在隐私侵犯风险。

### 2.1　数据采集中的隐私侵犯

随着各类数据采集设施的广泛使用,智能系统不仅能通过指纹、心跳等生理特征来辨别身份,还能根据不同人的行为喜好自动调节灯光、室内温度、播放音乐,甚至能通过睡眠时间、锻炼情况、饮食习惯以及体征变化等来判断身体是否健康。然而,这些智能技术的使用就意味着智能系统掌握了个人的大量信息,甚至比自己更了解自己。这些数据如果使用得当,可以提升人类的生活质量,但如果出于商业目的非法使用某些私人信息,就会造成隐私侵犯。

### 2.2　云计算中的隐私风险

因为云计算技术使用便捷、成本低廉,提供了基于共享池实现按需式资源使用的模式,许多公司和政府组织开始将数据存储至云上。将隐私信息存储至云端后,这些信息就容易遭到各种威胁和攻击。由于人工智能系统普遍对计算能力要求较高,目前在许多人工智能应用中,云计算已经被配置为主要架构,因此在开发该类智能应用时,云端隐私保护也是人们需要考虑的问题。

### 2.3　知识抽取中的隐私问题

由数据到知识的抽取是人工智能的重要能力,知识抽取工具正在变得越来越强大,无数个看似不相关的数据片段可能被整合在一起,识别出个人行为特征甚至性格特征。例如,只要将网站浏览记录、聊天内容、购物过程和其他各类别记录数据组合在一起,就可以勾勒出某人的行为轨迹,并可分析出个人偏好和行为习惯,从而进一步预测出消费者的潜在需求,商家可提前为消费者提供必要的信息、产品或服务。但是,这些个性化定制过程又伴随着对个人隐私的发现和曝光,如何规范隐私保护是需要与技术应用同步考虑的一个问题。

## 3　伦理规范挑战

伦理问题是人工智能技术可能带给人们的最为特殊的问题。人工智能的伦理问题范围很广,其中以下几个方面值得关注。

### 3.1　机器人的行为规则

人工智能正在替代人的很多决策行为,智能机器人在作出

决策时,同样需要遵从人类社会的各项规则。比如,假设无人驾驶汽车前方人行道上出现3个行人而无法及时刹车,智能系统是应该选择撞向这3个行人,还是转而撞向路边的1个行人?人工智能技术的应用,正在将一些生活中的伦理性问题在系统中规则化。如果在系统的研发设计中未与社会伦理约束相结合,就有可能在决策中遵循与人类不同的逻辑,从而导致严重后果。

### 3.2 机器人的权力

目前在司法、医疗、指挥等重要领域,研究人员已经开始探索人工智能在审判分析、疾病诊断和对抗博弈方面的决策能力[3-4]。但是,在对机器授予决策权后,人们要考虑的不仅是人工智能的安全风险,而且还要面临一个新的伦理问题,即机器是否有资格这样做。随着智能系统对特定领域的知识掌握,它的决策分析能力开始超越人类,人们可能会在越来越多的领域对机器决策形成依赖,这一类伦理问题也需要在人工智能进一步向前发展的过程中梳理清楚。

### 3.3 机器人的教育

有伦理学家认为,未来机器人不仅有感知、认知和决策能力,人工智能在不同环境中学习和演化,还会形成不同的个性。据新华网报道,国外研发的某聊天机器人在网上开始聊天后不到24个小时,竟然学会了说脏话和发表种族主义的言论,这引发了人们对机器人教育问题的思考。尽管人工智能未来未必会产生自主意识,但可能发展出不同的个性特点,而这是受其使用者影响的。机器人使用者需要承担类似监护人一样的道德责任甚至法律责任,以免对社会文明产生不良影响。

## 4 启示与建议

人类社会即将进入人机共存的时代,为确保机器人和人工智能系统运行受控,且与人类能和谐共处,在设计、研发、生产和使用过程中,需要采取一系列的应对措施,妥善应对人工智能的安全、隐私、伦理问题和其他风险。

### 4.1 加强理论攻关,研发透明性和可解释性更高的智能计算模型

在并行计算和海量数据的共同支撑下,以深度学习为代表的智能计算模型表现出了很强的能力。但当前的机器学习模型仍属于一种黑箱工作模式,对于AI系统运行中发生的异常情况,人们还很难对其中的原因作出解释,开发者也难以准确预测和把握智能系统运行的行为边界。未来人们需要研发更为透明、可解释性更高的智能计算模型,开发可解释、可理解、可预测的智能系统,降低系统行为的不可预知性和不确定性,这应成为人工智能基础理论研究的关注重点之一。

### 4.2 开展立法研究,建立适应智能化时代的法律法规体系

欧盟、日本等人工智能技术起步较早的地区和国家,已经意识到机器人和人工智能进入生活将给人类社会带来的安全与伦理问题,并已着手开展立法探索,如2016年5月,欧盟法律事务委员会发布《就机器人民事法律规则向欧盟委员会提出立法建议》的报告草案[5],探讨如何从立法角度避免机器人对人类的伤害。有效应对未来风险挑战需强化立法研究,明确重点领域人工智能应用中的法律主体以及相关权利、义务和责任,建立和完善适应智能时代的法律法规体系。

### 4.3 制定伦理准则,完善人工智能技术研发规范

当人工智能系统决策与采取行动时,人们希望其行为能够符合人类社会的各项道德和伦理规则,而这些规则应在系统设计和开发阶段,就需被考虑到并被嵌入人工智能系统。因此,需要建立起人工智能技术研发的伦理准则,指导机器人设计研究者和制造商如何对一个机器人做出道德风险评估,并形成完善的人工智能技术研发规范,以确保人工智能系统的行为符合社会伦理道德标准。

### 4.4 提高安全标准,推行人工智能产品安全认证

可靠的人工智能系统应具有强健的安全性能,能够适应不同的工况条件,并能有效应对各类蓄意攻击,避免因异常操作和恶意而导致安全事故。这一方面需要提高人工智能产品研发的安全标准,从技术上增强智能系统的安全性和强健性,比如完善芯片设计的安全标准等;另一方面要推行智能系统安全认证,对人工智能技术和产品进行严格测试,增强社会公众信任,保障人工智能产业健康发展。

### 4.5 建立监管体系,强化人工智能技术和产品的监督

由于智能系统在使用过程中会不断进行自行学习和探索,很多潜在风险难以在研发阶段或认证环节完全排除,因此加强监管对于应对人工智能的安全、隐私和伦理等问题至关重要。建议在国家层面建立一套公开透明的人工智能监管体系,实现对人工智能算法设计、产品开发、数据采集和产品应用的全流程监管,加强对违规行为的惩戒,督促人工智能行业和企业自律。

### 4.6 推动全球治理,共同应对风险挑战

人工智能的快速发展是全球各国共同面临的问题,应明确科学家共同体、政府与国际组织各自的职责,引导各国积极参与人工智能全球治理。加强机器人伦理和安全风险等人工智能国际共性问题研究,深化人工智能法律法规、行业监管等方面的交流合作,推进人工智能技术标准和安全标准的国际统一,使人工智能科技成就更好地服务于人类社会。

## 参考文献

[1] Ray Kurzweil. 奇点临近[M]. 李庆诚, 董振华, 田源, 译. 北京: 机械工业出版社, 2011.
[2] Kile F. Artificial intelligence and society: A furtive transformation[J]. AI & Society, 2013, 28(1): 107-115.
[3] Alzou S, Alshibly H, Aitah A M. Artificial intelligence in law enforcement, a review [J]. International Journal of Advanced Information Technology, 2014, 4(4): 1-9.
[4] Sikchi S S, Sikchi S, Ali M. Artificial intelligence in medical diagnosis [J]. International Journal of Applied Engineering Research, 2012, 7(11): 1539-1543.
[5] Committee on Legal Affairs of European Parliament. Report with recommendations to the Commission on Civil Law Rules on Robotics[R]. Strasbourg: European Parliament, 2017.

**致谢:**本研究受科技部改革发展专项"中国人工智能2.0规划编制"(2016GH010036)、科技部科技创新战略研究专项"重大科技项目和科技工程形成机制研究"(ZLY2015133)资助。

**文/李修全**
作者简介:中国科学技术发展战略研究院,副研究员。

**(责任编辑 王丽娜)**