

对人工智能工程伦理的讨论

SA20225085 朱志儒

摘要

随着人工智能技术的快速发展，人们不断从中获得许多便利，但“去人类中心化”的风险也逐渐体现。本文从技术、责任、环境、利益四个方面展开论述，分析人工智能技术面临的伦理道德风险，基于人工智能发展的初衷，即增加人类健康和安全，需要对其进行预防和控制。

正文

人类社会正从以计算机、互联网、大数据等技术支撑的信息社会，步入以人工智能为关键技术支撑的智能社会，人类的生活方式和格局将由此发生巨大改变。人工智能技术在抗击今年的新冠肺炎疫情中起到了至关重要的作用，例如，阿里巴巴团队推出的“AI 防疫师”系统提供实时精准测体温、口罩佩戴识别、预警及追踪高危人群等防控功能，百度的智能外呼平台用语音机器人及时帮助政府机构快速完成居民排查和通知。第四次工业革命的一个标志就是人工智能，人工智能具有广泛的特点，其应用场景具有多样性，在科研行业发展迅猛，在日常生活中应用广泛。AlphaGo 击败人类的围棋冠军后，人们对人工智能有了新的看法，各大电商、新闻、娱乐平台的智能推荐系统甚至比用户更加了解自身的喜好，语音识别系统增加了更为人性化的设计。

个人和社会在享受人工智能带来的便利的同时，应该深入考虑人工智能技术带来的道德风险。在人工智能的发展初期，如果没有预设相应的伦理道德机制，可能会引发极大的不确定性和危害。比如，2016 年 3 月微软发布的聊天智能机器人 Tay 没有预设相应的伦理道德指导机制，在之后的 24 小时内被某位网民调教成一个极端主义分子，而这迫使微软不得不将其下线。同时，祸不单行，一辆自动驾驶的 Uber 汽车于 2018 年 3 月在美国亚利桑那州撞死行人，人们对自动驾驶技术产生信任危机。由此，自动驾驶技术陷入伦理道德困境，在无法避免交通事故发生时，先救乘客还是行人？只救一人还是救多人？自动驾驶技术将面对电

车难题。

伴随人工智能技术在实际生活中的不断应用，相应的社会问题也渐渐付出水面，人们意识到对人工智能伦理的考量将变得愈发重要，人工智能技术的研发和应用应该设置什么样的伦理边界、遵循什么样的原则、遵守什么样的规章制度，这些深刻影响着人工智能技术未来的发展方向。

董青岭提到的人工智能朝着“去人类中心化”方向发展带来的道德风险值得警惕，未来人工智能不再是人类生产生活的附属品和效率工具的延伸，而是以一种“类人”姿态参与人类社会的运转，机器人网络未来或许会成为社会决策的中心枢纽，人机关系可能会从现在的从属关系转变成未来的共生甚至是相争关系。那么，在以后的人工智能的研发和应用中，我们是否需要嵌入人类社会的道德准则呢？如果需要的话，是以多大的程度及范围呢？

著名科学家霍金指出，人工智能技术创造出的能够独立思考的机器对人类构成极大威胁。所以，我们不但要提高技术层面的要求，更要伦理道德层面的引导和规制，以设计和研发出更加安全可靠地人工智能系统。

基于算法和大数据的人工智能技术背后隐藏着伦理道德的危机。算法是工程师和程序设计者的逻辑和思想的产物，他们的政治立场和社会偏见会不可避免地影响他们设计的算法。对数据的预处理和特征提取的过程中，通常会存在偏差，在之后的训练过程中，这种偏差可能不会被消除甚至会被进一步放大。所以，工程师和设计者在研发初期就需要消除主观偏见，避免引入不公正的价值观，以消除算法和数据中的偏差，减少人工智能因技术因素带来的伦理道德风险问题。

在上面提到的自动驾驶问题中，体现了人工智能在责任承担与价值选择困境中存在风险。当机器像人一样作为自主决策者在面对电车难题时应该如何选择，这是在设计人工智能时应该考虑的问题。在无法避免交通事故发生时，自动驾驶汽车先救乘客还是行人？只救一人还是救多人？在我们实际生活场景中，驾驶者可根据当时具体情况和社会的规范约束做出自己的选择，但对于机器而言，预先设置的选择并不一定能最小化即将造成的伤害。当交通事故发生时，在传统驾驶方式下，有较为清晰的法律流程来确定责任主体，并配有相应的法律责任。但现在人工智能机器没有社会规范意识，只能根据预置的算法和当时的场景进行选择判断，没有能力为其行为承担相应的责任。所以，自动驾驶技术的发展瓶颈不仅

仅是技术，更多的是伦理道德问题的限制。

人工智能的发展在一定程度上会对生态环境造成损害。人工智能机器的研发需要不断从自然界索取不同的物质，加速了人类对自然资源的消耗，其更新换代被淘汰的产品在自然环境中很难降解，所以，人工智能的发展会对生态环境造成一定程度的损害，可能恶化原本严重的生态问题。根据人类追求的环境伦理的可持续性、人与自然和谐相处等基本原则，需要对上述问题加以约束和规范。

人工智能存在侵犯人类利益的风险。人工智能对人类来说应该是安全的、可靠的，自动驾驶技术的初衷是减少交通事故的出现，但近年来发生的自动驾驶汽车交通事故，引发人们对其伦理道德风险的思考。其次，人工智能利用大数据技术“掌握”用户的喜好和习惯，对用户进行“大数据杀熟”，如果用户数据落入诈骗分子的手中，会给用户带来隐私被侵犯的问题，让用户面临财产受损的危险。

参考文献

- [1] 高文, 黄铁军. 从信息社会迈向智能社会
- [2] 人工智能在新冠肺炎战“疫”中能帮什么忙?
- [3] 郭爽. 微软聊天机器人为何会“学坏”
- [4] 林雨佳. 自动驾驶事故中的过失犯罪分析
- [5] 董青岭. 人工智能时代的道德风险与机器伦理
- [6] 霍金. 人工智能可能使人类灭绝