新一代人工智能的风险防范和高质量发展

孙大伟[1] 张淑芬[2]
1. 中国社会科学院社会学研究所，北京 100732
2. 中国社会科学院金融研究所，北京 100028

摘要：当前，以算法、数据和计算能力为核心驱动力的新一代人工智能发展迅速，但是算法面临着算法歧视、“过滤气泡”、算法黑箱等问题，数据存在数据滥用、数据泄露、数据污染等多重风险。因此，应当加强人工智能发展的潜在风险研判和防范，明确人工智能的发展应遵循人类利益原则、人类自主原则、责任原则，完善相关法律规则，强化监管，推动新一代人工智能高质量发展。

关键词：新一代人工智能；算法风险；数据风险
中图分类号：TP18 文献标识码：A

未来学家阿尔文·托夫勒认为，现代人是第三次浪潮的参与者。较之 1 万年前掀起的第一次农业革命浪潮以及 300 年前惊天动地的第二次工业革命浪潮，第三次浪潮带给人们的是“有史以来最强烈的社会变动和创造性的重组”的“超级工业社会”（阿尔文·托夫勒，2018）[1]。这个“超级工业社会”的特点之一就是高新科技发展日新月异。其中，新一代人工智能以及与之相关的物联网、区块链、云计算、大数据便被认为是当前最具有标志性的新兴科技。

1 新一代人工智能被广泛应用

近年来，我国已经在新一代人工智能的五大核心技术：图像识别、语音识别、自然语言处理、知识图谱、机器人技术等领域取得重大进展，并将其广泛应用到了社会生产生活当中。

当前，在针对新型冠状病毒感染肺炎疫情防控中，新一代人工智能技术就发挥了重要作用。依托人工智能技术的仪器能快速精准地测量体温、识别高温人群，人工智能机器人可以承担消毒、清洁、药物配送等工作，极大地减少了人与人之间的接触概率，降低了感染风险。利用人工智能技术还可以进行病毒样本比对以及自动化诊断和分析等，提升精准筛查比例。

需要强调的是，技术是中性的，技术的发展会带来正反两方面的效应。新一代人工智能是以计算为中心，其核心驱动力是算法、数据和计算能力。推陈出新的算法、海量的数据、惊人的算力以及巨大的应用需求，深刻影响着人类的生产生活的同时，在其发展过程中也存在着诸多风险和挑战。

2 新一代人工智能的算法风险

众所周知，算法是人工智能的根基。随着算法模型的推陈出新，算法的自主决策性逐渐提高，人工智能技术得到广泛应用，极大解放了人类的劳动力，提高了工作效率。但是，算法同时也存在算法歧视、“过滤气泡”“信息茧房”和“回音室效应”、算法黑箱等问题。

2.1 算法歧视

算法带有一定的主观性。算法歧视是指在人工智能系统研发过程中，研发人员将其主观的价值判断植入人工智能系统，导致系统在编码、收集、选择或使用数据时，产生了带有偏见或歧视性的结果。

算法歧视主要分为三类：人为造成的歧视、数据驱动的歧视和机器自我学习的歧视（郭锐，2019）[2]。人为造成的歧视是指研发人员在研发人工智能时植入带有偏见或歧视的价值观。数据驱动的歧视是指采集的数据本身包含偏见与歧视。机器自我学习造成的歧视是指人工智能机器在学习过程中自我学习到数据的不

作者简介：孙大伟（1982–），男，博士、副研究员，研究方向：人才学、廉政学等。
张淑芬（1983–），女，博士后，研究方向：金融风险、金融监管、金融犯罪。

同特征，并将某些偏见或歧视引入到决策过程中。以上这些偏见与歧视可能会导致人工智能系统在决策过程中对某些群体或个人的直接或间接的偏见与歧视。

在人工智能应用中，搜索引擎、电子商务、社交媒体等平台均可能存在算法歧视问题。比如电子商务平台为向客户进行精准营销，往往会根据已有数据和算法，进一步挖掘客户的消费数据，产生价格歧视和“大数据杀熟”等问题。

2.2 “过滤气泡”

在信息爆炸的当下，人们需要通过“算法推荐”为自己提供个性化服务。但是搜索引擎平台、电子商务平台、社交媒体平台等网络平台使用“推荐算法”，容易制造“过滤气泡”问题，引发“信息茧房”和“回音室效应”。

“过滤气泡”指的是计算机记录用户进行网上搜索、浏览留下的痕迹，根据这些痕迹推断出用户的信息偏好，通过算法进行计算并向用户进一步推送相关信息，以实现用户信息的个性化定制，保证用户黏性(Eli Pariser，2011）。[3]“信息茧房”是指用户更愿意根据自己的需求，选择接触自己感兴趣并乐于接受的固定领域的信息，但这就像“茧房”一样封闭，充满了与自己意见一致的观点，虽然温暖而舒适，但是也会形成毫无根据的极端主义、偏激错误的观念甚至过度自卑的心态（凯斯·桑斯坦，2008）[4]6。“回音室效应”则是指用户在互联网上就某一话题或观点进行互动，一些相近或相同的信息会不断重复、强化、夸张和扭曲，从而让处于封闭环境中的大多数人认为这些不断重复的信息是事实的全部（凯斯·桑斯坦，2008）[4]。

“过滤气泡”引发的“信息茧房”和“回音室效应”，限制用户接触信息的范围和途径，使公众接收到的多是同质化的信息，难以获取全面信息，很难突破信息壁垒寻找其他类型的信息，容易导致用户信息窄化，甚至有可能会出现“网络群体极化”现象。

2.3 算法黑箱

“黑箱”是控制论的概念，作为一种隐喻，是指那些不为人知的不能打开、不能从外部直接观察其内部状态的系统（张淑玲，2018）[5]。“黑箱”中的整个技术操作过程并不透明，决策过程不可解释，决策结果不可预测。用户既不了解算法的目标和意图，也不知道算法的研发人员、实际控制者以及人工智能系统生成内容的责任归属等信息（叶韦明，2016）[6]，只能被动接受通过算法所产生的决策结果。但是，算法的研发人员、实际控制者等却可以利用所掌握的数据和算法技术，生成符合其自身需要的决策结果。

算法黑箱会产生很多问题：一方面，会在用户与算法、数据的实际控制者之间形成信息壁垒和数字鸿沟，导致用户难以了解个人信息的处理情况，无法掌握个人信息，而数据、算法的实际控制者则会拥有大量数据，并通过清洗、加工和分析数据，更新算法，获得重要的商业资源和竞争优势。另一方面，带来监管困难，一旦出现数据泄露、数据滥用、数据操纵、个人隐私权受到侵害等问题，很难得到救济。

3 新一代人工智能的数据风险

在实际的社会生活运行中，通过新一代人工智能技术，大数据使用者可以采集到用户的各种信息，在万物互联、大数据和人工智能三层叠加后，人类隐私似乎无所遁形（何波，2018）[7]。同时，政府和企业决策尽管对大数据采用了匿名化、加密化等预防措施，但是依然面临着多重数据风险。

3.1 数据滥用

数据是人工智能发展的基础，但同时，数据滥用问题也日益突出，特别是违法违规使用个人信息的问题十分严重。

首先，在人工智能时代，企业出于商业或其他目的，要求用户在互联网平台或手机 APP 上注册个人信息，这些海量信息经加工后形成巨大的数据库。部分企业会将此作为重要资源和竞争优势，用于商业化，从而引发诸如“大数据杀熟”、过度营销等问题。

其次，社交媒体平台为用户分享信息提供平台的同时，也有可能成为制造和传播谣言、进行人身攻击的重要渠道。由于现阶段个人信息保护乏力，且大多数用户缺乏个人信息保护意识，给网络暴力、人肉搜索、深度伪造、流量造假、操纵账号等非法行为带来了一定的可乘之机。

3.2 数据泄露

当前，数据泄露事件层出不穷。在人工智能应用中，如果相关主体操作不当，或黑客非法对人工智能系统进行攻击，都有可能导致海量数据泄露，对个人、企业和政府信息安全带来极大风险。

数据泄露具有以下特点：一是泄露范围广泛。个人数据、企业数据、政府数据在内的海量数据都具有泄露

风险。二是泄露原因复杂。既有内部管理疏漏或操作不当导致泄露，也有不法分子非法攻击造成泄露；既有管理缺陷，也有技术漏洞。三是危害严重。数据一旦泄露，极有可能会超越技术范畴和组织边界，严重危害个人、企业甚至国家的信息安全，侵害商业安全和国家安全，改变政治进程。

3.3 数据污染

当前人工智能处于海量数据驱动知识学习阶段，数据集的规模和质量决定着人工智能模型的质量（张宇光等，2019）[8]。对于人工智能系统来说，数据是其进行分析、判断、输出决策结果并进行行动的依据，如果通过数据投毒等方式污染训练数据集，将会影响人工智能模型的准确率，产生的决策结果可能完全不同。

一般来说，数据污染有两种方式。一种是训练数据集的污染，包括训练数据集本身已经被污染和被黑客攻击污染。另一种是人工智能模型的污染。一旦数据被污染，将会产生人工智能决策失误，出现对人工智能的信任危机，甚至会引发安全问题。

4 推进我国新一代人工智能高质量发展的建议

习近平总书记高度重视我国人工智能技术和产业的发展。他在2018年中共中央政治局第九次集体学习时强调指出，要加强人工智能发展的潜在风险研判和防范，维护人民利益和国家安全，确保人工智能安全、可靠、可控。因此，我们应积极应对新一代人工智能发展面临的一系列问题和挑战，积极推进我国新一代人工智能的高质量发展，助力经济高质量发展。

4.1 明确人工智能发展原则

不同国家、地区、国际组织等主体就人工智能发展原则发布的声明、报告、规划、指南等文件不胜枚举，其中规定不尽相同，但也有共通之处。总体而言，新一代人工智能发展应当遵循以下基本原则。

第一，遵循人类利益原则。这是人工智能发展的重要原则，任何技术都应当以人类的根本利益作为目标。

一是在算法方面，应当确保算法的透明性和可解释性，避免算法歧视和偏见，解决算法黑箱问题；制定公平的利益分配机制和机会平等的算法模型，缩小数字鸿沟。二是在数据方面，应当注意隐私保护，防止数据泄露、数据滥用、数据污染给人类造成损害，做到安全可控。三是应当尊重人类尊严，保障人的基本权利和自由，促进正义、公平、团结和民主。人工智能应当造福人类，不能损害人类利益，注意防止人工智能武器的军备竞赛。

第二，人类自主原则。即人类应当能够了解、控制和监督人工智能的每一个决策，不能反被人工智能系统操纵。

从科幻小说家阿西莫夫提出“机器人三法则”起，对人工智能是否具有主体性就一直存在争议。很多学者对此持支持态度：科林·艾伦（Colin Allen，2000）[9]认为，人工智能越来越接近自主智能体的标准，希望设计能通过道德图灵测试的人工道德智能体。肖恩·拜仁（Shawn Bayern，2017）[10]等认为，机器人的法律地位可以借鉴公司人格的形式，被纳入法律主体之中。王荣余（2019）[11]认为，人工智能的本质是算法，基于算法的“深度学习”和“神经网络”使智能机器人有可能成为非完全的主体或独立主体。刘宪权（2019）[12]认为，将具有辨认能力和控制能力的强智能机器人作为刑事责任主体不仅有其合理性，且有利于发挥刑法的机能。

更多学者认为人工智能尚不具有自主性，不需要赋予其主体地位。巴尔托什·布罗热克（Bartosz Brozek）和马利克·杰库比克（Marek Jakubiec）（2017）[13]认为，从技术角度来看，赋予人工智能法律主体地位具有可能性，但是实践上不具有可行性。李醒民（2019）[14]认为，强人工智能嵌入的伦理算法或程序不是实践伦理学的具体道德行为；强人工智能的行动不具有道德意义，不具有道德主体地位。

人类自主原则的基本要求如下：一是人工智能要以人为本，尊重人类选择，由人类决定是否、如何、何时、何地将何种事项委托给人工智能系统进行决策和行动，人工智能不能脱离人类的控制。二是人工智能模拟、增强、深化、拓展、补充人的智能——包括认知、识别、记忆、学习、理解、推理、行动等——应以促进人类社会发展为目的，不能欺骗、操纵人类。三是人工智能要具有透明性和可预测性，即人类应当知道人工智能技术、决策过程以及能够预测其决策结果。

第三，责任原则。责任原则是人工智能研发和应用的基础，具体是指人工智能在研发和应用中都应当有明确的责任体系。

一是从技术层面来说，应当明确人工智能研发、设计和制造过程中的责任主体。人工智能的研发、设计和

制造者对人工智能在使用、误用和产生决策结果时产生的伦理影响以及人身或财产损害，应当承担相应的法律责任。二是从应用层面来说，应当建立合理的责任分配和承担机制，遵循权责一致的原则。首先，人工智能的数据、算法、运行过程、决策结果、应用情况等应当被合理记录并受到监督，一旦出现问题可以进行相关审查。其次，应当明确人工智能使用过程中的责任主体。人工智能的参与者应当根据其角色，场景和所起的作用，对人工智能的使用、误用和产生决策结果负责。

4.2 完善人工智能的法律规则

目前，我国与人工智能发展相适应的法律法规相对滞后，这在一定程度上制约了新一代人工智能的高质量发展和应用。因此，我们应当加强相关立法工作，建立健全保障新一代人工智能高质量发展的法律体系。

第一，当前对智能的概念、特征、技术水平、标准化体系、应用模式等尚没有形成统一意见，人工智能发展所带来的正负效应也有待时间检验，因此人工智能相关法律法规的建立健全，制度体系的完善也需要一定的时间，需要充分论证、考察、调研。

第二，应当充分关注人工智能的法律地位，设置合理的责任分配和承担机制。目前，人工智能尚不具备自主性，不应赋予其主体地位，应当根据人工智能研发、制造、使用等各过程中各方的职责、过错等，确定法律责任的分配和承担问题。

第三，关于算法规范问题。要通过立法增强算法的可解释性和透明性，应禁止采用带有偏见与歧视的算法，确保人工智能系统不会因为特定因素产生不同的决策结果。此外，还应确保各类网络平台向用户进行定向推送时，首先征得用户同意，并公开说明其应用的算法，保障用户的知情权。

第四，关于数据保护问题。进一步建立健全保障数据安全的法律法规和制度体系。应当对数据的采集、分析、存储、流转、使用等全过程进行规制，保障数据安全。强化个人隐私信息保护，未经个人同意，不得收集和向他人提供个人隐私信息，并对个人隐私信息采取延伸式保护，不得非法使用和存储个人信息，不得非法通过重组、关联、交叉等方式分析挖掘个人信息。

4.3 强化人工智能监管

对于人工智能的监管，不同国家采取了不同的监管模式。日本注重推动科技创新，倾向于“无需批准式”监管模式，除非有足够充分证据证明高新科技和新的商业模式的危险性，否则不被禁止；英国、法国等国家更注重安全性，采用了“审慎监管”模式，需先证明高新科技和新的商业模式不具有危险性，才允许被使用（张富利，2019）[15]。前者有利于推动科技发展和应用，但对高新科技和新的商业模式存在的风险监管有疏漏。后者可从源头监管，但对高新科技和新的商业模式的发展会产生一定负效应。我国应当立足本国国情，充分总结“无需批准式”监管和“审慎监管”经验教训，统筹规划人工智能监管体系。

第一，加快构建人工智能监管体系。制定人工智能的发展规划和监管策略，实施符合我国国情和人工智能发展规律的监管措施，对人工智能技术发展中存在的风险挑战问题进行治理，引领人工智能安全、健康的发展。

第二，政府、企业、行业等各主体协同一体。一方面，政府应当加强监管。政府应当对人工智能研发、制造、应用等全过程进行全面监管，防止人工智能被非法利用。加大对人工智能领域的算法歧视、“过滤气泡”、算法黑箱、数据滥用、数据泄露、数据污染等行为的惩戒力度。另一方面，应当推动人工智能行业自律，强化企业和行业对人工智能技术的算法风险、数据风险等责任意识，防止人工智能的研发、制造和使用偏离既定目的。

第三，分层次监管。对于不同的人工智能技术采用不同的监管策略。对于可能引发严重伦理、法律、社会风险及安全问题的人工智能技术及应用，采取更严厉的监管措施。对于仅有可能引发一般风险及安全问题的人工智能技术及相关应用，可以采取相对宽松的监管措施。

参考文献

[1] 阿尔文·托夫勒 . 第三次浪潮 [M]. 黄明坚，译 . 北京：中信出版社 .2018:1-10.

[2] 郭锐 . 人工智能的伦理与治理 [J]. 人工智能，2019(4):14.

[3] 郭小安，甘馨月 .“戳掉你的泡泡”——算法推荐时代“过滤气泡”的形成及消解 [J]. 全球传媒学刊，2018，5(2):77.

[4] 凯斯·桑斯坦 . 信息乌托邦：众人如何生产知识 [M]，毕竞悦，译 . 北京：法律出版社 .2008:6.

[5] 张淑玲 . 破解黑箱：智媒时代的算法权力规制与透明实现机制 [J]. 中国出版，2018(7):50.

[6] 叶韦明.机器人新闻：变革历程与社会影响[J].中国出版，2016(10):19.

[7] 何波.人工智能时代数据保护莫沦为皇帝新衣 [N] .人民邮电报，2018-1-31(5).

[8] 张宇光，孙卫，刘贤刚，等.人工智能安全研究[J].保密科学技术，2019(9):9.

[9] COLIN A，GARY V，JASON Z. Prolegomena to Any Future Artificial Moral Agent[J]. Journal of Experimental Theory of Artificial Intelligence，2000，12(3):251.

[10] BAYERN S，BURRI T，GRANT T D，et al. Company Law and Autonomous Systems: A Blueprint for Lawyer，Entrepreneurs，and Regulators[J]. Hastings Science and Technology Law Journal，2017，9(2):135-162.

[11] 王荣余.“机器人也是人”的法理拷问[J].社会科学动态，2019(11):22.

[12] 刘宪权.对强智能机器人刑事责任主体地位否定说的回应[J].法学评论，2019，217(5):113.

[13] BROZEK B，JAKUBIEC M. On the Legal Responsibility of Autonomous Machines(J).Artificial Intelligence and Law，2017，25(3):293.

[14] 李醒民.人工智能技性科学与伦理[J].社会科学论坛，2019(4):188-189.

[15] 张富利.全球风险社会下人工智能的治理之道——复杂性范式与法律应对[J].社会科学文摘，2019(9):73.

（上接第04页）

向影响，显示参股策略对于股东获利有所帮助。

5.2 研究建议

新团队新营销技术：在新营销技术方面，现在大多数群众都认为新营销技术为互联网金融，并且坚信新营销技术会带来新人脉客群（含非金融业异业结盟）。

降低理赔率的策略：多数的群众认为扩大保险覆盖率可达到降低理赔率，但是只有少数的群众认为公司将会设计保险互联网核保制度来降低理赔率，部分群众认为公司仍然应该实地查核，显示联网保险趋势应该加快设计保险互联网核保制度。

在征信制度方面：在征信部分，公司要建立征信团队，利用股东关系企业提供征信资料，降低理赔率。整体而言，台资保险公司改组原因为公司会利用股东关系企业提供征信资料，降低理赔率[10]。

参考文献

[1] JAMES C，HAO，CHOU L-Y. The estimation of efficiency for life insurance industry: The case in Taiwan[J]. Journal of Asian Economics，2005(16):847-860.

[2] CHOU L-Y，CHU S-H M. The effect of capital requirement policies on X-efficiency and profitability: the case of Taiwan life insurance industry[J]. The empirical economics letters，2008，7(3):1-14.

[3] LAI G C，CHOU L-Y，CHEN L R.The Impact of Organizational Structure and Business Strategy onPerformance and Risk-Taking Behaviorin Insurance Industry[J]. Applied Finance and Accounting，2015(1):107-128.

[4] CHOU L-Y. The impact of going public strategy on free cash flow: the case of Taiwan life insurance industry[J]. Journal of Insurance and Risk Management，2008，12(4):1-34.

[5] HELWEGE J，PIRINSKY C，STULZ R M. Why do firm become widely held? An analysis of the dynamics of corporate ownership[J]. The Journal of Finance，2007，62(3):995-1028.

[6] LAI G C，LIMPAPHAYOM P. Organizational structure and performance: evidence from the nonlife insurance industry in Japan[J].Journal of Risk and Insurance，2003，70(4):735-757.

[7] LAMM-TENNANT J，STARKS L. Stock versus mutual ownership structures: the risk implications[J]. Journal of Business，1993，66(1):29-46.

[8] CHOU L-Y，CHU S-H M. Decision to go public: Empirical analysis on financial industries in taiwan[J]. International journal of Applied Business and Economic Research，2016(14):9147-9173.

[9] 尤胜平.大陆投资对于台湾保险企业获利影响的研究[D].台北：台湾东吴大学，2017.

[10] 周林毅，尤胜平.海峡两岸互联网保险合作与优劣势发展探讨[J].保险职业学院学报，2017，31(2):70-76.