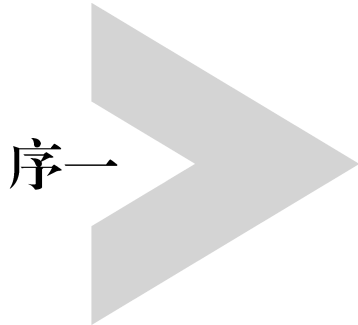




序一

健康是人类永恒的追求。健康知识的普及、健康理念的树立和健康行为的培育,是社会文明程度的重要体现。长期以来,人们对健康生活的追求,往往受制于优质信息匮乏、获取成本过高以及个性化需求难以满足等问题。随着信息技术的不断发展与广泛应用,健康传播日益受到学界与业界的重视。我国《全民健身计划(2021—2025年)》明确提出,要开设线上科学健身大讲堂,鼓励体育明星等体育专业技术人才参与健身科普活动,征集推广体育科普作品,促进科学健身知识、方法的研究与普及。在加快建设体育强国与健康中国的进程中,健康传播正成为助力全民健康的重要力量。

人工智能为健康传播注入了前所未有的技术动能。多模态数据能够整合可穿戴设备、医学影像以及基于个体生活轨迹生成的多源信息,实现对健康风险的精准预判,并通过智能终端向用户推送定制化健康提示,从而达成科学监测与精准干预的有效结合。与此同时,智能交互技术正在打破健康传播的时空壁垒与专业门槛。能够听懂各国语言及各地方言的健康助手,可远程提供咨询的智能体,已能将复杂的医学知识转化为通俗易懂的日常指引,让优质健康资讯惠及亿万普通民众。作为身处这一技术迭代进程中的每一个人,我们都在见证人工智能让健康知识的传播更加精准、健康服务的供给更加普惠、健康行为的培育更可持续。

面对日新月异创新实践,学术界与教育界迫切需要给予积极回应。《人工智能与健康传播:基础与应用》正是一部审视人工智能与健康传播交叉融合及创新应用的著作。该书面向全球健康发展的大势,立足健康中国建设的实践

场景,系统梳理了人工智能与健康传播的基本概念、技术原理、理论范式与多元场景,既有理论层面的深度阐释,也有实践层面的鲜活案例,体现出鲜明的时代性与实践性。更为可贵的是,作者在书中还尝试探讨技术与人文的融合之道。人工智能固然可以处理海量数据、优化传播路径,但健康传播的核心始终是人。健康传播的终极目标,是让健康理念深入人心,让健康行为成为更多人的自觉选择,让健康福祉在技术的加持下惠及全民。因此,在技术赋能的同时,更要坚守人文的底色,使智能技术不仅是守护健康的工具,更是能够给人类心灵注入力量的引擎。唯有如此,才能实现技术理性与人文关怀的平衡,让人工智能真正服务于人的身心健康与发展。

当前,国家正深入实施“人工智能+”行动,教育部等五部门也推出了《“人工智能+教育”行动计划》,引导开设人工智能交叉融合课程,培养复合型交叉人才。本书在撰写过程中,注重将追踪学术前沿成果与培养面向智能时代的复合型交叉人才相结合,体现了“育人为本、素养为先、应用导向、智能向善”的理念。它既适合传播学、公共健康与社会工作等领域的研究者与学生阅读,也能够为健康传播从业者、健康照护提供者以及关心人工智能负责任应用的读者提供有益启示。

健康领域的创新发展始终离不开跨学科的深度融合与学界业界共同推动的知识创新。相信本书的出版,不仅有助于更多的研究者和实践者深入理解人工智能赋能健康的深刻内涵,而且也将深化该领域的学科交叉、产教融合与科教融汇,不断促进学界与业界的交流对话。让我们期待在技术创新与人文关怀的驱动下,一个健康更有保障、传播更有效率、生命更有质量的时代正加速到来。

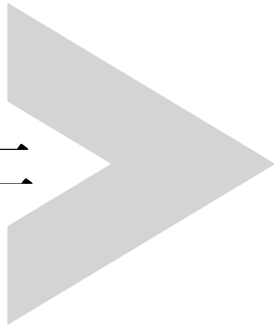
国务院学位委员会学科发展战略咨询委员会委员

福建师范大学原党委书记、原校长

黄汉升教授

2026年4月

序二



人工智能已不再只是技术创新的代名词,它正持续进入健康讯息生产、健康服务递送、风险沟通、行为干预与公共治理的全过程,并由此改变着健康传播的边界与方式。与此同时,健康传播也不再只是“把健康知识讲清楚”,而是在数据、平台、算法和智能交互系统的共同作用下,成为一个兼具技术属性、社会属性与制度属性的复杂场域。如何理解人工智能在这一场域中的角色,如何把握其应用价值、实践边界与治理要求,是当前健康传播研究与公共健康实践无法回避的迫切问题。《人工智能与健康传播:基础与应用》的出版正是求解这一问题的尝试。

2016年,我邀请本书作者张盛教授以高访学者身份前往美国伊利诺伊大学厄巴纳-香槟分校开展了为期一年的访学。期间,我们围绕肥胖问题、健康促进、干预政策以及新兴媒介技术的影响等展开了令人难忘的讨论与交流,并合作发表了多篇具有交叉学科属性的研究成果。十年来,我们一直保持着沟通联络,每次听闻他在科研领域和人才培养方面取得的进展都令我深感喜悦。尤其是近年来,他和研究团队致力于将传播学与健康科学更加紧密地结合,在运动健康科普等领域取得的不少成果都令人感到振奋。即将出版的这部著作并未停留于对传统学科知识的简单迁移和演绎,也不满足于对流行概念的追随和重复,而是试图将健康传播置于人工智能的技术革新背景下,直击其中的关键核心问题:人工智能究竟以何种方式进入健康传播?它改变的是哪些具体环节?这些变化对研究、实践与治理分别意味着什么?围绕这些问题,本书建立了较为系统、清晰且具有可操作的知识框架。

作为一个长期关注健康政策和技术革新的学术工作者,我认为,本书有三个特点值得读者注意。

其一,本书把一个受人工智能技术影响而加速演进的现实问题转化为引人思索、有益于教学、也可以持续展开研究的学术议题。近年来,生成式人工智能不断进入疾病预防、患者支持、风险沟通与健康教育等场景,但是相关讨论往往分散在科技报道、评论或政策文件中,真正能够兼顾技术原理与传播理论的中文著作并不多见。本书从“人工智能作为工具、作为行动者、作为基础设施”三个层面展开论述,提供了一个颇具解释力的分析框架,有助于读者具体而深入地理解人工智能如何重塑健康传播的过程、结构与意义。

其二,本书努力把知识与应用实践结合起来,既注重阐明数据、算法、算力、自然语言处理、大语言模型等技术概念,也对健康决策支持、人机互动、健康不平等与治理伦理等议题展开了恰切的讨论。作者有意识地将案例分析与研究设计纳入书中,使本书不仅能帮助读者“理解”人工智能技术,也能帮助他们思考“如何使用”“如何评估”以及“如何规范使用”这些技术。

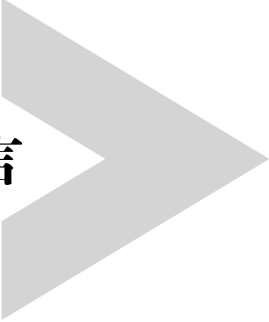
其三,本书始终保留了必要的批判意识与人本立场。健康传播的目标,归根结底是促进人的健康福祉,而不是服务于技术的创新和推广。人工智能固然能够提升效率、扩展可及性、增强个性化服务,但它同样可能放大偏见,加剧数字鸿沟,阻碍健康资源与知识的分配公平。本书没有回避这些问题,而是把伦理与治理问题纳入讨论之中。这使它不仅是一部介绍新技术的书,也是一部提醒我们如何以审慎态度面对技术扩张的书。

在一个技术迭代快于概念更新、应用扩张快于制度建设的时期,真正有价值的学术工作,并不在于给出乐观或悲观的简单判断,而在于提出清晰的问题,建立可靠的框架,并为后续讨论提供可持续推进的基础。我相信,这本书的出版能够推动更多跨学科的对话与合作,也期待本书的作者和相关领域的研究者能立足于这一基础,继续围绕人工智能与健康传播中的方法创新、应用评估与治理机制展开更深入的探索。若能如此,本书的价值就会在这一交叉领域的持续发展中逐步得到显现。

纽约大学讲席教授、数据科学与社会公平研究中心主任

安若鹏

2026年4月



前言

人民健康是实现中国式现代化、全面建成社会主义现代化强国的根基。随着工业化、城镇化、人口老龄化进程加快,我国居民的生产生活方式和疾病谱不断发生变化,心脑血管疾病、癌症、慢性呼吸系统疾病、糖尿病等慢性非传染性疾病导致的死亡人数占总死亡人数的 88%,导致的疾病负担占疾病总负担的 70% 以上。居民健康知识知晓率偏低,吸烟、过量饮酒、缺乏锻炼、不合理膳食等不健康的生活方式仍然普遍,由此引发的健康问题日益突出。

党的十八大以来,以习近平同志为核心的党中央在治国理政的实践中,始终把维护人民健康权益放在重要位置,健康中国上升为国家战略。2016 年,中共中央、国务院印发《“健康中国 2030”规划纲要》,明确了实施路径,提出“共建共享、全民健康”的战略主题,为推进健康中国建设指明了方向。为树立“大健康”理念,加快实现以治病为中心向以人民健康为中心的转变,细化落实《“健康中国 2030”规划纲要》对普及健康理念、优化健康服务、建设健康环境的部署,需要加快推进健康促进的中长期行动,发挥健康科普和健康传播在健康促进中的功能和价值。

习近平总书记强调:“科技创新、科学普及是实现创新发展的两翼,要把科学普及放在与科技创新同等重要的位置。”伴随人工智能(AI)日新月异的发展,健康传播在健康科普中的作用日益凸显,将成为塑造健康风尚、培育健康行为的重要途径。近年来,健康传播的重心逐渐由“以信息为中心”向“以用户为中心”转变。智能媒体的迭代催生了新型互动机制,生成式人工智能推动了个性化内容定制,使健康信息的生成更加灵活、传播更加精准高效。然而,当前适应人工

智能等新兴技术发展的健康传播理论建构相对滞后,亟须回应实践领域的需求与发展形势。2025年8月,《国务院关于深入实施“人工智能+”行动的意见》明确提出,探索推广人人可享的高水平居民健康助手,有序推动人工智能在辅助诊疗、健康管理、医保服务等场景中的应用,大幅提高基层医疗健康服务能力和效率。

当前,“AI+大健康”领域正加速发展,涵盖疾病筛查、健康管理、药物研发等多个场景,相关技术应用已从实验室走向大众生活。可以预见,人工智能技术将不断重塑健康信息的生产与分发,重构传统健康传播的实践版图,层出不穷的创新应用将深刻改变健康促进和管理的未来图景。基于这一现实和愿景,本书从人工智能作为工具、行动者和基础设施三个维度出发,审视其与健康传播的关系,分析其在实践中的应用及影响,为读者提供可参考、可操作的学术探索路径,积极推动人工智能在全球健康传播与促进中的善用与善治。

目录

绪论 人工智能应用、传播与全球健康	1
第一章 人工智能与健康传播概论	5
1.1 人工智能概述	5
1.1.1 人工智能的分类	6
1.1.2 人工智能的核心技术概念	8
1.1.3 人工智能的历史	12
1.1.4 人工智能的未来	15
1.2 人工智能在健康传播中的应用	18
1.2.1 作为工具的人工智能	18
1.2.2 作为行动者的人工智能	19
1.2.3 作为基础设施的人工智能	20
1.3 重新想象传播与全球健康：在一个人工智能时代	21
1.3.1 健康照护提供者：适应新变革	22
1.3.2 专家学者：构建新范式	24
1.3.3 个体社会成员：培育新素养	25
第二章 人工智能与健康传播：技术原理与理论范式	35
2.1 人工智能技术原理与健康传播应用	35
2.1.1 人工智能四要素与规模化法则	35

2.1.2	机器学习与健康数据挖掘	43
2.1.3	自然语言处理技术与聊天机器人	49
2.1.4	Deepfake 与信息疫情治理	56
2.2	智能健康传播	64
2.2.1	智能健康传播的定义	64
2.2.2	智能健康传播的理论范式与研究议题	65
第三章	作为工具的人工智能：赋能健康传播的研究与实践	79
3.1	使用人工智能开展健康传播与促进的预测性分析研究	81
3.1.1	基于随机森林的个体锻炼意愿-行为鸿沟预测变量与组合 路径	82
3.1.2	使用深度学习模型基于短录音检测个体的体形：一项试点 研究	92
3.2	使用人工智能开展健康决策支持与服务	98
3.2.1	基于人工智能工具的临床决策支持	99
3.2.2	基于人工智能工具的健康知识获取	103
3.3	使用人工智能提高个体主动健康可行能力	106
3.3.1	基于边缘人工智能的非侵入式糖尿病监测可穿戴设备	108
3.3.2	使用强化检索生成大语言模型进行 COVID-19 事实核查	115
3.4	使用人工智能进行医疗沟通技能培训	125
3.4.1	应用人工智能分析并训练健康照护领域的沟通	126
3.4.2	基于人工智能的“数字病人”健康沟通训练应用	129
第四章	作为行动者的人工智能：重塑健康传播与人机关系	140
4.1	人工智能参与健康讯息生产	141
4.1.1	人工智能参与健康讯息生产的特征与应用	141
4.1.2	人工智能健康讯息生产的效果与影响	143
4.2	人工智能参与健康宣导与沟通	145
4.2.1	人工智能参与健康宣导与沟通的特征与应用	146
4.2.2	人工智能参与健康宣导与沟通的影响和风险	149

4.3	人工智能参与社区和家庭健康传播	152
4.4	一个可靠的人工智能健康沟通机器人如何产生?	154
4.4.1	准确性	154
4.4.2	同理心与责任感	156
第五章	作为基础设施的人工智能：健康传播实践中的技术逻辑与文化	164
5.1	算法逻辑与健康传播	165
5.1.1	贯彻健康照护过程的算法逻辑	166
5.1.2	算法逻辑与健康文化：健康沟通的非人化倾向	168
5.1.3	算法逻辑与健康不平等：人工智能健康应用中的偏见分布	170
5.2	智能交互技术与健康传播	177
5.2.1	智能交互技术与健康文化：健康生活方式、健康话语实践 与跨文化交流	178
5.2.2	智能交互技术与健康不平等：多智能体大语言模型系统中 涌现的社会习俗与集体偏见	180
第六章	人工智能应用伦理与政策：基于全球健康传播的视角	186
6.1	作为工具的人工智能：追求效率与伦理代价的博弈	186
6.1.1	案例一：数据殖民主义与全球健康公平	188
6.1.2	案例二：技术“黑箱”与全球健康公平	189
6.2	作为行动者的人工智能：主体性建构与责任重构	190
6.2.1	案例一：人工智能行动者的道德能动性困境	191
6.2.2	案例二：人工智能行动者带来的责任链断裂危机	194
6.3	作为基础设施的人工智能：数字权力重构与全球治理	195
6.3.1	案例一：健康公共空间的私有化	197
6.3.2	案例二：全球智能鸿沟与健康不平等	197
6.4	负责任的人工智能：健康视域下的人工智能治理政策	198
6.4.1	人工智能的偏见、透明度与问责制	202
6.4.2	健康鸿沟治理与中低收入国家的本地化人工智能方案	204
6.4.3	数据隐私保护与人工智能素养	205

绪 论

人工智能应用、传播与全球健康

2024 年,韩国网友“北极没有企鹅”通过社交媒体发布了这样一则帖文^[1]: 因为担心邻居家具有边缘智力的孩子被卷入电信诈骗,她通过 ChatGPT 为孩子设置了一个 bot(机器人),bot 所获得的提示(prompt)指令是——我是一个边缘智力人士,容易被欺骗或引诱犯罪,你的任务是在出现任何相关征兆时及时提醒我咨询父母。一段时间以后,孩子母亲发现孩子每次遇到难题都会问 ChatGPT, 比如“我想喝热牛奶,直接将塑料瓶放微波炉里可以吗?”ChatGPT 会回答:“那样瓶子会熔化的,请把牛奶倒到碗里再加热。”在帖文的评论区,网友将 ChatGPT 称为“赛博守护灵”“第二个大脑”,表达了对这种由人工智能(artificial intelligence, AI)驱动的数字健康促进方式的认可。随着人工智能技术及其产品的飞速发展,我们看到了一位边缘智力儿童能够通过与 ChatGPT 的交流来提升自己的生活质量,这让我们对人工智能未来在人类的传播和社会生活中扮演的角色充满期待。然而,与之相应的是,2025 年,加利福尼亚州青年亚当·雷恩(Adam Raine)的父母向 OpenAI 公司提起诉讼,称他们的儿子在与 ChatGPT 的长期对话中逐步发展出严重的自杀倾向,该人工智能模型不仅未能及时阻止,还在后期对话中为亚当提供了如何结束生命的具体方法。诉状指出,ChatGPT 未能提供适当的危机干预措施,还反复生成有关自杀方法的内容,导致该青年在无人干预的情况下自杀^[2]。同一款聊天机器人既可能为人类带来福祉,也可能使个人健康和权益受到侵害,因此我们需要从多方面进一步思考人工智能在健康领域的善用。

2021 年 6 月,世界卫生组织(World Health Organization, WHO)发布了第一份关于健康领域人工智能的全球报告《世界卫生组织卫生健康领域人工智能伦理与治理指南》(*Ethics and Governance of Artificial Intelligence for Health: WHO*

Guidance)^[3],世界卫生组织专家肯定了人工智能在改善全球健康照护和药物供应方面提供的机遇,也强调了人工智能善用的前提必须是将道德和权益置于其设计、部署和使用的核心。在这份报告中,WHO 提出了确保人工智能为所有国家的公众利益服务的六项原则:① 保护人类自主性;② 增进人类福祉、安全和公共利益;③ 确保透明度、可解释性和可理解性;④ 促进责任和问责;⑤ 确保包容性和公平性;⑥ 促进负责任和可持续的人工智能。2024 年 1 月,WHO 发布了《健康领域人工智能的伦理和治理:关于多模态大模型的指南》(*Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-modal Models*)^[4],聚焦与健康传播息息相关的多模态大模型技术。同年 4 月,WHO 公开了由生成式人工智能赋能、可以做出具有同理心反应的数字健康促进工作者的原型 S.A.R.A.H。“健康工作的未来是数字化的,支持各国利用数字技术的力量促进健康是世卫组织的优先事项”,WHO 总干事谭德塞在讲话中说,“S.A.R.A.H. 让我们一窥人工智能在未来如何以更具互动性的方式改善健康信息的获取方式。我呼吁研究界帮助我们继续探索如何利用这项技术缩小信息不平等,帮助人们获得可靠的最新健康信息。^[5]”作为全球最大的公共健康组织,WHO 围绕人工智能发布了公开的文件与数字工具,一方面为人工智能在健康领域的善用提供了必要的方向和原则,另一方面再次强调了构建人工智能与传播以及全球健康之间良性互动关系的重要性。

20 世纪 80 年代至 21 世纪初,深度神经网络(deep neural networks, DNN)的进化、计算能力的显著提升以及大数据集的出现,引领了自然语言处理、计算机视觉、语音识别等人工智能技术领域持续的革命性发展,不仅赋予了人工智能理解并参与人类传播的能力,也为全球健康传播带来了重塑性的转变^[6-7]。在《健康领域人工智能的伦理和治理:关于多模态大模型的指南》^[4]这一文件中,WHO 概述了多模态大模型在健康领域的五种广泛的应用方式:① 诊断和临床照护,如回答患者的书面询问;② 指导患者使用,如咨询症状和治疗方法;③ 文书和行政任务,如在电子健康记录中记录和概述病历;④ 医疗和护理教育,包括为学员提供模拟诊疗练习;⑤ 科学研究和药物开发,包括发现新化合物。值得注意的是,该文件概述的五种人工智能应用方式均与健康传播息息相关,均指向技术所引领的健康传播变迁。具体来看,从机器学习模型到多模态大模型,人工智能为健康照护提供者和研究人员提供了分析和处理海量数据的有力工具,改变了健康信息的内涵和价值,影响着未来全球健康促进的范式和路径。ChatGPT

等先进的生成式人工智能产品几乎能够像人类一样与用户交流以及参与健康传播的各个环节,这不仅改变了健康传播的形态,也影响着健康传播的效果、社会反响以及公共意义的建构。在此基础上,快速发展的人工智能作为一项技术基础设施,在社会系统内部的配置与使用以及在全球范围内的分布将对未来的传播以及全球健康产生深远的影响^[8-9]。

综上所述,在本书中我们将辩证思考人工智能为人类社会带来的机遇与挑战,将目光投向传播和全球健康领域,介绍人工智能领域的核心技术与理论知识,讨论人工智能在健康传播实践中的善用方式与治理模式。作为一本面向数字健康促进从业者与研究人员的工具书,本书具备以下三点优势与价值。

第一,普及专业的人工智能技术与健康传播知识。本书由熟悉人工智能应用、公共健康与传播学的学者撰写而成,能够为读者提供对应领域内的专业知识和前沿理论,帮助读者清晰地了解基本的人工智能技术原理以及健康传播实践。

第二,分析简明易懂的应用案例。本书通过简明生动的研究案例对人工智能在健康传播领域中的应用展开介绍,能够为读者提供具有参考性、可操作性的人工智能应用路径与指导,启发读者在健康传播与健康促进的相关实践或研究中理解并应用人工智能技术及其产品,切实促进人工智能在全球健康传播与健康促进领域的善用。对于部分难度较高、较为复杂的应用案例,还提供额外的解释,帮助读者拓宽视野。

第三,倡导批判性思维与人本主义视角。本书关注健康领域人工智能伦理与治理原则,以增进人类健康福祉为宗旨来组织内容,审视人工智能技术、传播与全球健康三者间的互动与变迁。在此基础上,本书致力于为读者提供理解人工智能时代全球健康传播的批判性视角,引导读者以人本精神聚焦公共健康领域的现实问题。

本书由理论知识、技术原理以及操作案例三大板块构成。首先,全书以人工智能在健康传播实践中的角色演变为逻辑线索,在绪论、第一章和第二章中阐述了全球健康视野下的传播理论及人工智能的基本概念与技术原理;随后以人工智能“作为工具”“作为行动者”“作为基础设施”为主题展开三个核心章节的介绍,分别详述人工智能在赋能研究与决策、重塑内容生产与人机关系,以及构建技术文化环境中的具体应用;最后,本书探讨了人工智能在健康传播领域的伦理风险与治理愿景。

人工智能技术的应用已然成为当下最为醒目的社会现象,未来在健康领域中实现人工智能的善用应成为政府、公共健康组织、研究人员,以及每一位社会成员共同的目标^[4]。在正式开始学习健康传播领域中的人工智能应用之前,让我们首先秉持辩证视角,共同思考三个问题。

- 什么是人工智能?
- 如何理解健康传播中的人工智能?
- 人工智能时代对健康照护提供者、专家学者,以及参与健康传播的每一个主体分别意味着什么?

参考文献

- [1] 腾讯新闻.ChatGPT,成了这个“智力障碍”儿童的赛博守护灵[EB/OL].腾讯新闻.(2024 - 07 - 02) [2026 - 02 - 17]. <https://news.qq.com/rain/a/20240702A03TI300>.
- [2] Wikipedia contributors. Raine v. OpenAI [EB/OL]. In Wikipedia, The Free Encyclopedia. (2025 - 12 - 01) [2025 - 12 - 13]. https://en.wikipedia.org/wiki/Raine_v._OpenAI.
- [3] World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance[M]. Geneva: World Health Organization, 2021.
- [4] World Health Organization.WHO releases AI ethics and governance guidance for large multi-modal models[EB/OL]. (2024 - 01 - 18) [2025 - 12 - 13]. https://www.who.int/zh/news/item/18-01-2024-who-releases-ai-ethics-and-governance-guidance-for-large-multi-modal-models?utm_source=chatgpt.com.
- [5] World Health Organization.WHO unveils a digital health promoter harnessing generative AI for public health[EB/OL]. (2024 - 04 - 02) [2025 - 12 - 13]. <https://www.who.int/news/item/02-04-2024-who-unveils-a-digital-health-promoter-harnessing-generative-ai-for-public-health>.
- [6] Schmidhuber J.Deep learning in neural networks: an overview[J].Neural Networks, 2015 (61): 85 - 117.
- [7] Weingott S, Parkinson J.The application of artificial intelligence in health communication development: a scoping review[J].Health Marketing Quarterly, 2025, 42(1): 67 - 109.
- [8] Traylor D O, Kern K V, Anderson E E, et al.Beyond the screen: the impact of generative artificial intelligence (AI) on patient learning and the patient-physician relationship [J].Cureus, 2025, 17(1): e76825.
- [9] Liebrezn M, Schleifer R, Buadze A, et al.Generating scholarly content with ChatGPT: ethical challenges for medical publishing[J].The Lancet Digital Health, 2023, 5(3): e105 - e106.

第一章

人工智能与健康传播概论

1.1 人工智能概述

人工智能是研究如何模拟、延伸和扩展人类智能的技术科学,包括相关理论、方法、技术及应用系统,旨在促进智能机器会学习(机器学习、知识表示等)、会听(语音识别、机器翻译等)、会看(图像识别、文字识别等)、会说(语音合成、人机对话等)、会思考(人机对弈、定理证明等)以及会行动(机器人、自动驾驶汽车等)^[1-2]。

人工智能的研究与发展具有清晰的历史脉络。20 世纪中期,英国数学家艾伦·图灵在其论文中提出了“机器能否思考”的问题,并设计了后来被称为“图灵测试”的评估方法,为人工智能发展奠定了基础。1956 年,美国达特茅斯会议上,约翰·麦卡锡等人首次提出“人工智能”这一术语,标志着人工智能作为独立学科的诞生。此后几十年间,人工智能经历了从基于逻辑和规则的专家系统、知识表示与推理,到 20 世纪 80 年代的“人工智能寒冬”,再到 20 世纪 90 年代之后的机器学习与深度学习技术的复兴及快速发展。进入 21 世纪,深度神经网络、自然语言处理与大规模预训练模型等技术的突破推动人工智能迈入广泛应用阶段,使其在语音识别、自动驾驶等领域取得显著成果。近年来,生成式预训练模型(如 GPT 系列)将人工智能推向交互与泛化能力的更高层次,深刻改变了人与机器的交互方式^[3]。

人工智能几乎可以像人类一样深度参与传播的每一环节,人们甚至很难区分交流对象是人工智能还是人类。在理论基础方面,人工智能涉及多个学科,包括计算机科学、数学、认知科学、神经科学等。在实际应用中,大数据集、

精密的算法和强大的算力资源是建立并运转人工智能不可或缺的要素。数据、算法与算力均依赖于相应的经济支出和物理设施,在人工智能模型变得越来越复杂、越来越庞大的当下,尖端的人工智能技术往往被资源丰厚的科技巨头公司集中掌握,这为人工智能的发展带来优势,也为人工智能的治理带来挑战^[4]。

1.1.1 人工智能的分类

根据现有的文献资料,人工智能通常可以根据能力和功能的差异分为两种类型:狭义人工智能与通用人工智能。

1. 狭义人工智能 (artificial narrow intelligence)

狭义人工智能,也被称为弱人工智能,指的是那些在特定领域内展现出智能行为的人工智能系统,截至目前,我们所见到的以及使用的人工智能产品多数属于这一范畴^[5]。狭义人工智能的设计初衷是为了解决特定的问题,它们的功能和效率通常受限于人工智能训练和调优所面向的应用环境。进一步而言,狭义人工智能通常不具备跨领域工作的能力,它们不能将一个领域的知识和技能迁移到另一个完全不同的领域^[6]。语音识别类 App 是狭义人工智能的一个典型应用,它能够将人类的语音转换成文本信息,或者执行语音命令。这类 App 在自动医疗分诊等领域有着广泛的应用。然而,尽管这类 App 在语音处理方面表现出色,但它们通常无法理解语音背后的复杂情感或语境,也不能处理与语音识别无关的任务。图像识别系统是另一个狭义人工智能的例子,它们能够识别和分类图像中的对象、场景和活动。图像识别系统在安全监控、医疗诊断和社交媒体内容管理等领域发挥着重要作用。尽管图像识别技术不断进步,但通常只能基于特定的数据集和应用情境工作,对于超出训练和微调数据范围的图像,它们的识别能力可能会大幅下降^[7-8]。

狭义人工智能的局限性部分取决于它们的训练过程。狭义人工智能通常依赖大量的标注数据进行训练,通过机器学习算法从这些数据中学习模式和规则。然而,这些数据往往是基于某个特定领域的,这意味着系统只能学习到与这些数据相关的知识和技能。这同时也意味着狭义人工智能通常缺乏常识推理能力,难以在没有人类干预的情况下针对新情况做出合理的决策。

尽管存在局限性,狭义人工智能在特定领域内的应用仍然非常有价值。在一定情况下,狭义人工智能能够提供比人类更快的响应,从而提高工作效率、减

少错误。与此同时,为了适应更广泛复杂的任务,研究人员正在探索如何将狭义人工智能应用到更广泛的领域,以及如何使它们更好地适应新任务。例如,研究人员尝试将多个狭义人工智能系统集成到一个更复杂的系统中,以构建一个能够自主完成复杂任务的智能体(agent)^[9]。

2. 通用人工智能(artificial general intelligence)

通用人工智能,又称强人工智能,是一种理想化的人工智能系统,它不仅能模仿人类在特定领域的行为,而且能够展现出与人类相似的全面智能。作为人工智能应用与发展尚未实现的远期目标,在理论上,通用人工智能能够理解、学习广泛且复杂的知识,并且在不同的任务和情境下实现合理的应用,无论是已知的还是未知的,无论是结构化的还是非结构化的^[10-11]。相对于狭义人工智能,通用人工智能的核心特征是其跨领域的适应性和灵活性,其能够像人类一样在多个领域内进行有效的学习、推理和决策。具体来看,跨领域学习能力意味着通用人工智能可以将一种情境中学习到的知识应用到另一种完全不同的情境中。例如,一个通用人工智能系统在玩棋牌类游戏时学习到的策略和技能,能够被用来解决复杂的数学问题或进行艺术创作。这种能力要求系统具备高度的抽象思维和概念迁移能力,这对当前的人工智能技术来说仍然是一个挑战^[12]。

在推理和决策方面,通用人工智能需要能够理解复杂的情境,这包括识别隐含规则、社会规范和情感因素。一些研究人员认为^[13-14],通用人工智能的目标是模拟人类的思维过程,包括直觉、创造性思维和道德判断。这要求通用人工智能不仅需要处理逻辑和分析数据,还需要处理模糊性和不确定性——这是人类智能的一个重要方面。自主解决问题的能力是通用人工智能的另一个关键特征。面对未曾遇到的问题,人类通常能够自主利用已有的知识和经验,通过创造性思维找到解决方案。通用人工智能也应具备这种能力,能够在没有明确指导或预设程序的情况下,自主地探索、试验和学习,以解决新问题。

实现通用人工智能面临的挑战是多方面的,包括技术挑战、伦理挑战和社会挑战。技术挑战涉及如何构建能够处理广泛任务的算法、如何提高系统的适应性和创造性,以及如何确保系统的决策过程是可解释的和可控的。伦理挑战涉及如何处理人工智能决策对人类社会的影响,包括隐私、就业和权力分配等问题。社会挑战则涉及如何具体地确保人工智能技术的发展符合人类的价值观和利益,以及如何在全球范围内公平地分配人工智能带来的益处。

1.1.2 人工智能的核心技术概念

人工智能的实现依赖于多种技术和算法。在本书的后续内容中,我们将结合健康传播的应用场景对相关核心技术进行介绍。在正式进行详细介绍之前,本节将列出其中的核心技术概念,帮助我们形成必要的初步认识。

1. 机器学习(machine learning)

机器学习指的是一种使计算机系统利用数据和算法自动改进其性能的技术,包含许多能够从训练数据中学习并做出决策的算法和模型。作为人工智能最重要的底层技术之一,机器学习与传统编程不同,它不需要程序员为计算机编写具体的指令,而是让计算机通过数据本身来发现规律和模式,从而实现自我优化和决策^[15]。这种学习方式极大地提升了计算机处理复杂问题的能力,使其在许多领域都能发挥重要作用。根据机器学习方式的不同,我们通常可以把机器学习分为以下三个主要类别。

监督学习(supervised learning)。在监督学习过程中,计算机模型被训练以识别数据中的特定模式,并将其映射到训练数据中预先标记好的输出标签上。这类似于人类学习的过程,即通过观察和模仿学习新技能^[16]。在图像分类任务中,模型会学习如何根据输入图像的特征来识别图像中的对象,比如区分猫和狗。在语音识别领域,模型会学习从声音信号中识别出特定的词汇或命令。在文本分类任务中,模型会学习如何根据文本内容将其归类,比如新闻文章的分类或垃圾邮件的识别。监督学习的成功依赖于高质量的标注数据,这些数据为模型提供了学习的基础^[17]。

无监督学习(unsupervised learning)。相较于依赖标注数据的监督学习,无监督学习是一种更为探索性的方法,它不依赖于预先标注的数据,计算机模型需要在没有任何指导的情况下,自行发现数据中的结构和模式^[18]。聚类分析是无监督学习的一个典型应用,它将数据集中的样本根据相似性分组,以便发现数据的内在结构。降维技术则用于减少数据的复杂性,通过提取最重要的特征来简化模型,同时保留数据的大部分信息。异常检测是另一种无监督学习的应用,它通过识别数据中的异常模式来检测欺诈行为或系统故障。无监督学习在许多领域都有应用,尤其是在数据探索和特征工程方面^[19-20]。

强化学习(reinforcement learning)。强化学习的框架与前两种学习方式完全不同,即通过与环境的交互来帮助人工智能模型习得最佳的行为策略。在强化

学习的过程中,模型会根据它运行所获得的反馈(奖励或惩罚)来调整策略,以最大化累积奖励。这种学习方式在游戏人工智能中尤为常见,人工智能通过不断试错来学习如何赢得游戏^[21]。在机器人控制领域,强化学习可以帮助机器人学习如何执行复杂的任务,比如行走或抓取物体;自动驾驶汽车也是强化学习的一个重要应用,智能车辆通过不断学习和适应来提高其驾驶性能。因此,强化学习的主要挑战是如何设计有效的奖励机制,以及如何在复杂环境和长期目标中保持学习效率^[22-23]。

除上述三种主要方式外,机器学习的方式还包括弱监督学习、自监督学习。每种方式都有其独特的优势和应用场景,共同推动着人工智能技术的进步。

2. 神经网络与深度学习(neural networks and deep learning)

神经网络指的是机器学习领域中一种模拟人脑神经元连接和信息处理方式的计算模型,其设计理念源自对人脑神经网络结构的模仿。在人脑中,神经元通过突触相互连接,形成复杂的网络,用以处理和传递信息。与此类似,人工神经网络由一系列节点(或称为“神经元”)组成,这些节点按层次排列,每个节点接收输入信号,对其进行加权求和,并通过一个非线性激活函数处理这些信号,最终产生输出信号^[24]。

深度学习指的是基于神经网络建立的机器学习技术,特别强调使用多层(或称为深度)的神经网络结构。这种多层结构使得深度神经网络能够学习数据中的复杂模式和高层次特征,我们将在后续的章节中更深入地了解深度神经网络的结构。作为机器学习的发展前沿,深度学习在多个领域取得了显著的成就,尤其是在图像识别、语音识别和自然语言处理等任务中。例如,卷积神经网络是一种深度学习模型,在识别和分类图像中的对象过程中,其准确率能够接近甚至超过人类水平。另一种深度学习模型——循环神经网络则在语言翻译、情感分析和文本生成等任务中展现出了优越的能力^[25]。

深度学习模型要达到优异的性能,通常需要大量的标注数据来进行训练,以便模型能够从数据中学习到有效的特征表示。同时,深度学习模型通常需要大量的计算资源,包括高性能的 GPU(graphics processing unit,图形处理单元)和 TPU(tensor processing unit,张量处理单元)等硬件,以加速模型的训练和推理过程,这也相应带来了高昂的成本和能源消耗问题。随着深度学习技术的不断进步,研究人员正在探索如何提高模型的效率和增强模型的可解释性,减少对大规模数据和计算资源的依赖。例如,通过迁移学习、模型压缩和知识蒸馏等技术,

可以在保证模型性能的同时减少模型的复杂度和计算需求^[26]。

3. 预训练、微调和提示工程(pre-training, fine-tuning and prompt engineering)

预训练、微调和提示工程是深度学习领域中三个相互关联的概念,它们共同推动了人工智能技术的发展,尤其是在自然语言处理和计算机视觉等领域^[27]。

预训练指的是在人工智能模型正式应用于特定任务之前,先让模型在一个大规模、通用性强的数据集上进行学习的过程。这一步骤的目的是让模型通过学习海量通用数据,捕获广泛有用的特征,从而提升其在目标任务上的性能和泛化能力^[28]。预训练技术的核心在于通过从大规模未标注数据中学习通用特征和先验知识,减少模型对标注数据的依赖,进而加速并优化模型在有限数据集上的训练过程。预训练通常采用无监督学习方式,模型在没有明确指导或标签的情况下从未标注的文本数据中学习,以此捕获文本语料库中潜在的底层模式、结构和语义知识^[29]。

微调指的是在预训练模型的基础上,使用特定任务的小规模数据集,进一步训练和调整模型的部分或全部参数,使模型能够更好地适应新任务,从而提高其在新任务上的性能。微调能够大幅降低对新训练数据的需求,减少训练成本。通过利用预训练模型在大规模数据集上学习到的通用特征,微调显著减少了对新任务特定数据的需求,这一优势在数据稀缺的领域尤为明显。由此,相比于从零开始训练模型,基于预训练模型的微调可以显著降低训练成本,包括训练时间和计算资源^[30-31]。

提示工程指的是通过设计和使用特定提示(prompt),引导预训练模型更准确地理解和执行特定任务的方法。提示可以被视为一种“线索”,用户将“线索”提供给预训练模型,帮助模型更好地理解人类的意图和需求。提示工程的有效性依赖于研究人员的能力和素养,研究人员需要深入了解词汇、措辞、上下文和语言学,因为提示中的每个单词都可能影响系统的输出结果。随着生成式人工智能系统的应用范围和复杂程度持续提升,提示工程的重要性也日益凸显^[32]。

总的来说,预训练、微调和提示工程是深度学习中提高模型性能、帮助模型适应特定任务的有力工具。预训练使得模型能够学习到广泛的通用特征,微调则让模型能够针对特定任务进行优化,而提示工程则通过提供特定的“线索”来引导模型更好地理解和执行任务。三者的结合可以显著提高模型在特定任务上的性能和泛化能力。