



Editorial

迈向更通用赋能的人工智能

吕跃广^a, 吴飞^b^a Department of Information and Electronic Engineering, Chinese Academy of Engineering, Beijing 100088, China^b College of Computer Science, Zhejiang University, Hangzhou 310027, China

吕跃广



吴飞

近十年来,人工智能犹如历史上蒸汽机、电力、计算机和互联网等通用技术一样,正以史无前例的速度深刻地改变人类社会和促进经济发展。这场人工智能革命为释放人类创造力和促进经济增长提供了巨大的机会,使工人不再从事最危险和最卑微的工作。此外,许多研究人员正在寻求更通用的人工智能系统,这些系统在一些智力任务中与人类智能相当——这种努力被称为通用人工智能(AGI)。换句话说,更通用的人工智能系统可以与人类智能相当的水平理解、学习和执行任务[1–4]。

近年来,一些大型语言模型(如OpenAI的ChatGPT和Google的PaLM)已被证明在各种领域和任务中比以前的AI模型表现出更通用的智能。这些语言大模型可产生新颖甚至是意料之外的结果——这与早期仅仅只会生硬产生可预测结果的人工智能模型有很大不同。然而,语言大模型严重依赖统计模式识别和机器学习算法,因此难以完

成需要深入理解上下文的任务。此外,语言大模型所合成的内容有时包含非事实信息(这被称为人工智能幻觉)。从真正的因果解释的角度来看,语言大模型是现象级人工智能产品[5]。

随着人工智能产品不断与人类社会深度结合,有必要考虑应用通用人工智能系统应该遵从的道德规范。因此,开发和部署值得信赖、可靠和可解释的人工智能应放在优先位置,以确保公平分享人工智能进步所带来的益处并减轻潜在风险。我们认为,AGI的目标不仅是技术通用,更是让智能技术更好地赋能社会、造福全人类[6]。追求更通用的人工智能系统是一项复杂且具有挑战性的工作,但对人类的潜在益处可能是变革性的。例如,更通用的人工智能系统可改变医疗保健、教育、交通和制造等领域服务和生产模式,还可以带来更高效、更准确的疾病诊断和治疗,并帮助发现新科学知识、助力个性化学习体验以及实现能够适应动态变化需求的全自动工厂。

本期人工智能专题报道了最近对更通用人工智能思考的研究,包括理论研究、算法模型和实用方法。潘云鹤在题为《群体智能系统的结构分析》的文章中分析了不同类别群体智能系统结构的一般特性,对三类重要群体智能系统的结构、机制和实例进行了充分讨论。在《从计算机科学理论审视意识和通用人工智能》一文中,Lenore Blum和Manuel Blum讨论了意识图灵机(CTM),旨在研究理论计算机科学(TCS)中有关意识的问题,CTM为AGI的创建提供了深思熟虑的思考。

袁路遥和朱松纯的《通讯式学习——统一的机器学习模式》一文提出了一种通讯学习（CL）的形式化方法，它统一了现有的机器学习范式，如被动学习、主动学习和算法学习等，并促进新机器学习的发展。这种形式主义将学习视为一种交流过程，并将教育学与新兴的机器学习领域相结合。王海峰等的《预训练语言模型及其应用》一文指出预训练语言模型正在引领从监督学习到“预训练+微调”的机器学习范式转变，该文全面回顾了自然语言处理（NLP）领域的代表性工作、最新进展和未来挑战。

方璐等的《记忆成像》一文提出了一种称为记忆驱动摄像的双通道成像机制，成像过程在激发-抑制和关联状态之间交替进行工作。在激发-抑制状态下，像素级细节会动态巩固或受到抑制，以加强对物体实例的记忆。在关联状态下，空间和时间一致的内容在其记忆驱动下被合成。黄铁军等的《基于普通器件实现快1000倍的相机与机器视觉》一文提出了Vidar概念，以记录和重建任意特定时刻的场景辐射亮度。

魏颖等的《人工智能在化学逆合成中的应用》一文综述了人工智能技术的逆合成预测技术，可自动从现成的实验数据集中学习化学知识，以预测化合物反应和逆合成路线。

近年来，数据已成为数字经济最重要的资源之一，《面向数据权利、数据定价和隐私计算的数据驱动学习》

一文回顾机器学习技术中的数据权利、数据定价和隐私计算问题。以人工智能造福人类社会为长远目标，人工智能算法将接受数据保护法规（即隐私保护法规）的评估，帮助构建值得信赖的用于日常生活的人工智能系统。

感谢审稿专家对本期论文提出的宝贵评审意见，感谢王真（加拿大英属哥伦比亚大学）、陈怡然（杜克大学）、黄恒（匹兹堡大学）、刘小明（密歇根州立大学）、韩军功（亚伯大学）、张含望（南洋理工大学）、苏昊（加州大学圣地亚哥分校）、Fisher Yu（苏黎世联邦理工学院）、刘军伟（香港科技大学）、李学龙（西北工业大学）、姜育刚（复旦大学）、李玺（浙江大学）、卢策吾（上海交通大学）、朱军（清华大学）、乔宇（上海人工智能实验室）、李纪为（浙江大学）和况琨（浙江大学）等专题编委的辛勤工作。

References

- [1] Fjelland R. Why general artificial intelligence will not be realized. *Humanit Soc Sci Commun* 2020;7:10.
- [2] Pan Y. Heading toward artificial intelligence 2.0. *Engineering* 2016; 2(4): 409–13.
- [3] Pan Y. Special issue on artificial intelligence 2.0. *Front Inform Technol Electron Eng* 2017;18(1):1–2.
- [4] Pan Y. 2018 special issue on artificial intelligence 2.0: theories and applications. *Front Inform Technol Electron Eng* 2018;19(1):1–2.
- [5] Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, et al. Sparks of artificial general intelligence: early experiments with GPT-4. 2023. arXiv:2303.12712.
- [6] Lyu YG. Artificial intelligence: enabling technology to empower society. *Engineering* 2022;6(3):205–6.