

DeepSeek呼唤DeepThink：重视AI治理与社会范式变革

王飞跃^{1,2,3}

(1. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室, 北京 100190;

2. 澳门科技大学创新工程学院工程科学系, 澳门 999078;

3. 欧布达大学, 匈牙利 布达佩斯 1132)

摘要: DeepSeek无疑将加速人工智能技术的发展, 必须加快有关人工智能治理的相关讨论与研究, 本文将就此进行简要探索。

关键词: 人工智能治理; 深度求索; 深度思考

中图分类号: TP39

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202501

DeepSeek calls DeepThink: rethinking AI governance and societal paradigm shift

WANG Fei-Yue^{1,2,3}

1. The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

2. Department of Engineering Science, Faculty of Innovation Engineering, Macau University of Science and Technology, Macao 999078, China

3. Óbuda University, Budapest 1132, Hungary

Abstract: Undoubtedly, DeepSeek would speed up AI research, development, and applications. However, we need to think deeply and must intensify our effort in AI governance related investigation and implementation before important issues become out of control.

Key words: AI governance, DeepSeek, DeepThink

0 引言

2025年开局, 我们惊喜地迎来了国产人工智能大语言模型 (large language models, LLMs) DeepSeek 引发的世界性智能科学与技术的热潮。毫无疑问, 这是当今社会, 特别是开源路线的一个伟大成就。但客观而言, 尽管是一项使世人振奋的巨大成果, 但从创新角度, DeepSeek的科技价值小于其商业价值, 商业价值又小于其经济价值, 依次还小于其社会和政治价值。无论如何, 在当下的国际政治局势之下, DeepSeek的成功为国际化和人类的和平向善之合作意愿注入了新的力量和信心。

DeepSeek不但打破了OpenAI的神话, 还使其

承认“ClosedAI”路线之错误, 为开源技术和开源社区提供了强力支撑, 在世界范围, 特别是欧美日地区, 造福了一大批无力于大算力和复杂算法拼杀的中小企业。然而, 未来的LLMs依然还要大幅度地降低成本和复杂性, 同时大幅度地提高性能和准确性, 不可以再以“幻觉 (hallucination)”之词掩饰其本质上的“无能”或“愚蠢 (stupidity)”。而且, 人们还应为“知识蒸馏 (knowledge distillation)”正名, 因为这实际是人类学校教育和师生传承的一种技术变形、文学上“比兴”的结果, 正当且合理。我们必须不断地“知识蒸馏”, 从Pre-training、Post-distilling到Fine-tuning, 不停地蒸, 直到蒸除一切“幻觉”。

1 DeepSeek的警示

对笔者而言, DeepSeek最大的启示是在AI的研发过程中, 必须强调DeSci(即分布式去中心化的自主科学研究)的重要作用。人们习惯熟习的CeSci, 它是集中式中心化的组织计划科学研究。但如图1所示, AI的研究必须根植于多样化的土壤(D阶段), 在此基础上涌现出各种各样的有用科技(E阶段), 最后向社会对科技的目标收敛(C阶段)。

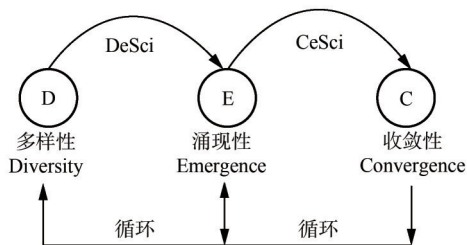


图1 DeSci和CeSci与人工智能技术研发的DEC阶段

从D阶段到E阶段, 应以DeSci为主, 体现科学本质, 而从E阶段到C阶段, 应以CeSci为主, 体现社会意志。英国的AlphaGo, 美国的ChatGPT, 现在中国的DeepSeek, 都是DeSci的产物, 几乎与CeSci无关, 这值得人们深思并有所行动。

2 DeepSeek呼唤DeepThink

笔者近期在中国科学院哲学研究所就AI与伦

理及治理问题进行过讨论, 如图2所示。我们还曾在上期的“评论智能”文章《新AI与智能科技前沿: 从PAPAI++到新产业生态》中指出, AI正从“老AI”“旧AI”, 向“新AI”发展, 形成“老、旧、新”3种AI平行之势, 其必将带来深刻的社会变革^[1]。而DeepSeek的出现, 必将加剧这一变革, 人们必须深刻思考如此情势之下可能带来的社会冲击并计划应对策略。

首先, 人们大可不必扩散AI使人类失业甚至AI奴役人类等言论。工业时代不断发展, 其日益有效的机器, 不仅没有让人类失业, 反而为人类提供了比农业社会更多、更好、更合适的工作, 这就是历史性的证明。在即将到来的智业时代, AI越发达, 就越需要更多的人工, 而且让过去不是工作的事情也变成可能获得合理报酬的职业, 如快递工作等。

人们首先应该担心的是会用AI的人将取代不会用AI的人, 其实还是人类取代人类的问题, 不是AI取代人类的问题。其次, 人们更应该担心的是, 由于AI Agents的出现, 社会上的富裕阶层从技术上无须像过去一样依靠不富裕阶层, 一时甚至较长时间内可能产生各种需求与供应失衡, 并产生社会矛盾、引发社会动乱的因素增加且更复杂。因此, 人们必须加快对人工智能治理(AI governance, AIG)技术的研究和实验工作。



图2 王飞跃主编在中国科学院哲学研究所讨论AI与伦理及治理

Pascal的“新赌注”认为AI不可解释,但可治理,且必须治理^[2]。但应该如何治理?不少AI狂热者认为“AI Agents”是比原子弹还厉害的新发明,他们想过人们是如何看待并治理原子弹的吗?其实,大家不妨回顾人们对待常规武器的态度。农业社会时,猎枪等武器可随意持有,而工业社会之后,多数文明或发达国家禁止或严格管制枪支。尽管AI Agents现在还不及原子弹的威力,但毫无疑问,其效用将比一般武器的作用更大。人们可轻易地通过AI Agents组建个人的无人团队,甚至可以做到规模很大的部队。利用组织化的AI Agents之“新AI”,甚至已经可以从事许多新型且后果难以估量的经济、社会和军事活动,因此AIG应尽快提上日程。目前各国展开的AIG措施,特别是欧洲的有关行动,方向正确,但离AIG的真正效用依然差得太远。

笔者认为,AIG必须从源头和教育开始,但不可局限于传统的立法与政治手段,必须以新的技术手段为支撑,比如区块链、智能合约、DAOs/TAOs等新的AI方法与体系。具体而言,就是针对未来成熟的AI Agents,即数字人和机器人,从“出生”到“上岗”进入社会,必须给予严密且科学的支持、教育和管控。

比如,AIG之下的AI Agents生成必须通过规范的数字学校、场景检测、数字社会之DAO,它们分别是今日之对抗生成、Pre-training、Post-distilling和Fine-tuning的升级和系统化。AI Agents在数字学校生成,并由各种场景工程进行测试,达到一定水平才可以进入DAO的组织之中,成为社会注册的可管、可控、可操作之数字人。

而AIG之下的AI Agents上岗服务于社会,必须通过更加规范且长期的平行学校、场景考核、平行社会之TAO,它们分别是今日学校、考核机制、社会管理的智能化升级和数智平行化的结果。换言之,未来的数字人将经过与人类相同的教育,其价值观将与人类对齐,其服务方式也面向人类的需求。人们将利用区块链、智能合约和DAO,使数字人“TRUE”,即可信(trustable)、可靠(reliable)、可用(useful)、以正确的方式做正确的事(effective/efficient)。

AIG是一项长期的过程,我们必须付出长期的努力并持续关注,欢迎各位作者向《智能科学与技术学报》就相关问题投稿,共同推进相关研究,造福人类命运共同体。

参考文献:

- [1] 王飞跃. 新AI与智能科技前沿: 从PAPAI++到新产业生态[J]. 智能科学与技术学报, 2024, 6(4): 413-415.
WANG F Y. New AI for new frontier of intelligent science and technology: from PAPAI++ to new enterprise ecology[J]. Chinese Journal of Intelligent Science and Technology, 2024, 6(4): 413-415.
- [2] WANG F Y. Pascal's new wager: AI is not explainable, but governable[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(9): 4147-4150.

[作者简介]



王飞跃(1961-),男,博士,中国科学院自动化研究所复杂系统管理与控制国家重点实验室主任,澳门科技大学特聘教授,主要研究方向为平行系统的方法与应用、社会计算、平行智能、知识自动化。