医学科学报 Medical Science News

主管单位：中国科学院
主办单位：中国科学报社
学术顾问单位：
中国人体健康科技促进会
国内统一连续出版物号：CN11-0289

地址：
北京市海淀区中关村南一条乙3号
邮编：100190
编辑部电话：010-62580821
发行电话：010-62580707
邮箱：ykb@stimes.cn

广告经营许可证：
京海工商广登字20170236号
印刷：廊坊市佳艺印务有限公司
地址：
河北省廊坊市安次区仇庄乡南辛庄村
定价：2.50元
本报法律顾问：
郝建平 北京灏礼默律师事务所

院士之声

梅宏：牢记科技向善、以人为本

◉本报记者 高雅丽 见习记者 江庆龄

“智能是人类区别于其他动物的专有特征，特别是认知能力。我们可以接受机器在感知能力方面超越人类，但对机器认知智能的研发，需要高度审慎。从科技伦理的视角来看，我们为什么要造出一个替代我们认知的东西？”

近日，在以“跨越边界的科技伦理”为主题的第二届中国科技伦理高峰论坛上，中国科学院院士、北京大学教授梅宏直指当前数字技术面临的伦理挑战。

从智能制造到智慧城市，从医疗健康到金融服务，当前大语言模型火爆，人工智能（AI）风头正盛。梅宏认为，在AI热潮中，泡沫太大，仍处于技术成熟度曲线的高峰阶段，喧嚣埋没理性，需要一个冷静期。

“手机读不到有价值的内容”

从脸书公司（Facebook）数据泄露到大模型生成内容引发侵权纠纷和虚假信息传播，数字技术在为人类社会经济发展带来益处的同时，也带来了数据隐私保护、算法偏见、责任认定等一系列伦理问题。

例如，就当前大语言模型的技术路线而言，“黑盒”导致的不可解释性是其最大“罩门”。如果不加任何规制而大量应用，可能导致人类知识体系面临严峻挑战。训练语料的质量缺陷、概率统计的内生误差等因素会导致大模型产生幻觉，生成错误内容；再加上人为干预诱导，极易生成虚假内容。

“通过算法，平台可以个性化推送内容，但也可能形成用户的信息茧房。我最近最大的困扰就是拿着手机却读不到想读的有价值的内容。”梅宏表示，目前几乎大部分网络平台都在AI算法和大数据驱动下运营，这就带来对算法和数据应用的有效监管问题，这些亟待通过建立完善的治理体系加以解决。当涉及平台跨境时，还需要有相应的国际治理体系。

现实情况不尽如人意

当前，社会对“AI+”或“AI for everything”（一切皆人工智能）抱有很高的期望，然而，现实情况却不尽如人意。

梅宏

“雷声隆隆，雨点并不大。”梅宏坦言，“从当前的热潮中，我看到了太多‘炒作’和‘非理性’导致的AI‘过热’现象，也对当前AI发展技术路径多样性的欠缺产生了一些担忧。”

“大语言模型的成功依赖于人类长时间积累的庞大语料库，文生视频的成功也依赖于互联网上存在的海量视频。然而，其他行业的数据积累尚未达到这个量级。获取全数据，关键要跨越足够的时间尺度。”梅宏表示，AI的应用还需要经历一段时期的探索、磨合和积累，才可能迎来繁荣。

“在我看来，AI当前的问题有3个：泡沫太大，仍处于技术成熟度曲线的高峰阶段，喧嚣埋没理性，需要一个冷静期；以偏概全，对成功个案不顾前提地放大、泛化，过度承诺；期望过高，用户神化AI的预期效果，提出难以实现的需求。”梅宏说。

面对AI技术发展及其应用的现状，梅宏建议，在尚搞不清如何应用、用到何处时，不妨先积累数据，“可采尽采、能存尽存”。

没有跳出概率统计框架

那么，大语言模型能走向通用人工智能吗？梅宏认为，从基本原理来看，目前的大语言模型没有跳出概率统计这个框架。

梅宏并不认为现在的AI有所谓“意识”或者知识涌现能力。以大语言模型为例，模型本身无法产生新的东西，其生成的内容取决于对大量文本内容的统计，如果某些内容反复出现，它们大概率就会将之视为“合理存在”的内容。

“就这个意义而言，大模型可被视为是由已有语料压缩而成的知识库，生成结果的语义正确性高度依赖于数据的空间广度、时间深度以及分布密度，更高度依赖于数据的质量。”他提到，学术界的研究更应关心大模型构建过程的可重复性和可追溯性，尽可能保证结果的可解释和可信任。

“大胆预测，作为压缩了人类已有的可公开访问的绝大多数知识的基础模型，大语言模型将像互联网一样走向开源。全世界共同维护一个开放共享的基础模型，尽力保证其与人类知识同步。”梅宏说，“这也是自己的一种期望。”

调整伦理审查复核清单

“发明技术的最终目的是为了让人类的生活变得更好，毫无疑问应该充分考虑技术可能带来的伦理问题。”梅宏表示，在AI快速发展的热潮中，需要对可能的风险进行研判，并提醒科技工作者时刻牢记科技向善、以人为本。

随着数字技术的发展，科技伦理治理也应当跟上。梅宏强调，目前应当完善科技伦理治理体制机制，明确数字技术领域的治理重点和安全护栏，对数字技术进行分级分类治理，同时建立健全AIGC（生成式人工智能）的主动披露标注制度，并开展相关技术研究。

那么，数字技术领域是否应该存在研究禁区？梅宏认为，在基础研究阶段一般不做限制，涉及对人的认知能力调控、违背人的自由意志的研究应列为禁区。在技术和产品应用阶段，要根据具体场景及影响确定。

“例如大规模远程监控、自主决策的社会评价体系、操纵个人意识和行为并造成个体或他人身体或心理伤害的技术，以及以超越或替代人的认知能力为目标的AI技术研发，应当受到限制。”梅宏说。

同时他表示，数字技术领域的研究禁区应该建立动态调整机制，适时调整伦理审查的复核清单。

此外，他呼吁建立AI生成内容的披露标准机制，建立涵盖大模型开发者、创作者、使用者的标注责任机制；鼓励多条技术路线推进大模型生成内容标注技术研发；完善大模型内容审核制度规范和检测技术开发；推进相关标准和规范制定并将之国际化。