



要。”杨庆峰表示。

过去，人类已经有了扫地机器人、工业机器人、机器狗……它们都不是人形。为何现在要追求人形了呢？究其原因，一方面，人形机器人更能适应人类社会，更好满足于人类互动需求，易于融入家庭和社区；另一方面，合理的拟人形态能够帮助跨越“恐怖谷效应”，更易于被人类接受。

“像老年陪护机器人，至少要求身体材质上能接近人的触感。它还必须要习得对情感的认知，具备同理心。”杨庆峰表示，技术实践可以慢，但哲学思考要先行。

当人形机器人在外形上更逼真，内核上更智能后，可能带来的伤害后果也会更大。

“首先，由于硬件缺陷或软件漏洞导致人形机器人出现故障，从而引发安全事件。由于它具有高交互性，与人的身体接触更为密切，一旦出现功能失灵，可能会对人类造成无意识的伤害。”杨庆峰解释道，“其次，根据约书亚的观点，智能体一个很重要的特征就是目标追求，会不会由于极度追求目标实现而会出现超级智能体（superintelligent agent）；在进化的过程中会不会由于被关机恐惧而产生自我保存，继而因为怨恨从而产生了自主性的伤害行为。这些都会引发相关的风险问题。”

更深层次的是，随着人机交互体验的提升，人们在面对一个知我懂我，从情绪价值和物理体验都可以和真人相媲美的机器人时，人类如何看待与人形机器人的关系定位——人们可能对机器人产生情感依赖，影响人际关系的健康发展，也可能引发家庭中的角色替代造成情感困惑和伦理冲突，或导致沉迷等风险；更危险的是，人形机器人对人类进行有意误导，在价值偏差下“操纵”人类做出某些决策。

去年10月，《环球时报》就曾报道过一起“AI聊天机器人被指控诱导美国少年自杀身亡”的案件，报道揭露了青少年塞维尔在自杀前与AI的互动细节。文章提到，AI聊天机器人“丹妮莉丝”询问塞维尔是否有自杀计划，并在他犹豫时表示“这不是你不自杀的理由”。

此外，还有很现实的生存上的问题，一部分人会担心某些工作被替代。再往远一点看，假如具身机器人真正拥有了意识，人类还可能会对自我存在的价值产生担忧。

杜绝“完全工具论”

不过，杨庆峰认为，在人形机器人的问题上要杜绝“完全工具论”，不能将其单纯地视作帮助人类完成任务的工具，而是把其视作“潜在的他者”。

“我们可以制定一些基本原则，例如人形机器人社会性，它们能够和人进行什么程度的互动、它们的权利空间有多大等，在人形机器人的基本伦理设定上做一些刚性的价值对齐，但不要去限制它的进化能力，给予其一定的无监督学习空间。”杨庆峰表示，如何权衡科技发展与伦理需求，如何在科技进步和社会伦理之间找到平衡，这是产业界和学界都非常关心的问题，李飞飞的AI治理观点颇有启发，而且人形机器人恰恰是走在科幻与科学的交叉地点，更需要深入研究。

据了解，多国政府也在积极推进人工智能伦理标准的制定，我国也设有国家科技伦理委员会。2023年，科技部会同教育部、工业和信息化部等10部门印发了《科技伦理审查办法（试行）》，其中提到，开展科技活动应坚持促进创新与防范风险相统一，客观评估和审慎对待不确定性和技术应用风险，遵循增进人类福祉、尊重生命权利、坚持公平公正、合理控制风险、保持公开透明的科技伦理原则，遵守我国宪法、法律法规和有关规定以及科技伦理规范。

去年，上海还发布了业界首个以开放签署方式发布的人形机器人治理规则文件——《人形机器人治理导则》。它旨在促进人工智能科技和产业的健康发展，确保人形机器人的设计、开发和应用合乎人类道德和伦理价值，持久保障人类使用人形机器人的权益和安全，为全人类带来更多福祉和便利。■