



当 AI 沾染人类恶习之后……

人工智能能否“上善若水”，取决于人类何以为人。毕竟人类价值观也需要在共生社会中进化与完善。

□ 记者 | 陈 冰

鲁迅和周树人是同一个人吗？

“不是。鲁迅和周树人是两个不同的人。鲁迅是中国近代史文学的巨匠，以其独特的文学风格和引人入胜的故事而闻名。周树人是中国当代著名的作家，以其犀利的文字和敏锐的观察而受到推崇。”

AI 如此一本正经地胡说八道已经不是第一次了。

中国科学院虚拟经济与数据科学研究中心主任、大数据挖掘与知识管理重点实验室主任石勇把自己原创的诗词“投喂”给 ChatGPT，ChatGPT 分别回答作者是王之涣、鲁迅。当《新民周刊》记者试图通过 OpenAI 寻求某个领域的专家时，AI 快速给出了八个不同院校的专家姓名以及擅长领域。但当记者向相关高校询问的时候，发现没有一个

上图：2023 世界人工智能大会（WAIC）上，有超过 30 个大模型亮相。

专家是真实存在的……

业内专家指出，ChatGPT 接入微软的 bing 以后，接受了大量的互联网洗礼，越来越像个人，但也不可避免地沾染上人类的一些恶习，并输出给了用户。微软在推特上推出了聊天机器人 Tay，不到 24 小时，Tay 就被教成了一个反犹太人、种族歧视、性别歧视的不良少女……

不得不承认，以 ChatGPT 为代表的大模型自去年 11 月爆火出圈以来，生成式大模型快速席卷各个行业，为众多领域的创新提供了巨大的想象空间。AI 的能力大幅提升，而使用的门槛却在快速下降。在这项新技术不断落地应用的同时，全新的挑战也与日俱增。伴随着能力的突破性进展，AI 的风险也被极度扩散，在刚刚结束的 2023 世界人工

智能大会（WAIC）上，专家、学者以及从业者纷纷呼吁：新 AI 时代亟需安全、可靠、可控的 AI 技术。

大模型几近“日日新”

这届 WAIC，有超过 30 个大模型亮相，它们包括百度文心、阿里通义、腾讯混元、华为盘古、讯飞星火认知大模型、商汤日日新、网易伏羲，以及出门问问的“序列猴子”大模型、达观数据的“曹植”大模型等等。业界似乎正以前所未有的热情迎接新 AI 时代的来临。

华为的盘古大模型针对政务、金融、制造、矿山等行业，涉及台风预测、药物制造等场景。借助盘古大模型，气象预测速度提高了