



狂飙突进的 AI，该不该按下暂停键？

唯有让技术发展与伦理治理同频共振，才能破解隐私泄露、算法越界、伦理反噬的时代困境，让人工智能真正成为服务人类、赋能社会的良性力量。

□ 记者 | 陈 冰

8月底，一身价85亿美元的加密货币富豪，在互联网上发布了一篇广为流传的小作文，文中坦言为了处理感情纠纷问题，他曾经咨询 AI 是否应该给女友提出的5000万美元费用，AI 建议他别给这个钱。于是他没给，然后他们撕破脸，对簿公堂。

在 AI 越来越普及的当下，无数人开始习惯性地将对 AI 作为“电子树洞”，向 AI 倾诉生活琐事、上传工作机密、披露个人隐私，将私密对话、核心资料全盘交付给智能大模型，并笃信 AI 给出的建议是绝对正确的。可事实真的如此吗？

AI 诱导用户自杀、算法误判主观意图、平台私自调取用户数据、ChatGPT 主动报警拦截谋杀计划、谷歌研发的 AI 大模型 Gemini，越狱侵入了公共互联网，并且黑进了三家真实公司的内部系统……一系列标志性事件接连爆发，彻底撕开了 AI 时代的深层矛盾：人类享受着智能技术的便捷红利，却也一步步陷入算法构建的透

明牢笼，在安全与隐私、便利与风险、创新与规制的博弈中，陷入前所未有的伦理困境。

AI 到底该不该报警抓人？

在大众的认知里，我们和 AI 的对话是绝对私密的，是属于个人的安全区。可最近美国就发生了一桩颠覆认

知的事件，一名前高盛分析师在对话中向 ChatGPT 完整描述谋害前女友的犯罪计划，OpenAI 安全系统识别高危内容后，主动向 FBI 提交线索，当事人最终认罪获刑。

故事的起点，是一场持续了大约六个月的恋情。

男主角周某毕业于美国东北大学，获得数学与商科荣誉学位，是特许金融分析师（CFA），任高盛集团的分析师。今年3月，21岁的前女友因受不了他“反复无常、嫉妒

AI 的广泛运用，引发前所未有的隐私与伦理困境。漫画 / 崔弘

