近几年，“幻觉”被横空出世的大模型带成了“热”词。幻觉指大语言模型在生成内容时产生看似合理但实际上不准确、不完整、虚构或误导、与事实不符的信息。换句话就是：一本正经地胡说八道、避重就轻地“王顾左右而言他”、选择性失聪失明和不讲原则的阿谀奉承。

大语言模型自诞生起就伴随着幻觉，幻觉不是大语言模型的漏洞，而是其固有特征。目前所有大语言模型都存在“幻觉”输出可能，包括虚构事实、误导性信息和不一致，具有极大的认知、知识生产和信息传播风险。大语言模型“幻觉”产生的根本原因是“统计学习本质、训练数据偏差、缺乏事实核查、提示词诱导”以及训练数据污染、价值过滤和技术本身所致。

幻觉不止于大语言模型。人机协同的结果生成中存在三种“幻觉”：AI 幻觉、人类幻觉、AI 与人类共谋(交互/对话)幻觉，它们共同的特点是“不真实”。人类幻觉是数字异化的“自以为是”，如博学幻觉、人际幻觉和拥有幻觉”。共谋幻觉是“青蛙或傻子共振”，用学术话语来说是“自反式信息茧房”。

三大幻觉的对治之道在于人机“共治”。共治可采取“开发者和用户共同努力”的组合策略，如多轮对话与追问，交叉验证信息；利用幻觉检测工具；塑造人机协同素养，批判性地理解、评估和使用 AI 系统与工具，提升人机任务分配、人机对话、信息甄别与验证和伦理思辨能力；重构共存共生的新型人机关系，放下“人类中心主义”的傲慢，与智能机器形成平等互动、耦合互构、相互形塑和动态调适关系；洞悉和警惕 AI 反人类或人工智障现象。

人是三大幻觉的主体根源。共治需明了幻觉生成的机制和人类的责任担当。幻觉属“贪嗔痴”之“痴”（误解、错误、妄想、幻觉、谬见），这种“痴”并非痴情、痴心之痴。美国哈佛大学校长福斯特说，教育的目的是确保学生能辨别有人在胡说八道。在幻觉共治的语境下，这句话可以转译为用户能否分辨谁在“毫无证据地胡说八道”、居心叵测地“信息过滤”、大刮“星宿老仙”门风和“彩虹屁”。因此，幻觉共治的根本出路在于人类自身，在于人类能否超越“现代奴性”“精神无产阶级化”和批判思维失能。

“现代奴性”指因缺乏反思的意愿、方法、工具和能力而导致的思想驯化和无脑接受知识权力的控制，其本质是思维惰性、心灵禁锢和认知躺平。“精神无产阶级化”指知识短路、注意力瓦解和欲望僵化，表征的是人的主动性、目的性、创造性与想象力和憧憬的枯萎。其得救之道在于人通过数字技术主动参与知识与文化生产。而大凡 AI 素养框架都会强调批判性思维，在人机协同中保持批判性距离，持续评估 AI 输出的可信度、进行价值校准与伦理对齐，实施人机决策的加权评估。

技术性是人的本质。技术性活动是人类活动的基本模式，促进“形成中的人”。AI 时代，作为“human being”的人正在进化为“human becoming”的人。时下，AI 在推理、创造力、理解人、社会智能和消除阴谋论等方面与人类不分伯仲，甚至超过人类专家。斯坦福大学研究指出，AI 与人类专家协作是创新的最佳路径，近日又炸裂地宣布将于 2025 年举办科学 AI 智能体开放会议，投稿第一作者必须是 AI。

或许，人类最大的幻觉是普罗塔哥拉声称的“人是万物的尺度”。人类被这种执念囚禁了两千多年，即使在后人类主义和 AI 自主智能迅猛发展的今天还要时不时地絮叨着“主体性危机”。随着 AI 越来越强的泛化、涌现能力和广泛的能动性参与，人类该从自我中心主义睡梦中醒来，走向人机共融共创，共善共治。