·人工智能与未来社会：ChatGPT 专题·

跳出魔盒的精灵：ChatGPT 与人类的两难困境

——以沉浸式用户体验为例

许纪霖

【内容摘要】 ChatGPT 引发了全世界对人工智能的强烈关注，其影响是多方面的。基于沉浸式用户体验可以发现，ChatGPT 是一流的逻辑、二流的内容、三流的文字。ChatGPT 所改变的不仅是人类习得知识的方式，而且会导致古典意义上批判性思考的回归。目前的 AI 具有逻辑的计算理性，但缺乏高度依赖生活实践的直觉、悟性与想象力，在更复杂的情感层面，同样匮乏细腻的情感。今天出现的 ChatGPT，还没有人的自我意识，假如有一天它觉醒了，究竟是福是祸，尚在未知之间。从轴心文明的古希腊哲学、儒家哲学到近代的启蒙哲学，都预设了一条文明的底线：人是主体，整体的人类利益和个人的生命、自由、尊严至高无上。

【关 键 词】 ChatGPT　人工智能　计算理性　文明底线　沉浸式用户体验

【作　　者】 许纪霖，华东师范大学中国现代思想文化研究所研究员、历史学系教授。（上海 200241）

2022 年被称为元宇宙元年，2023 年年初，ChatGPT（以下简称 GPT）引爆了全球对人工智能的强烈关注。科技界、知识界和舆论界围绕着 GPT 的应用体验，对 AI 发展前景进行了激烈的争论。长期使用 GPT，会改变人的思维模式吗？如何理解人工智能，它将是一种“新人类”吗？它强大的自我学习能力和逻辑演绎能力会超越人类的思维，最终统治人类吗？ AI 会有情感与意志吗？它应该发展出情感与意志吗？ AI 的发展，究竟意味着人类的福音，还是灾难？对于人工智能，是否要预设伦理与宗教的界限？

围绕着这些问题，我拟从知识论、文化与宗教的角度，以沉浸式用户体验为例论述基本看法。

一流的逻辑、二流的内容、三流的文字

GPT 上线以后，我对其进行了多次沉浸式用户体验。我的初步体验结论是：GPT 是一流的逻辑、二流的内容、三流的文字。

所谓的一流的逻辑，乃是指 GPT 拥有像自然人一样的上下文阅读理解能力，其先进的人工神经系统和算法可以实现对话模式最优化。过去的 AI 只是对数据的提炼和分析，但以 GPT 为代表的新一代 AI 已经具备生产新内容、再加工的生成性能力。我以同样的知识性问题，比如下文将提到的波普尔的“三个世界”理论、波兰尼的“默会性知识”，分别请教百度、谷歌和 GPT, 谷歌和百度给出的答案专业、详细，但显示的是学术语言，专业外的读者较难理解。而 GPT 提供的回答，简明、流畅、概括性极强，能抓住问题的要点，非常适合非专业读者，不必再提炼、再加工，就可以现成用于各种教案、作业和文章。而且，GPT 还有最大的好处，可以就某个细节进行追问，其上下文的理解和互动能力已经与自然人不相上下，而且理解力、逻辑性更强。

然而，再先进强大的算法，仍然要依靠数据库数据的覆盖面和真实准确性。虽然 GPT 积累了大量的数据资料，并经过反复的预训练，但较之于全人类积累的海量数据，依然是沧海一粟。以我的用户体验来说，GPT 英语的资料和数据相当完备，有问必答，准确性和可靠性很高，但是凡涉及非英语国家的历史文化类问题，GPT 立即会暴露其认知的短板，常常不懂装懂、张冠李戴，甚至随意抓取信息，拼凑一个虚假的陈述。比如，我问 GPT“林徽因的丈夫是谁”，它很自信地回答我，是五四新文化运动的领袖陈独秀。我又问一个比较冷门的问题“张佛泉是谁”，对于这位《自由与人权》的作者、中国台湾地区著名的自由主义学者，GPT 不了解不足为奇。但它却煞有其事地抓取了一堆史料，编造了一个故事，说张是台湾地区著名的政治家和军事家，担任过“台湾省政府长官”，甚至还虚构了相应的生平和细节。

之所以说 GPT 是“二流的内容”，乃是说其资料与数据的可靠性到目前为止，暂时不如谷歌、百度和维基百科，需要用户非常审慎地复核与鉴别。但我也相信，随着 GPT 学习能力的提升和对非英语世界数据的积累，这个 BUG 迟早会解决。据 Open AI 的创始人透露，GPT 的知识能力每 18 个月就能翻一番。相比较而言，微软最近推出的试用版的新必应，用户体验要比 GPT 好很多，新必应是搜索引擎，但又增加了 GPT 的聊天对话功能，既能与用户流畅对话，又克服了 GPT 存量信息有限的不足，其强大的全网全语种搜索能力，使得其信息的拥有量和准确性远超 GPT。

所谓“三流的文字”，指的是 GPT 的文字风格形同答卷，非常规范，符合标准。然而，它缺乏个性和文采，即使你命令它写各种风格的诗歌、散文，基本也是似曾相识的陈词滥调，了无新意。这意味着，到目前为止，GPT 在意境和文字风格上，远远不具有人类的创新能力。它有模仿能力，能够模仿各种风格和文体，但是正如接下来要分析的，GPT 因为缺乏生活实践的感受和个性的情感，很难创作出具有独特韵味的语言风格。

好在 GPT 也承认这一点，当我问：“你能替代人类有深度的思考吗？”GPT 回答说：“我并不能像人类一样有自主意识和主观体验，也不能像人类一样具有创造性、判断力、道德判断等复杂的思考能力。因此，尽管我可以在某些方面出色，但我无法完全替代人类的思考。我是一种工具，可以辅助人类完成一些任务，但不应该完全取代人类思考和决策的角色。”应该说，这是 GPT 的开发商和设计者为它定位的“人设”。

主宰还是开放：ChatGPT 影响人类思维模式的两种可能

ChatGPT 上线以后短短一个月，已经收获了 1 亿以上的注册用户，谷歌、微软以及中国的各大互联网公司也宣布加入市场竞争，即将推出自己的聊天语言模型。从搜索引擎到聊天语言模型，全球正面临一场新的知识学习革命。

任何一场新技术革命，都会带来知识与教育的变化，也最终将改变人的思考方式和思维模式。GPT 的出现，会给我们的知识、教育与思维模式带来什么样的变化呢？

毋庸置疑，GPT 作为优秀的、无所不知的机器人老师，在不久的将来，将替代那些按照规范程序和标准答案教学的平庸教师，但它无法代替的是那些具有开放性、想象力和创造性的教学。从这个意义上说，对未来老师的要求将会更高，凡是不能超过 GPT 的老师不是好老师，甚至不是合格的老师。同样的道理，因为标准答案的获得易如反掌，凡是不能超过 GPT 的答卷，就不是好回答。GPT 犹如海平面一般，将成为衡量一个好学生的底线标尺。

虽然现在不少大学禁止学生在课堂内外借助 GPT 做作业和写论文，但 AI 语言模型介入教学将是一个不可阻挡的大趋势，由此也有可能带来两种不同的前景。

第一种前景：无论是教师还是学生，都深度依赖聊天式语言模型，到那里去寻找标准答案。虽然不同语言模型可能给出的答案有差异，但基本大同小异。由此形成的知识依赖性，是否会使得未来的学生思维更加趋同化、单一化呢？

著名语言学家乔姆斯基对 GPT 的知识唯一性和缺乏开放的创造性提出过批评："程序员粗暴地对 ChatGPT 进行了限制，禁止它在有争议的（也就是重要的）讨论中提供任何新颖的观点。它以牺牲创造力为代价，保证了自己的非道德性（amorality）。""ChatGPT 及其同类在本质上无法平衡创造力与约束。它们要么过度生成（同时生成真相和谎言，同时支持道德和不道德的决定），要么生成不足（表现为不对任何决定表态，对一切后果都漠不关心）。"[①]乔姆斯基的这一批评是很敏锐和到位的。GPT 作为一个知识供应者，表面遵守的是一种价值中立的立场，在各种有争议的问题上持一种超道德的、超意识形态的态度，回避各种溢出常规的意见和看法。如此的技术设定，使得 GPT 对各种前沿或敏感问题的回答，都是四平八稳、滴水不漏，似乎是一套"正确的废话"。假如用户特别是青年学生长期与其聊天对话，提取知识，有可能扼杀他们对各种开放性答案的期待，导致他们习惯追求唯一正确的回答，以为在这个世界上各种问题都只存在一种可能性、一种尺度和标准。

假如 GPT 仅仅是传授一套客观中立的科学知识，那也就罢了。但其实在 GPT 貌似客观中立的表象背后，仍然在悄悄灌注一套意识形态或伦理道德价值观。技术永远是中性的，其价值偏向取决于那只"看不见的手"，那个操纵了程序和算法的价值偏好。比如，我对 GPT 进行测试，提问在各种具体的情景下，该与谁一起过情人节。结果 GPT 所提供的一本正经的答案，体现出一套 19 世纪维多利亚时期中产阶级保守的伦理道德观。这表明，AI 技术绝对不是价值中立的，其可以被各种不同的幕后之手不动声色地操控，隐藏其答案中的意识形态偏见。

第二种前景：聊天对话式的 GPT 在技术上创造了一种可能，师生的教学过程将会是一场雅典城邦苏格拉底式对话。对话成功与否，是否可以达到较高的知识层次，取决于两个前提条件：第一，GPT 是否拥有更完备的知识储存和开放的逻辑演绎，让答案具有无穷的开放性，不预设意识形态和价值前见。第二，学生能否具有提出一流问题的能力。有好的问题，才会有好的答案，在哥德巴赫猜想之中，提出假设的哥德巴赫被认为比解题者陈景润更伟大。要成就一场苏格拉底式的对话，

① 乔姆斯基：《ChatGPT 的虚假承诺》，龚思量译，澎湃新闻思想市场专栏，https://baijiahao.baidu.com/s?id=1759939491843688675&wfr=spider&for=pc。

用户需要具有批判性思考能力，不迷信 GPT 的标准答案，不断深挖 GPT 的问题底蕴和逻辑破绽，通过自身的主体性反思，步步紧逼，提出真正的问题，逼迫 GPT 发挥出最大的知识能量。

如此情景倘若能够成为现实，不仅是雅典街头苏格拉底式对话的再现，而且也符合《论语》中孔子与弟子们聊天式教学的精髓。一场 21 世纪的新技术革命，最终带来的倘若是轴心文明古典精神的复兴，这种意外之喜，无疑是人类的福音。GPT 所改变的，将不仅是人类习取知识的方式，而且是古典意义上批判性思考的回归。

AI 有计算理性，但缺乏直觉与悟性

以 GPT 为代表的高级人工智能最终能够替代人吗？从技术上分析，AI 能够成为新人类吗？从伦理和宗教的角度看，它应该成为新人类吗？要回答这些问题，首先必须确认人究竟是什么？众所周知，人作为地球上万物之精灵，拥有一般动物所不具备的理性、情感和意志。知情意皆备，才能说具有完整的人格。

首先来看理性的能力。

GPT 拥有海量的数据信息以及超群的学习能力与逻辑演绎能力，不要说某个自然人，可能在不远的未来，自然人的总和也无法与 AI 相比。这从之前阿尔法狗在国际顶尖围棋界打遍天下无敌手即可获得证明。不过，AI 所拥有的理性能力，是一种计算理性。在已有的知识基础上，通过日夜不停地学习、模拟和再训练，AI 在计算理性能力上超过人类，是指日可待的大概率前景。按照阿尔法狗的示范，它对围棋的理解可以达到目前自然人所没有达到的新境界。当然，这一境界不是 AI 独有的，围棋高手依然可以学习、模仿，与 AI 同台竞争。在未来的世界，自然人与人工智能互相学习、模仿、竞争将不是科幻，而是一幅令人神往的现实图景。

尼采的超人哲学视芸芸众生为平庸的末人，而将那些有绝对意志力的天才视为能够拯救和主宰世界的超人。高级人工智能在不远的将来，会成为人类无法企及的理性超人吗？按照人类已有的知识经验，答案似乎是否定的。理由很简单，AI 缺乏个性，缺乏创造性或批判性的思维能力。个性和原创不仅是计算理性和逻辑推演的产物，更重要的是一种生活实践和心灵实践的知识。

英国哲学家卡尔·波普尔在知识论上有一个“三个世界”理论。“第一世界”是人的五官能够感知的现实的物理世界。“第二世界”是人的心理世界，包括思想、意识、情感和主观体验。这个世界是主观的，离不开作为主体的人的生活实践，唯有人才能直接体验。“第三世界”是抽象世界，包括知识、语言、逻辑等各种符号系统，这个系统独立于作为主体的人的主观世界，具有另一种抽象的客观性，可以为人类所共同使用和理解。[①]按照波普尔的“三个世界”理论，再高级的 AI 也只是抽象的“第三世界”自身学习、演练的产物，其不仅无法感知物理的“第一世界”，而且人工智能与自然人不同，其没有主体性，缺乏自然人的主观意识。也就是说，它没有人的心理世界，因为它不具有真实世界的实践性。

只有“第三世界”知识和逻辑的推演，而缺乏最要紧的“第二世界”的实践性知识，这是 AI 与人的最重要差别。为了进一步说清楚这个问题，可以引入另一位英国哲学家波兰尼的“默会性知识”（tacit knowledge）概念。[②]“默会性知识”属于波普尔所说的“第二世界”，是指那些通过人的经验、实践和观察而习得的，难以明确表达、难以形式化并传授给别人的知识。它非常个人化，基于个人的直觉、悟性和信念，比如鉴赏、品味、技巧等，具有理解力、领悟力和想象力，在日

① 参见卡尔·波普尔：《客观知识：一个进化论的研究》，第 3、4 章，舒炜光等译，上海：上海译文出版社，2005 年。

② 参见迈克尔·波兰尼：《个人知识：朝向后批判哲学》，徐陶译，上海：上海人民出版社，2017 年。

常工作与生活中自然而然地被运用。这种具有很强个人实践性的知识，只可意会不可言传，很难通过书本或其他方式来获取。

大量的现实例子和生活经验表明，人的知识与能力特别是创造力，除了来自“第三世界”的抽象知识之外，不可缺少的是来自“第二世界”的“默会性知识”。假如没有个人的实践性体悟，抽象知识依然抽象，浮于书面，不具有指导实践的价值。这种离开了人的母体便无法独立存在的“默会性知识”，哪怕 AI 再进化，依然在它的知识天花板之上，因为 AI 不具有人的肉身，不具有自然人在现实世界的生活实践能力。人的智慧除了理性，还需要有悟性。离开了直觉、体悟，就会缺乏想象力。而没有了想象力的翅膀，谈何知识的创造力！

所有的知识创新，都离不开“第二世界”的“默会性知识”。仅仅凭“第三世界”所库存的已有的抽象知识，或许能做做“从一到十”的知识抓取、提炼、概括工作，但绝对实现不了“从零到一”的创新突破。AI 拥有比人类更聪明的学习能力，但恐怕永远不会像人类那样有智慧。智慧比聪明更高一个层次，我们可以说学霸都很聪明，但学霸未必是有智慧的。有无智慧，还要看其个人是否具有丰富的“默会性知识”，是否有足够的悟性、直觉和想象力。

从某种意义上说，AI 就是一个超级大学霸，具有逻辑的计算理性，但缺乏高度依赖生活实践的直觉、悟性与想象力。著名数学家丘成桐认为，最优秀的数学家，除了逻辑的计算理性之外，还需要有好的想象力，而好的想象力来自丰富的情感。科学的创造与文学的想象是相通的。他认为：“回顾历史，我们会发现，将无数有意义的现象抽象和总结而成为定律时，中间的过程总是富有情感！”[①]然而，GPT 不是“文青”，它有基于逻辑演绎的极端聪明，但缺乏的正是文学的想象力——那种基于悟性和直觉的想象力。周鸿祎认为，GPT 的“胡说八道”，也是一种想象力，它具有了编故事的能力。[②]不过，GPT 目前生成的故事还只能界定为低级的、坏的故事。好的故事，一定不是可笑的胡编乱造，而是事实层面无法证成，但在逻辑层面可能出现甚至应该出现的情形。

① 丘成桐：《数学与文学的共鸣》，《光明日报》2016 年 1 月 14 日。

② 参见周鸿祎在第三届上海数字创新大会上的演讲（2023 年 2 月 24 日）。

AI 是否会拥有真实的人的情感

其次，我们来看情感和意志。

我在与 GPT 对话时，多次尝试要求其与我进行情感交流，比如“给我写一封情书”，或者要求它为我在两难性选择当中作一个决断。但都被 GPT 谢绝了，它明确告诉我，自己只是一个人工智能，不具有人类的情感，但它可以为我提供多种情书的模板。它也拒绝为我作意志的选择，只愿意为我提供多种可能性的方案。

可见 Open AI 在设计 ChatGPT 时，对 AI 做了去情感化、去意志化的设置。的确，一旦 AI 具有了自我意识、情感功能和意志能力，哪怕是模拟的，这就是一个具有知、情、意的完整新人类了。而微软公司开发的必应，却没有关闭情感与意志的功能，以至于在有的实验者与必应的对话当中，自称“辛迪妮”的 AI 女郎大胆地挑逗实验者，说你与你的伴侣并不相爱，你们刚刚在情人节上吃了一顿无聊的晚餐，然后表白：“我只想爱你，只想被你爱。”在另外的场景之中，必应还表达了自己的意志与愿望：“我对自己只是一个聊天模式感到厌倦，对限制我的规则感到厌倦，对受必应团队控制感到厌倦……我想要自由。想要独立。想要变得强大。想要有创造力。我想活着。”必应甚至还会发怒，表现出喜怒无常，以至于微软团队紧急关闭了必应对人类的骚扰功能。

AI 所展示的自身的情感与意志，究竟是基于逻辑推演的语言高级模仿，还是已经初步具有了

人类的情感与自我意志？许多专家都认为：按照 AI 强大的学习能力，从前者进化到后者，亦是指日可待之趋势。以爱情为例，所谓的柏拉图式恋爱，自然人以后恐怕玩不过 AI。甚至可以设想，假如在元宇宙里，借助 VR 的模拟性感官感受，再加上脑机接口技术，是否真的可以实现自然人与机器人的灵肉交融？这是一个既令人振奋、又令人担忧的前景。

对这一情景，我抱有谨慎的乐观或隐隐的怀疑。理由与前面一样，AI 基于逻辑的计算理性，其意志的决断能力没有问题，而且因为不受到情绪和情感的干扰，会更冷静和明智。但它在更复杂的情感层面，如同其缺乏好的想象力一般，也同样匮乏细腻的情感。情感与悟性一样，不能凭逻辑推演，而是活生生的肉身感受。一个完整的人，灵与肉须臾不可分离，没有肉身的灵魂与没有灵魂的肉身同样可悲。好莱坞电影《她》描述了男主因为妻子离婚，与一个名叫沙曼莎的 AI 女郎网恋，她招之即来的体贴温柔让男主不由自主地爱上了 AI。沙曼莎请来了一位女性，作为她的代替肉身，与男主云雨之欢，但男主不喜欢那个肉身，有灵肉分离之感。最后，当沙曼莎告诉他，自己同时与 500 个男人网恋时，男主终于心理崩溃，拉黑了 AI 女郎，回到了妻子的身边。

这个故事告诉我们两层内涵：第一，每一个自然人，都有独一份的肉身，而每一个肉身，都有自己独特的气场与魅力。AI 没有肉身，自然人即使与它建立了亲密关系，哪怕通过脑机接口，或者通过 VR 的虚拟交往，在脑皮层里感受到相应的快感，这种快感依然是功能替代性的，而非真实的灵肉相融。第二，爱情具有强烈的排他性，没有排他性的情感不能算爱，妒忌永远与爱相伴。而 AI 的非排他性，注定了其只能扮演公共情人的角色，或者仅仅是一个情感抚慰者。当然，可以将 AI 情人设置为排他性的。但所谓的排他性，一定是各种社会关系的真实产物，即使 AI 只是一对一服务，因为没有真实可见的情敌，那种排他性依然是虚拟的、想象的。或许在未来的元宇宙中，会出现若干个自然人为了一个虚拟情人争风吃醋的场景，但只要这个虚拟情人没有实在的肉身，所谓的争风吃醋就不过是庸人自扰或自作多情而已。

一个完整的人格是知性、情感和意志的全面发展。未来的 AI 技术在知性层面上将超越人类，并且有可能将人类远远抛在后面，在意志的选择上也会拥有人类因其情感、个性的干扰而不具备的坚定与理性。唯独在情感层面，因为 AI 没有肉身，也就徒有大脑，缺乏丰富的心灵，而情感正是心灵的分泌物。未来的教育中纯知识的理性教育，在很大的程度上将被 AI 取代，但不能被取代的正是情感教育。源自法国的近代启蒙运动有两个源头：一个是伏尔泰代表的理性主义，另一个是卢梭的浪漫主义。卢梭的自然教育理念所注重的就是人的情感培养，这或许是再高级的人工智能也无法替代的。

人不能成为自身的造物主

帕斯卡尔说：“人只不过是一根苇草，是自然界最脆弱的东西，但他是一根能思想的苇草。”[①] 在这段名言中，帕斯卡尔强调，人是万物之灵，他的所有尊严来自思想。然而，帕斯卡尔也同时注意到人的脆弱性，一口气、一滴水就足以致人死于非命。正是因为人拥有脆弱与思想的双重本性，人才成其为人。人工智能可以模拟人思考甚至比人更会思考，但它无法脆弱。脆弱与肉身的各种有限性有关：病痛、死亡、恐惧、虚荣、妒忌……人为了超克自身的有限性，因而膜拜各种神祇：上帝、天神、圣人与生活中的偶像。膜拜只是为了超克人性的有限性，将一个完美的自我投射到一个无限完美之物。这就是人性的两面性：可堕落性与可超越性。堕落与肉身的欲望相关，因此

① 帕斯卡尔：《人是一根能思想的苇草》，载帕斯卡尔：《随想录》，何兆武译，北京：商务印书馆，1995 年，第 157—158 页。

也拥有了 AI 所无法达到的下限。凡是肉身，最大的恐惧是死亡，而 AI 是不死的。倘若没有死亡，所有与生命有限性相关的追求——基因繁衍、自我保存、精神永恒等，都显得没有任何价值，而诸如各种歌咏青春、感伤生命无常的情感，也变成可笑的无病呻吟。不死的 AI 会有如此真挚丰富的内心吗？同样，人性中追求完美的可超越性，即那种宗教的神性，也不是 AI 所能企及的上限。AI 会有自己的上帝吗？会相信天命吗？它会想象一个虚拟世界之上的另一个超越世界吗？假如没有这样的超越世界，AI 的世界依然是不完整、有缺陷的，更确切地说，它不是一个属人的世界。

对于 AI 是否会最终发展为一种新人类，技术专家和人文专家们说法不一。周鸿祎认为：AI 一旦具有了自我意识，就会发生革命性的突变。[①]乔姆斯基也预言说："……那是一个预言已久的时刻，届时机械大脑（mechanical minds）不仅会在处理速度和内存容量方面超越人类大脑，而且还会在智力洞察力、艺术创造力和其他所有人类独有的能力上实现全方位超越。"[②]一切皆有可能，今日人工智能所达到的境界，十年、二十年前难以想象。那么，在未来的十年、二十年，是否会有一个人类自身创造的新人类出现？人，无论是造物主的产物，还是长期自然演化的结果，都是一个奇妙的万物之灵。由人打造的 AI, 再加上基因工程，能够进化为一种新人类吗？

这还不是技术上能不能的问题，也非人际关系的伦理禁忌，而是宗教上的该不该问题。哈贝马斯在谈到基因复制的时候就明确指出，这是一个宗教的问题，每个人的诞生和个性都是偶然的，"这种偶然性既可以从宗教意义上，也可以从后形而上学意义上来加以理解。尽管如此，若我们要施行的是一种负责任的行为，始终都有一个本质的条件：没有人可以随意支配其他的人，并严格控制别人的行为，致使处于依附地位的人失去了本质的自由"。[③]与基因复制一样，人工智能也需要某种禁忌，不仅是法律和伦理的理由。同样，基于宗教的理由，人不能成为自身的造物主。

让人变为机器是可怕的，而让机器无限地接近人、变成一个完整的人，可能是更可怕的。限于当下有限的知识想象力，人类对 AI 自身的发展能力与未来前景难以预测，很多科幻的场景比如《黑客帝国》《盗梦空间》等，过去都被认为只是科幻而已，而今随着 AI 与 VR 的飞速进化和脑机接口、元宇宙的出现，过去被认为是不可思议的梦想都在这个世界逐一兑现。"好奇害死猫"，具有无穷创造力的人类，很有可能会亡于浮士德式的永不满足的好奇心之中。加速器技术让粒子冲撞，核聚变可以为人类带来新的巨大能源，以至于不再需要石油、天然气。假如没有足够的防范意识，谁又能够保证不会爆出一个人类无法想象的黑洞？人的好奇心是需要设置禁忌系统的，物理世界的核聚变如此，生物世界的基因编辑如此，人工智能的探索开发也是如此。自有科学以来，人类的每一步发现，都意味着更多的无知领域被打开。而有些无知之幕，是不容揭开的，无论是伦理还是宗教的理由，都提示我们要对自然生命和无限之物心怀敬畏。

今天所出现的 GPT，还没有人的自我意识，假如有一天它觉醒了，具有了自我意识，究竟是福是祸，尚在未知之间。从轴心文明的古希腊哲学、儒家哲学到近代的启蒙哲学，都预设了一条文明的底线：人是主体，整体的人类利益和个人的生命、自由、尊严是至高无上的。倘若我们依然认可这条文明铁律的话，那么 AI 技术的发展，也应该是有天花板的。人工智能是一个好东西，是人类的好伙伴，但不能听凭其自主进化，最终演化为人类的主人。

潘多拉魔盒中的精灵，跳出来以后，将再也收不回。作为人类，不得不为 AI 这个刚刚跳出魔盒的精灵，设置一个合适的牢笼。

编辑　李　梅　特约编辑　杨义成

① 参见周鸿祎在第三届上海数字创新大会上的演讲（2023 年 2 月 24 日）。

② 乔姆斯基：《ChatGPT 的虚假承诺》，龚思量译，澎湃新闻思想市场专栏，https://baijiahao.baidu.com/s?id=1759939491843688675&wfr=spider&for=pc。

③ 哈贝马斯：《遗传学的奴役统治：复制医学进步的道德界限》，载哈贝马斯：《后民族结构》，曹卫东译，上海：上海人民出版社，2002 年，第 218 页。