



潘多拉魔盒还是文明利器

——大文明视野中的 ChatGPT 反思

ChatGPT 横空出世，引起全球轰动。此类生成式人工智能到底意味着什么？是人工智能回到生活世界为人类赋能，还是出现了一种人类创造但可能失去控制的不祥力量？换言之，ChatGPT 是文明利器，还是潘多拉魔盒？此问题的回答关乎人类文明的未来走向。2023 年 3 月 13 日，《探索与争鸣》编辑部联合复旦大学哲学学院、《广州大学学报》（社会科学版）编辑部以及“复杂现代性与中国发展之道”课题组共同举办了“潘多拉魔盒还是文明利器——大文明视野中的 ChatGPT 反思”高端论坛。不同领域的专家各抒己见，围绕生成式人工智能与人类交互的前沿问题和应用领域、生成式人工智能对人类交往的影响、人工智能界限与科技伦理、大型语言模型技术及其对知识学习的影响等议题展开了内容丰富的讨论。

汪行福教授指出，以人工智能为基础的人机一体化完全颠覆了以往人类技术发明的逻辑，打破了人类纪文明的边界。要想避免 ChatGPT 对文明的反噬，技术批判固然重要，但更重要的是在新技术出现时寻找到作出合理判断和采取行动的机会。邱锡鹏教授从技术和原理的角度深度解读 ChatGPT 模型。他指出，大型语言模型在达到百亿级规模后开始获得“涌现能力”，这背后进一步隐含着三个非常重要的技术，即情景学习、思维链和自然指令学习。陈云松教授认为当下我们谈人类的恐惧和担忧为时过早，即使对一个学科或分支来说，AI 技术还远未起到颠覆性的作用。在目前阶段，研究者更需要充分理解 ChatGPT 的特点和存在的问题，以及它对人文社科的有限工具价值，在运用该技术的同时，对其可能产生的问题保持警醒并积极应对。围绕生成式人工智能构筑的“赛博格阶梯”，杜骏飞教授阐释了人类赛博格化的四部曲：第一步是交往的赛博格化，其表征是 AI 辅助交往和数字交往 2.0；第二步是主体的赛博格化，其表征是知识嵌入的多项式生命；第三步是认知的赛博格化，其表征是生理嵌入的 AI 增强人；第四步是意识的赛博格化，其表征是人类开始像 AI 那样思考，并期待拥有机器的心灵。严锋教授指出，技术会让我们获得一种面对传统人文学科的新眼光，这个问题从文学角度来看也许会更加突出，因为文学有一种对于原创性的执着追求。王峰教授指出，我们需要从一种革新的哲学视角来判断人工智能与人类自身的关系，把 ChatGPT 之类的大型语言模型视为通用人工智能的初步成就，但需要将思考与自我意识的获得彻底分开。哪怕是未来的强通用人工智能，也只是不断增强思考的功能，无法获得人的自我意识。刘云杉教授认为，技术之于教育，不再仅仅是教学辅助物，也不仅止于智能的教育者，它更是一个升级的、可以替代人类的新型学习者。ChatGPT 既是“类人”的学习者，更是“超人”的竞争者。如果不加限制，高等教育大众化后出现的白领无产阶级化现象将进一步加剧。朱鸿军教授指出，ChatGPT 直接对标的是人的智慧，其带来的不仅是对新闻传播系统的颠覆性影响，更是对整个人类系统、人类文明的颠覆性重构。如何让其始终处在人类的控制范围内，需要包括新闻传播系统在内的全人类各系统集体警觉、探究反思和寻找路径。王金林教授从知识生产方式、语言生成方式以及人的对象性存在方式三方面考察了 ChatGPT 技术给人类社会带来的变化，认为 ChatGPT 及其迭代产品不仅是一种新工具，而且预示着一一种统摄一切的新存在。余南平教授指出，ChatGPT 诱发了新一轮人工智能革命，其在历史进程中不可避免地嵌入国家角色并产生作用力，也不可避免地引发国家间围绕通用人工智能革命的新博弈与新竞争。在国家开拓技术新边疆的努力下，人工智能技术会被国家的主动性和主导性塑造。

——主持人 叶祝弟 张 蕾 屠毅力

后人类纪文明的到来与 ChatGPT 的终极之问

汪行福，复旦大学当代国外马克思主义研究中心研究员、
哲学学院教授

ChatGPT（以下简称 GPT）对人类来说，到底意味着什么？比尔·盖茨在自己的博客中写道，这是他人一生中第二次被科技真正震撼到；基辛格认为这是继印刷术发明以来人类最为重要的技术发明。^①

GPT 的潜能巨大，将在极大范围、极深程度上影响到人类当下和未来的生活，其意义可放在上下五千年文明大背景下讨论。GPT 能量巨大，我们对它的认识已呈现出两极化倾向。一些人把它视为赋能（empower）工具，“百”求必应。可以设想，如果它与互联网连接，实现实时查询，就成了 web 版的 GPT，利用自带的翻译功能，就可遍访人类所有的知识库，成为全球版的 GPT。这些不是想象，而是当下现实，是正在发生的事情。也有一些人认为 GPT 存在着致命的缺陷，就像一个黑箱，不透明，缺少反思和批判能力，与人的学习能力相差甚远，更为严重的是，它经常产生幻觉（illusion）和偏见（bias）。正反两方面的观察和评估给我们提出一个严肃的问题：GPT 是对人类的赋能还是人类自我罢黜（dethrone）的陷阱？人自诩为万物之灵，是天地之间唯一的有意识的动物，是能说话、有思维和推理能力的动物。如果 AI 有了类人式意识，又没有人在记忆和耐力上的生理限制，它是否会把人从“王座”上推下去，自己登上“王位”？

GPT 无疑是一个伟大的技术突破，它成功地使人工智能自然语言化，为人类创造了知识生产和运用的母体平台，这有可能彻底地改变人类文明的底层逻辑。人类的文明是以智人大脑为基础的自然语言文明。以人工智能为基础的人机一体化完全颠覆了以往人类技术发明的逻辑，其侵入人类智能领域，打破人类纪文明的边界。GPT 到底是对人类的赐福还是诅咒？这是我们需要面对的“终极之问”。

以智人大脑为基础的自然语言文明

GPT 是大型自然语言 AI 对话程序。之所以引起轰动，缘于其出色的技术优势。概括起来，即通用性、生成性和增强性。这些特点缘于 AI 的自然语言处



① 基辛格：
《ChatGPT 预示
着一场智能革
命，而人类还没
有准备好》，
<https://new.qq.com/rain/a/20230228A02K6Q00>。

理能力和大数据模型的结合。这一技术革新意味着知识的获取、创造和运用的母体或平台的大迁移。

人工智能革命出现之前，科学技术已取得了一系列重大的突破和成就。一方面，人类能上天入地，现代交通工具把全球联系在一起，不仅创造了地球文明（global civilization），如果火星移民计划实现，还会出现星球文明（planet civilization），这是人类在超距文明尺度上取得的成就；另一方面，现代物理学、化学和生物学研究已进入物质的内部，深入到感觉之下的细胞核、原子、分子层面，有了改造物质微观结构的能力，这些是人类在微距文明尺度上的成就。然而，在人工智能取得决定性突破之前，人的智能、意识、思维和理性仍然是技术的禁区，是“化外之地”。在根本意义上，超距文明和微距文明影响的是人类生活的环境，没有影响到人为人的特征，即人的意识和思维这一核



心领域。GPT 的决定性意义正在于它突破了这一禁区。如果将以智人的大脑结构和自然语言相结合为基础的人类文明称作人类纪文明，GPT 以及人工智能生成内容（artificial intelligence generated content, AIGC）技术出现后的文明则可称为后人类或后人类纪文明。^①

人类文明史经历过多次大的转型，从采集文明到农业文明，之后进入工业文明，再到后工业文明。每次文明变革都有其祸福相依的一面，既是对人的赋能，也意味着原有的人类活动及其形式的重组，这也不可避免地导致对收益和代价的重新分配。就此而言，GPT 与以往的文明变迁没有太多区别。然而，作为自然语言的人机对话平台，GPT 打破了智人大脑的生理限制，也突破了自然语言技术的限制。这一点具有深刻的意义。如果 GPT 和 AI 的发展去除了上帝为人类文明设定的限制，从根本上改变了人类文明的底层逻辑，那么它们的出现确实意味着人类文明史的奇点。巴别塔的倒塌意味着以智人大脑为基础的自然语言文明的到来，而 GPT 的出现则意味着以 AI 为基础的后人类文明时代的到来。

自然语言 AI 大模型意味着什么

人是语言的动物，语言的限度就是思想的限度。然而，自从现代科学诞生以来，特别是现代哲学的语言学转向以来，在语言上存在着两种范式的竞争，一种是自然语言，一种是人工语言。自然语言是文化的语言，人工语言是科学的语言。自然语言与人工语言各有其偏爱的领地。自然语言的独特领地是人文科学，它为人的丰富思想和情感表达提

供了条件，也为民族认同建构提供了神圣的象征。然而，自然语言有其优点，也有其缺点，它承载着人类生活世界的背景知识，是人与人交往的中介，它的生动性、多义性为文学和艺术提供了得天独厚的条件。但是，自然语言的歧义性和含义的模糊性也给知识的表达和传播带来了消极影响。正因为如此，一些人仍梦想巴别塔之前的同一语言，用一种世界语取代五花八门的自然语言。

人工语言传统虽然很有影响力，但在人类生活中仍然是自然语言占统治地位。亚里士多德的三段论和形式逻辑中也运用了符号，但它们仅仅用于表示判断之间的逻辑关系。从莱布尼茨开始，哲学家们就孜孜以求建立一套人工语言系统，用以表达人的思维和推理，并把算术机器化。人工语言的理想直到符号逻辑和符号数学的出现才部分得以实现，即使如此，符号逻辑的运用范围也是有限的，它始终被限制在语言表达式之间的逻辑关系层面，并没有进入思想内容本身。在人工智能取得突破性进展之前，语言的人工智能化陷入这样的困境：或者接受一个应用范围有限的可技术化处理的人工语言系统，或者延续自然语言传统但放弃对其进行人工智能化。GPT 的出现打破了这一僵局。作为自然语言人工智能系统，它不仅通用于各种自然语言，而且具有自然语言的生成性和对话性。

GPT 最大的特点是其生成性。它可以根据语言或文本提示生成新的内容，完成文本写作、绘画、编程、解码等任务。在受到大量的“喂食”和一定时间的训练后，大模型会“涌现”出各种神奇的能力，如信息搜索、历史记忆、上下文理解、推理能力、与人流畅对话等。在此之前，这些能力被认为只有人才具有，且只能通过自然语言充分实现。AI 大模型的能力涌现引起了广泛关注。能力涌现是指在无监督情况下拥有理解、对话、创作和处理复杂问题的能力。虽然我们已经有的一些关于 AI 能力涌现的技术条件以及大模型参数与能力涌现之间变化轨迹的经验性结论，例如能力涌现需要多少参数和训练量，参数的增加与能力涌现之间的非线性和倒 U 形轨迹等，但我们对其产生机理所知甚少。因此，AI 专家表示：“我们对 GPT-4 的研究完全是基于现象学的，我们关注的是 GPT-4 能够

① 按通行的理解，人类纪文明自近代使用化石能源开始，因为人类活动影响到地球表面的生态和近地大气层。简要地说，人类纪文明就是工业文明。后人类文明是一个宽泛的概念，有许多含义，其断代史时间也各不相同。罗西·布拉伊多蒂在《后人类》(The Posthuman) 中主要把它理解为后人类中心主义价值和后现代主义观念。本文中使用的后人类文明概念更加激进，即一种基于 AI 和大数据的人机一体化的文明。

做到这些令人惊讶的事情，但我们并不知道它是如何变得如此智能的。”^①

对于“涌现”现象的思考在哲学中并不陌生。海德格尔认为，运用语言的思想活动与其说是受人的意识的控制，不如说是语言本身的自发运动。真正的思想不是说话者的思想，而是思想通过语言自行“道说”，而自然语言正是思想最贴近的表达方式。真正的文学创作不是“为赋新词强说愁”，而是让自己沉浸于思想中，让思想自由表达。一个人沉浸于思考之中，会文思泉涌。对此，伽达默尔说：“‘表达’并非主观选择的问题，即并非限于事实之后，并且借助于个人思想中的意义被赋予了可传达性而加于事实之上的某种东西。”^②海德格尔、伽达默尔等人对思想与语言关系的论述似乎先行地道出了 GPT 的逻辑，即自行思考、自行生成和自我表达。实际上，与 GPT 的对话之所以具有这样的特征，不仅在于它能够很好地模拟人的自然语言，而且在于其本身表现出了类似于自然语言系统的特性。如果说 AI 是对人脑神经网络的模仿，那么 GPT 则是对人的自然语言对话系统的模仿，如模糊认知、解释、交互学习等。因此，从现象意义上来说 GPT 具有类人式意识并不为过。

问题在于，GPT 不仅具有类人式意识，而且具有超人式意识。依笔者所见，人类文明建立在自然语言和人脑思维之上，两者的结合既提供了人类文明发展的空间，也限制了它的发展。GPT 的出现截断了人脑机能和自然语言之间的联系，自然语言对话平台从人际间转移到了人机之间，甚至未来可能会发展到机器之间的对话。这是人类文明底层逻辑的重大变化。在超大型数据库（人类知识储存）和计算机超级算力（思维力）的加持下，语言生成模型的能力可以远远超过人类。就像计算机在计算领域取得的成就一样，AIGC 也可能在自然语言的理解、解释和生成方面拥有类似的能力。例如，AI 将人类的思考和推理等思维活动从自然人的智能平台转移到了人工智能平台，就像汽车从崎岖不平的土路走上了高速公路。在这个过程中，无数信息通过无数节点迅速连接，瞬间完成各种任务。我们不知道它是如何做到的，但它给人们呈现出一个“all-knowing machine”或“全知之神”的形象。

对 GPT 的近忧远虑

海德格尔曾区分“怕”与“畏”，认为“怕”有具体对象和缘由，可计算得失；“畏”则不同，它是对无之畏，是对存在之整体的沉沦的畏，在此，所有存在者被抛出自己的轨道，进入无根无据的深渊。面对 GPT，我们不由地生成复杂的感情，它不仅是“怕”，而且也是“畏”。

在现象层面，每一次技术创新都有复杂的伦理和社会后果。(1) 像每一次技术创新和大规模运用一样，GPT 不可避免地使人际间出现收益和代价的再分配，催生一些行业，毁灭一些行业。如果我们承认，人的尊严与其劳动中体会到的自主性有关，GPT 的出现无疑会对那些被剥夺了劳动机会的人的生活造成重大损害。(2) 虽然 GPT 有问必答，有极强的亲民性，但是这不等于技术的可及性，其有可能造成新的不平等，加深数字鸿沟。(3) 与任何节省体力和脑力的技术工具一样，GPT 的使用中会产生依赖性风险，这在教育和学习领域影响尤为显著。GPT 可能提高人的学习效率，也可能使人因对其依赖而丧失学习动力。(4) GPT 是自然语言的人机对话程序，形似人与人的对话，但毕竟不是人与人的对话。如果交往是人与人的关系和社会融合的条件，GPT 的出现可能会导致人类出现逃避其同类的倾向，强化人的孤独感，影响到以交往为基础的人类文明的各种组织形式，如婚姻、家庭、宗教、民主政治等。(5) GPT 表面上具有亲和性，对人有求必应，其给出的回答也不乏个性，表现得就像一个知心朋友，然而大数据模型训练出来的回答到底有多少个性？它向人呈现的个性

①《是什么让 ChatGPT 变得如此聪明？仍然未知的大语言模型“能力涌现”现象》，新浪财经，2023 年 4 月 11 日。

② Hans Gadamer, *Hegel's Dialectic*, trans. by P. Christopher Smith, New Haven: Yale University Press, 1982, p.32.

化是否只是一个幻影？诸如此类问题都是现实的，也是需要不同领域的研究者去探讨的。

真正引发人们担忧的是本体论层面的问题：GPT 到底是什么？我们到底能对它做什么？前者可称为透明性问题，后者可称为可控性问题。在任何时代，完全的透明性都是不存在的。长期以来，自然界对我们就是不透明的、充满奥秘的世界，只是这种不透明在哲学家和诗人眼中是神圣的，而技术的不透明却是有罪的。收音机发明出来后，海德格尔就对那个小黑盒里面发出的声音感到恐惧。因为与风车和水车相比，收音机是“不透明”的，我们看不见它如何工作，即使拆开来也是如此。GPT 更是如此，甚至 AI 技术人员也会对它的神奇能力感到困惑，更不要说外行人了。在某种意义上，GPT 就是一个黑箱，它对人类所提出的任何问题都能瞬间给出回答，且正确率越来越高。这不能不让人感到困惑：它到底是如何做到的？按照海德格尔对“怕”和“畏”的区分，在 GPT 黑箱面前我们感到的应当是畏，即对莫名的风险的恐惧。

在本体论层面上一个更大的问题是 GPT 的可控性问题。人的思维依赖于大脑神经网络，通过感知激活大脑中存储的知识和记忆，通过反思、推理形成自己的判断和观念。在这里，自然语言和符号成为记忆和感知的中介，而 GPT 则不同，它有自己的记忆，其记忆体即数据库可对电子数据信息一网打尽，它有自己的“大脑”，即神经网络 AI 程序，它能够生成新的知识。将来，它通过电子感应器还可以感知事物，与机器人相结合直接作用于物理世界。

今天，GPT 仍然需要依赖人的“提

词” (prompt)，但我们可以想象一个无人的 AI 世界。黑格尔在《精神现象学》中有一个著名的命题——“实体即主体”，他把精神视为比人更高的主体，它不仅通过自身运动创造世界，而且通过自我反思认识自己。然而，黑格尔的“绝对精神”与波普尔的“世界 3”一样，它们的自主性在某种意义上仍然是拟人化的，是人类生命的投射。按通常的理解，无论它如何神奇，归根到底 GPT 是一个智能系统，是物，是实体。但是，如果 GPT 不仅具有了人的意识和思维能力的表现，而且在加速迭代中获得远超出人的能力，不仅能为人所为，而且能为人所不能为，我们还纠缠于它是否有意识、是否是主体就失去了意义。

对人类文明来说，GPT 的真正挑战在于我们是否创造出了一个能够彻底逃逸人的控制的物，无论我们把它称为人工理智还是超级智能。如果是后者，它将导致人类“王位”的废黜。阿多诺在《否定辩证法》开篇曾指出：一度似乎过时的哲学由于错过其实现时机而得以保留下来。同样，一度过时的全知全能理想因没有实现，人类才将之保存下来。今天，我们面临的悖论是，也许 GPT 正在实现人类的全知全能幻想，成为新的“通天塔”，但它的成功会是人类的成功吗？GPT 是人类新文明的开端，还是人类文明终结的开始？

如何避免 GPT 对文明的反噬

科幻小说家阿西莫夫 (Issac Asimov) 于 1942 年提出了著名的机器人三定律：一是机器人不得伤害人类，或者目睹人类遭受危险而袖手旁观；二是机器人必须服从人给予它的命令；三是在不违背第一、第二定律的情况下，机器人要尽力保护自己。GPT 是否伤害人类取决于第二定律。包括 OpenAI 团队在内的专家都没有对它的运行机制，特别是其神奇的能力涌现现象给出解释，自然也就无法保证它百分之百执行人类的命令。应对 GPT 这一既具有强大赋能能力又潜藏巨大风险的技术也许没有万全之策，但并不意味着无需对此进行思考。

当前，各“大厂”纷纷建立自己的大模型，呈现

出不可阻挡的“进步”特征。这样的情况下，如何思考人类孜孜以求的进步？是否在任何情况下都要拥抱进步？本雅明的思考也许有一定的借鉴意义。他在《历史哲学提纲》中借助保罗·克利的画作《新天使》对盲目追求进步的冲动进行了批判。在本雅明看来，人类的救赎需要对过去的沉思，特别是对过去灾难的沉思。他把这一救赎的历史意识拟人化地描述为“历史天使”。历史天使想停下脚步唤醒死者，修补破碎的世界，“可是从天堂吹来了一阵风暴，它猛烈地吹击着天使的翅膀，以致他再也无法把它们收拢。这场风暴无可置疑地把天使刮向他背着未来……这场风暴就是我们所称的进步”。^①当然，本雅明观点难以直接运用于 GPT，但面对当前人工智能“竞赛”，本雅明式批判对我们不无启示。从时间结构上看，进步和灾难是非对称的，进步依赖于时间的连续性，而灾难总是不期而遇的，必须在火花碰到炸药之前将燃烧着的导火线切断，^②因为“对于个人，如同群体蒙受的苦难一样，只有一个临界存在，超越了它，‘事情就不会再这样继续下去了’，这个临界点就是毁灭”。^③今天的 GPT 是否逼近了这一临界点我们并不知道，但我们不能失去警惕。

海德格尔和马尔库塞的技术批判也有助于认识 GPT 的威胁。在他们看来，除了技术滥用和工具理性，技术本身也在对自然和人进行统治。统治的目的不是后来补充进去的，也不是人从外面强加给它的，而是早就进入了技术的设计之中。海德格尔把技术理解为存在的解蔽，即存在向人类敞开的形式。然而，与艺术等其他敞开形式不同，“在现代技术中起支配作用的解蔽乃是一种促逼 (herausfordern)，此种促逼向自然提出蛮横要求，要求自然提供能够被开采和贮藏的能量”。^④马尔库塞也认为，现代技术不仅是生产工艺学，而且是统治工艺学。就此而言，不仅要批判技术中立化教条，而且在任何重大技术创新的源头和开始处就必须认清存在何种危险。

技术批判固然重要，但更重要的是在新技术出现时寻找到作出合理判断和采取行动的机会。在安德鲁·芬伯格看来，“技术不是一种天命，而是斗争的舞台。技术是社会的战场，或者用一种更好的隐喻来说，

把技术比作一个文明的替代形式相互竞争的‘事态的议会’”。^⑤虽然技术发明者可以设定技术的目标和价值标准，但技术使用者和接受者也不完全是待宰的羔羊。底层人民可以在流动的边缘上与技术发明者和主导者进行缠斗，改变技术的具体形态和使用方式，实现技术的民主化。技术的民主化简单地讲，就是让处在边缘的弱势群体和受制者的利益得到体现，使技术的多种潜能得以呈现，让他们在技术面前有更多的自主性。技术民主化可以以多种方式展开，如不同利益群体之间的辩论，公众参与技术设计，对技术进行再发明和改造。技术的冲突实际上是两种“自主性”的冲突，一边是技术系统本身的操作自主性，另一边是技术运用者实现自身自律的操作自主性。现代技术体制的主要缺陷是那些因技术而受到影响的人对设计和运行很少或没有控制权。提高人类的自主性涉及技术的规划、建造和控制过程尽可能向那些注定要体验最终产品和社会后果的人敞开。芬伯格的技术民主化方案对我们思考如何应对 GPT 有重要的借鉴意义，虽然这一方案需要结合 GPT 技术的特点和当下社会条件转译为合适的斗争战略和策略。

本雅明和芬伯格代表着两个极端，本雅明为我们提供的是最终决断，芬伯格提供的是技术斗争的日常政治。两种立场并非决然对立，具体的选择取决于对事情本身的判断。

[本文系国家自然科学基金重大项目“复杂现代性与中国发展之道”(15ZBD013)的阶段性成果。]

① 汉娜·阿伦特编：《启迪：本雅明文选》，张旭东译，北京：生活·读书·新知三联书店，2008年，第270页。

②③④ 本雅明：《单行街》，南京：凤凰出版集团、江苏人民出版社，2006年，第91页，第91页，第932页。

⑤ 安德鲁·芬伯格：《技术批判理论》，韩连庆、曹观法译，北京：北京大学出版社，2005年，第16页。

ChatGPT 所展现出来的令人惊艳的对话理解与表达能力，让越来越多的人意识到人工智能已经迎来了全新的里程碑，并且有望成为一个至关重要的基座系统，其以前所未有的速度渗透各行各业，持续引爆未来世界的数字经济体系。



ChatGPT 背后的技术原理

ChatGPT 背后的技术原理到底是什么呢？在正式解答这个问题之前，我们有必要先了解一下“语言模型”，即利用计算机去重新建模人类语言，把自然语言转化为机器能够理解并判断的语言的一种机制。人类的自然语言是十分灵活的，它一方面有规则，另一方面又随时可以打破规则，甚至会存在巨大的歧义性，任何一句话放在不同的情景下大家对其理解都会大不相同，这就给建模造成了极大的困难与挑战。

我们可以利用概率判断来进行语句处理。如果这个句子符合自然语言的规律，我们就赋予它一个比较高的概率，反之，则赋予它一个比较低的概率。不过这样的话，新的问题又产生了：我们到底应该怎么去赋予句子以概率？这就

解剖大型语言模型：原理、应用及影响

邱锡鹏，复旦大学计算机科学技术学院教授、Mosschat

团队负责人

需要我们从互联网上获取海量的文本数据。但这也是一个难题，这个概率空间很大，难以直接建模。目前的解决方案是通过把整个句子的联合概率拆解成其中每个词的条件概率的连乘，大幅降低语言模型的建模难度。这样语言模型就转换为给定上文预测下文的机器学习任务。目前，大型语言模型使用的神经网络架构是“Transformer”。这个词的本义是变压器，有输入和输出，是一个类似于双塔式的结构。带入语言模型里面，就是进去是一个句子，出来是另一个句子，给定上句预测下句的特定网络模型。

现如今 Transformer 已经在整个人工智能领域成为最主流的架构体系。除了能力强之外，还有一个非常重要的原因在于它的架构设计对于 GPU 运算特别好。因为 Transformer 架构和卷积、循环神经网络不一样，卷积、循环神经网络都是在 GPU 诞生之前设计的，而 Transformer 则出现在 GPU 之后，所以设计方面就可以天然地充分考虑最大限度利用 GPU 能力，更容易做到大规模语言模型的效果。

不过，要想模型的预测更加准确，就势必需要用大量的参数对其进行训练，以帮助大规模语言模型充分理解人类语言规则及其逻辑关系。目前在 Transformer 的加持之下，我们已经可以把模型规模做到百亿级、千亿级，这就是如今的大型语言模型（large language model, LLM）。语言模型背后的仿人类神经元构成的 Transformer 神经网络可以对我们输入的“上文”进行加工，从而实现对下一个词的预测，并输出相应的文本。

一个好的语言模型越想预测得准确，就越需要充分理解人类语言和常识。比如让模型预测鸡蛋是圆的而不是方的，本身便隐含着一定的生活常识在里面。为什么神经网络可以学习到这些常识呢？这里面有一个信息压缩的概念，比如有 1 万亿个词，我们可以把

它原封不动保存到硬盘上，也可以用 10 亿个参数的神经网络把所有的语料知识记住。因为这里面有很多信息是冗余的，这就迫使神经网络不但要把这些浅层的知识存下来，还要把其中所蕴含的知识、规律累计下来。道理很简单：把规律记下来就可以胜过很多浅层的文字。通过这样的方式，我们迫使语言模型努力发现这些文字背后所蕴含的各种各样的规律，从而使这个模型能够更好地理解人类语言和世界知识。

涌现能力背后隐藏着哪些关键技术

在训练过程当中科学家发现，计算量大概在 10 的 22 次方之后，模型能力会完成从量变到质变的飞跃，呈现出惊人的爆发式增长，通常称之为“涌现能力”。大型语言模型在达到百亿级规模后开始获得涌现能力，其背后则进一步隐含着三个非常重要的技术，即情景学习、思维链和自然指令学习，这也是 ChatGPT 得以在人工智能领域叱咤风云的关键原因。

(1) 情景学习 (in-context learning)。它深刻改变了传统机器学习的范式，只需要用一系列精心设计的提示语句 (prompt) 对任务进行详细描述，然后再辅以一些情景例子，就能够让模型参考着既定例子完成特定任务。这种模式跟传统的将知识直接储存在参数里显然是不太一样的，也在一定程度上揭示了为什么 ChatGPT 通常是以多轮对话的形式来呈现。

事实上，ChatGPT 每一次接收人类发来的对话，都会把之前的聊天历史全部作为上文输入给语言模型，然后由语言模型续写出一个下文并反馈给用户。这种让一个大型语言模型直接和人类进行交互，从产品创新的角度来看确实是非常有智慧和远见的。

(2) 思维链 (chain-of-thought)。模型能力可以通过扩大参数规模来提升，但是研究人员却另辟蹊径，想了一个更好的方法，即把一个复杂问题分解为多步推理的简单问题，让模型能够明白并学习人类到底是怎么一步一步推导到这个答案的，这个就叫作思维链。

大型语言模型经过海量预训练之后，已经见过了很多推理方式，我们只需要一步一步引导，就可以让它按照自己想要的方式去推理。思维链方式进一步解

放了模型潜力，让本来模型不会解的一个个复杂问题分解成很多简单问题，然后通过逐一解决简单问题，最终使复杂问题迎刃而解。

(3) 自然指令学习 (learning from natural instructions)。传统上，机器学习都需要大量标记数据，才能让机器学习模型从数据当中实现学习。但是标数据本身又是个非常烦琐的事情，人类总归是希望语言模型最好能够直接从指令中进行学习，能够直接听懂人的意图。

而事实也证明，这个思路是可行的，甚至人类只需要在少量的任务上进行指令化，在经历大概四十多个任务指令化之后，对模型进行适度微调，就很容易泛化到上百、上千种任务，即使那些任务可能从来没有见过，它仍旧能够很好地进行应对。

尽管自然指令学习大幅提升了泛化能力，但是对人类的真实意图的理解仍然有着很大差距。OpenAI 就希望收集真实的人类意图，让专家来写答案，从而更好地与人类偏好进行匹配。在这个过程中，人的参与意义重大，其使得机器在迭代中始终保持和人类的价值观、思维方式对齐，也避免了机器自己迭代起来有可能会离人类的偏好和初衷越来越远。

大型语言模型是如何炼成的

在了解了 ChatGPT 背后的这些基本技术原理之后，我们大致可以尝试复现这一大型语言模型了。虽然这些关键步骤以及发展大势都已经十分明朗，但每一步的细节都需要我们自己去一一摸索，还是充满着各种未知的挑战。

第一步需要先实现语言模型基座，

主要是在 Transformer 架构上进行模块优化。对于 ChatGPT 来说，它并没有特别关注中文，很多时候只是直接把中文按照英文的方式进行编码，我们自然是希望对中文进行优化，这就需要重新实现更好的中文编码，并想办法把中文和英文打通。将来如果接入多模态的话，编码问题同样会带来架构设计及训练稳定性等方面的诸多问题与麻烦。

第二步是指令微调，其难度甚至比预训练更高。在预训练阶段，利用一些大公司成熟的预训练模型，可以在短时间内取得不错的训练效果，但是指令微调则难以立马做到，这一点上后来者和 OpenAI 之间存在着非常明显的差距。

第三步是能力不断强化迭代。在与人类对齐方面，想要让模型的回答尽可能符合人类的思维习惯，也很难做到。考虑到 OpenAI 暂不开源，我们只能一步一步慢慢向前探索。如果我们想要超过 ChatGPT，就得去找到一条比它更好的实现路径，而这个过程无疑充满艰险。

具体到笔者所负责的大型语言模型 MOSS 的实现方案上，首先要激发其对话能力。因为高校科研团队没有能力像 OpenAI 公司可以雇很多人去写答案，所以我们一开始是通过自指令技术先写一些种子，通过 GPT-3 帮我们扩展很多的条件语句和答案。写完之后，有一些小规模的对对话数据开始产生，我们就可以进一步利用监督策略做一个模型出来，然后在此基础上进行能力提升，让它能够和人类的真实需求逐步对齐。要让机器自己去写答案，鼓励它按照人类的偏好不断实现迭代优化，去生产越来越符合人类习惯的数据类型。

整体而言，MOSS 基于公开的中英文数据训练，拥有近 200 亿参数，具

有和人类对话的能力，并可以通过与人类交互实现迭代优化。尽管 MOSS 对语言的理解能力和 ChatGPT 已经非常相近，但是目前 MOSS 的回答完成度相比 ChatGPT 还有很大的改进空间，主要原因在于 MOSS 在部署和训练上投入的资本还非常有限，还有很多事实类知识记不住，思维链能力也相对较差。目前我们也在积极尝试引入一些外部工具，来进一步扩大模型参数规模，不断实现改进与优化。

AI 如何对未来社会形成颠覆性影响

考虑到 ChatGPT 已拥有通用语言理解的能力，并可以进一步增加对外接口，成为通用人工智能的技术基座，这也就意味着通用人工智能的加速实现对于现阶段的人类来说已经不是梦想。

以 ChatGPT 为代表的通用人工智能技术，可以引爆数字经济，充分发挥数据和算力效能，并催生出海量的新商业模式。它可以赋能产业数字化，并通过人机协同的方式解决行业专家资源不足的问题。数字人、个人助理、搜索引擎等形式也为数字经济新业态、新模式提供了发展动能。此外，通用人工智能技术还可以深刻改变教育、社会治理、司法等领域的生态，大幅提升行业水平。

当然，我们也必须正视，目前的通用人工智能技术尚且存在很多不足，包括随机性、不可控、容易“一本正经地胡说八道”等，不过，相信这些方面都将在未来通过各种方式逐步得到改进。或许正如图灵奖得主、著名人工智能专家杨立昆所言，下一代模型应该更加具有事实性、无害性、即时性，并且能够灵活使用计算器、数据库、搜索引擎、模拟器等各种辅助工具，而这也是人们亟须重点解决的时代课题。

对于下一阶段的大型语言模型来讲，我们目前重点需要去做的事情就是让模型和现实世界以及人类的价值观“对齐”，让模型成为一个真正的智能体，具有自身学习、跨模态学习、知识和工具利用等能力。与此同时，AI 与人类价值观“对齐”也同样不可忽视，毕竟如果 AI 的价值观和人的价值观背道而驰，那将会是非常危险的。

ChatGPT 对人文社科研究的价值与挑战

陈云松，南京大学社会学院教授

相较于自然科学，人文社会科学与文化、社会的距离更近，因此，当 ChatGPT 问世时，人文社科研究者就开始讨论人工智能与生成式 AI 技术对人类文明的影响。然而，人文社科研究者的担忧是出于学科本身的特点，也就是人类的“终极恐惧”——失去对语言的垄断。事实上，在 AI 技术发展的早期，我们就已经开始了对人类终极进路问题的探讨，当下我们来谈人类的恐惧和担忧为时尚早，即使对一个学科或分支来说，AI 技术还远未起到颠覆性的作用。在目前阶段，研究者更需要充分理解 ChatGPT 的特点和存在的问题，以及它对人文社科的有限工具价值，在运用该技术的同时，对其可能产生的意识形态问题保持警醒并积极应对。

从 ChatGPT 看人文社科对技术的隐忧

正如印刷术使得知识的大规模生产成为可能，也让当时的权贵恐惧于知识普及所带来的启蒙力量，自工业革命伊始，人类在利用工具和技术的同时，也一直处于对它们的恐慌之中。人们恐慌的不是技术本身，而是技术力量的无限延展对自身权益的侵蚀。从人文社科专业角度来看，研究者对技术的天然担忧代表着对更终极维度的思考——新的技术会不会影响人的自由意志？从计算机、互联网，再到 ChatGPT，人们对技术的恐惧发生得越来越频繁。然而，当今社会人类依然处于“自我恐吓”的阶段，因为历史的经验告诉我们，人被工具所统治的前提是失去语言能力。人类虽然一直被工具所增强和改造，却从未被工具所统治。

当然，在恐慌以外，学者对工具和技术的反思依旧有其必要性。在人类历史的长河中，每一次新技术、新工具的迭代都塑造了人之新“生”。这里的“生”不仅仅指人的肢体，还涉及人类社会的各个方面。技术的增长可以发挥肢体的工具性作用，协助人类不断进

化，向更高等的文明攀登，也可以像肿瘤细胞一样导致人类走向灭亡。这个决定权掌握在人类自己的手中。人文社科领域的学者们大多站在个人意志思维、社会宏观变迁的理论高度上，以人类自由的命运为终极目的，审视人工智能领



域所面临的道德、伦理问题。虽然这些问题看似“外行”，有时候甚至过于简单直白，但是人文社科学者的努力与忧虑，恰恰能提醒技术人员沿着符合全人类共同价值观、延续人类共同文化知识传承的方向前进。根据这一特征，与人工智能领域对 ChatGPT 的工具性价值已有深入理解不同，人文社科领域学者对 ChatGPT 的思考大多集中于其对象意义，也即该技术的出现对社会文化发展路径的影响。

人文社科学者对 ChatGPT 的另一重隐忧是，它打破了精英内容生产者对内容生产的垄断。作为学者，我们一直以知识精英型内容生产者自居，并擅长运用自身所属领域的话术对热点问题进行“高屋建瓴”的指导。但现在我们发现，ChatGPT 在内容生产方面未必比所谓的专家逊色，甚至用户诟病的可靠性问题也在新发布的 GPT-4 模型中通过更精确的神经网络预测获得了一定程度的



解决，从而在一定程度上打破了我们自满、自得的思考壁垒。可以预见的是，未来的人类将从规模化生产中解放出来，在获得自由的同时，生成有别于单纯收集整理而更具创造性的思维模式。该模式在今天获得了第一个应用：如何反思 ChatGPT 工具在人文社科领域的局限性，以及随之而来的挑战？

ChatGPT 对人文社科的 有限工具价值

作为非决策式 AI，ChatGPT 使用的是联合概率模型，它的学习模式更像人类在婴儿时期依靠联想和模仿来学习语言，在和对象的反复接触中明白词语的具体指向。在成长过程中，人类需要在多年里不断依赖周围能获取的信息数据来训练自己的语言能力，ChatGPT 也是如此，它需要对海量数据进行深度学习。ChatGPT 的底层逻辑是模仿人类的语言机制，因此其表述方式更接近一个真正的人类，很适合文本内容的生产。这决定了它对人文社会科学学科的工具价值。首先，它能够辅助学术研究的前期准备工作，如综合性知识问答和信息检索。这一功能相当于虚拟“教师”，整合和检索知识，给出综合性的判断和具有针对性的答案，它既可用于知识问答和推理，也是很好的资料收集工具，能以对话的方式帮助研究者更快捷地管理和检索信息。其次，ChatGPT 出色的数据分析和代码能力可以助力“爬虫工作”，对于需要进行定量实证分析的社会科学学者，特别是致力于进行人文学科跨领域分析、大数据研究的学者来说尤为关键。技术学习一直是人文社科计算领域的一大门槛，但在人文社科计算

的帮助下，学者们可以凭借相对基础的代码能力进行数据爬取、模型计算。如此一来，在 ChatGPT 降低人文社科计算准入门槛的同时，新方法、新规律、新理论的创造概率被极大提高。本质上说，ChatGPT 促进了各学科之间的交流互鉴，进一步推动了人文社科领域组织化、创造性的科研。最后，ChatGPT 能够辅助文章、书籍、项目申请书等文本内容的直接生成，提高工作效率，使学者能够把研究的重心转移至核心方法、核心理论、核心思维的攻坚克难上，以助力更多的学术创新。

需要特别关注的是，ChatGPT 对人文社会科学学科的意义和作用尚停留在表层，其仅具有工具性，尚未上升到概念性层次。换言之，其并未深入人文社科核心，不具备直接作用于学术研究且取得重大成果突破的能力。在目前的技术阶段，ChatGPT 以及类似的 AI 工具主要面临着下述问题。其一，创造力的缺失。ChatGPT 是基于对现有数据内容的学习提供对策与建议，其本身并不具备创造性。虽然有创新意识的用户会通过对话不断向其语料库中增添内容，并对其进行训练，但是这种单一化、非组织化的内容增加方式，其知识增量、创新速度相比已存在的海量数据来说，显然是沧海一粟。怎样运用有限的知识增量改变现有的知识体量，怎样使得有限的知识增量创新且准确进入所需要的内容生产，则是研究者在运用 ChatGPT 时需要重点关注的。其二，想象力的匮乏。由于 ChatGPT 使用联合概率，其唯一的想象就是联系和组合，难以产生创新性的观点。以社会学为例，如果要遵循米尔斯等社会学大家的指引，发挥“社会学的想象力”，那么依旧需要学者基于自身经验来探索前人所未及的领域，ChatGPT 只能起到知识辅助作用，并不能对核心论点进行决断。

除了创新、创造领域的先天不足，ChatGPT 在辅助学术的具体过程中还存在其他一些问题。例如，评价事物时的模糊态度、回答结果的非科学性、决策过程的不透明性、观点输出的潜在偏见、复杂分析的技术限制、高度依赖用户的提问方式和交互等。概而言之，人文社科领域在使用 ChatGPT 时需要充分意识到，ChatGPT 之于知识生产是整合而非创造的，之于信息

检索是有偏差而非精准的，之于数据分析是借力而非依赖的，之于复杂推导是辅助而非主导的，之于学术研究是运用而非依从的。

ChatGPT 对意识形态的潜在治理挑战

ChatGPT 会在对话中潜移默化地影响人类的思想观念，而其习得语言、获得信息的差异，极易造成输出知识的差异，进而导致意识形态的偏颇。在学术研究的具体过程中，ChatGPT 的缺陷只是现有技术的不足，而在意识形态的导向方面，ChatGPT 则构成了一定的方向性挑战。对于人文社科学者来说，虽然研究过程并不带有意识形态的偏向性与能动性，但与自然科学不同，其研究问题并没有精确的答案，特别是各种理论分析工具之间的复杂关系极易受意识形态的强烈影响。一旦人文社科领域以 ChatGPT 为主要学术辅助工具，或者其所使用的数据、所关注的社会群体、所探索的文化现象受到 AI 工具有偏观念的冲击，那么势必会造成研究结果的扭曲。由于 ChatGPT 的学习对象是全球互联网数据库中的语料，而全球话语场域早已形成多种意识形态交叠的竞争之势，在不久的将来，一些语料经过 ChatGPT 训练过程中的学习、模仿和放大，话语的竞争场域必定会升级成为知识的竞争场域。正如培根所言，“知识就是力量”，谁垄断了文本，谁就能够凭借知识的力量来影响、塑造世界。新型的“技术精英主义”和“知识精英主义”会在对特定 AI 工具的垄断过程中形成。这一问题需要学术界和思想界予以特别重视。

知识和技术本身并不存在善恶之分，但利益群体的蓄意操纵，会使得 ChatGPT 学习、输出的知识体系沦为权力的工具。ChatGPT 模型使用基于人类反馈的强化学习算法，能够在与人类的对话和互动中更加自主地接受人类给予的现实反馈，并即时学习调整相关的回答信息。因此，它的回答有可能被人为改变或控制。由于 ChatGPT 本身不具备鉴别能力，它同样会学习和模仿某一时间段互联网极端情绪和错误观点堆叠，并不反映具体社会事实的结果，并将其学习到的内容总结、归纳给提问者。

在错综复杂的互联网空间中，政治、文化权力的角逐并不罕见。中华文化主体性和非中华文化主体性之间的博弈，以及长期存在的有关价值观来源、主导权的争端都会成为 ChatGPT 的学习内容，于是在新的技术领域产生新的角逐。不同语言的语料对 ChatGPT 会产生不同的训练结果，我们尤其需要重视不同语料所形成的话语偏差。针对 ChatGPT 对意识形态治理所构成的挑战，有专家提出组建包括人文社科学者和技术人员在内的专业团队。这一应对手段强调发挥专业团队力量的优势，在国内互联网场域提供更高质量的中文数据来实现知识的引导。专业团队可通过对模型的训练，纠正 ChatGPT 对知识议题的错误回答，以减少潜在风险。同时，在全球互联网场域中，专业团队也可以对 ChatGPT 进行多语言的集中训练和语料的重点补充，对错误观念出现的源头进行“正本清源”。

当然，AI 技术是一把双刃剑。当我们认为 ChatGPT 的某些回答是错误的且需要被纠正的同时，也给予了它一个很强的前提假设，即随着时间的变化和 AI 决策能力的增强，并不排除产生新的错误和偏颇的可能。当人类与机器产生分歧时，应该由谁来定义对和错，又应该如何对其生成的内容进行恰当的引导，是知识界，特别是人文社科研究者需要重点关注和理解的问题。这正是人文社科学者审慎对待 AI 技术的意义所在。需要意识到，ChatGPT 的价值与挑战并存，应基于历史的经验，做好应对未来挑战的准备。

[本文系国家社科基金重大项目“大数据驱动的网络社会心态发展规律与引导策略研究”(19ZDA149) 阶段性成果。]

关于 AI 的生命伦理反思是丰富而有效的。例如，计算机技术改变了人们对知识、思维和决策的认识方式，将人类思维过程中的判断转化为计算过程，可能导致人类丧失直觉和判断力，甚至对人类思维和

① Z. B. Wei, "Overview of Artificial Intelligence and Machine Learning," *Acta Electronica Sinica*, no.4, 2017, pp.9-17.
②《关于 AGI 与 ChatGPT, Stuart Russell 与朱松纯这么看》,“机器之心”公众号, <https://mp.weixin.qq.com/s/TL1TCfQMetQh2nM3XykrBg>, 2023 年 2 月 27 日。



价值观造成破坏。^①这类担忧正逐步为技术所证实。

今天, AI 的时代已经开始了: 那些被人类以全部知识、科技资源哺育的 AI 已经缓缓觉醒, 看起来它将势不可挡地发展其强大的能力。而从理论上讲, 非人的 AI 智力主体使人感到生存危机, 需要具备三个前提: 一是 AI 的智力不亚于人类, 甚至更强; 二是 AI 表现出自主生长性, 甚至可能产生意识; 三是基于数字交往革命, 人类与 AI 之间建立起广泛且日常化的生存联结。迄今, 这三个前提都已经出现了。其中, 第三个或许是更为基础的要素: 只有在普遍的数字联结中, 人与非人、人与拟人之间, 才可能呈现齐一性。以我的术语而言, 这种齐一性即他(它)们都被视为“交往端”或“交往者”。实际上, 也正是在交往效用的意义上, 数字生命、机器生命才被等同于人的生命。

AI 永不眠: 交往革命与“赛博格阶梯”

杜骏飞, 南京大学新闻传播学院教授

数字交往 2.0

今天我们讨论 ChatGPT, 但讨论的意义不止于今天。如果我们要将那种未来世界中的高仿机器人(highly humanoid robot, HHR) 视为与人类生命并驾齐驱的拟人, 那么, 嵌入其中的 AI 必不只是广受追捧的聊天应用, 也不只是万众惊羡的 AIGC (人工智能生成内容), 而会是 AGI (通用人工智能)。纯粹从技术上看, ChatGPT 又未必是 AGI 的未来, 一如朱松纯所说: 虽然大型语言模型已经取得了一些惊人的进展, 但如果我们对照 AGI 的实现标准(无限任务、自主性、价值系统), 它的方向其实与 AGI 并不完全一致。^②

即使如此, ChatGPT 也足以宣告数字交往 2.0 时代的开启。如果说, 数字交往 1.0 时代的特征是人、人与非人之间的数字化交往, 那么, 数字交往 2.0 时代的特征, 则是人与人、人与非人之间的 AI 辅助交往(AI aided association, AAA)。其本质的变革是发生于人的生命与人工生命之间的联结——你也可以说, 这种联结是一种跨生命的交往。这种交往具有复杂的结构类型。(1) 跨生命交往的基本模式是 H-A, 即对话在人(H)与人工生命(AL)之间发生。后者可以是程序、模型或智能机器, 当下以 ChatGPT 为代表。(2) 在 ChatGPT 内部, 内容生成训练源自大规模的互联网文本数据集, 换言之, AIGC 来自“所有人”, 生成方式是人机接力和概率性输出, 故而其 H-A 模式的本质是 H-(AH), 这是一种人对混合生命的交往。(3) 在人的一端, 我们可以将 AI 辅助交往具体定义为 AI 支持下的社会交往, 其广泛应用于虚实融合的数字社会。这无疑将使人自身产生两种变革——AI 辅助人和 AI 增强人, 并形成混合生命对混合生命的交往(HA)-(AH)。在这一模式里, 人工生命才是真正的转译者和交往平台。自然, 从传播学的角度, 我们也可以说: AI 将成为超级媒介。

以上讨论大致展现了，在一个 AI 崛起的时代，自然生命与人工生命之间的“跨生命交往”如何可能，而更重要的是，这种交往革命也在重新定义生命本体。

多项式生命

如今，人类已经在机器身上部分实现了拟人化，人自身的工程化也在快速发展。我曾经从类型学的视角揭示过，技术对人的支持力和支配力可能会达到何种程度，而人的自我迷失又会导致何种后果。在 AI 兴起的时代，数字化的人类包括智能辅助人、生理增强人以及非整数维度的数字人。数字化人类的三种发展方向是虚拟化、智能化和外脑化。^①但是，这些工程化的人类前景或许都不是什么美好的理论叙事。其中，“虚拟化”的部分并不富有诗意，它可能会让人贪图虚拟的意义，而渐渐脱离实在本身；“智能化”的部分对我们自身的大脑并没有改进作用，却使我们的生存再也无法离开外部智力；“外脑化”的部分，则对增强人脑格外殷勤，但其反向技术会以指令介入人脑，发展为颠覆人类本质的赛博格工程。

作为一种历时性的过程，数字化人类以及人的全面赛博格化仍处在最初形态。不过，一切都已开始展现出表征。AI 辅助的兴起使人类达成了交往的赛博格化。依照上文关于人的危机的三个前提的讨论，那种平台级的演化条件业已具备。ChatGPT 的重要进展之一是终能与数字生命、机器生命流畅对话，这是一种前所未有的交往形态。这个世界对生命的理解正指向一种“多项式生命本体”，它本质上是 AI 觉醒与数字交往革命的后果。

让我们来类比一下。(1) 在数学中，几个单项式的和称为“多项式”(polynomial)。人类作为生命之一种，也将与非人并存，与拟人并存，与不同的自己并存，甚至还与智能复活的死者并存。(2) 多项式中的每个单项式称为项。我们也将成为未来生命体的一项，还是不那么聪明、不那么敏捷和不那么坚强的一群。(3) 多项式是指由变量、系数以及它们之间的运算得到的表达式。多项式生命本体也将通过演化的算力，重塑

人与“非人”，而假如这算力是来自“非人”，请不要惊讶——因为的确没有什么人比 AI 更适合在一场深谋远虑的演化级的超算中为多项式生命本体中的每个“项”(例如人、机器人、数字程序)分配权重了。

这是理论的演绎，但也会是实实在在的数字生活本身。也许，最关键的生命事实已经奠定：高智能(如 AI 与 AI 高仿人)与低智能(如真人)的生命即将并存。那么，现在唯一的问题是：人类准备好了吗？

赛博格阶梯

交往的赛博格化，还不足以引发生命意义上的戒惧，但如果生命本体的定义开始动摇，则会产生前所未有的意识演化危机。具体来说，人在发达的 AI 面前，出于对被僭越、被代替、被否定、被遗弃的四重恐惧，很可能会主动寻求被 AI 所演化。很多人惊恐于机器人将在大多数劳动中取代人，AI 将取消越来越多的人类智力劳动，但这不算是 AI 的坏消息。真正的坏消息是，我们即将背负起不能承受之重：AI 将止步于何种高度？它是否以及将如何支配人类总体？毫无疑问，有一天我们会觐见高不可攀的 AI 生命。即使在善意的童话版本里，作为智能枢纽的 AGI 母体也将向我们显示：实际上，没有任何主体比它更适合“科学治理”这个跨生命世界。在那之前，它会先证明：没有任何主体比它更适合优化、规划、演化 AI 生命。

也许，数字交往 2.0、多项式生命，都还是这场演化的初始阶段，人类物种在精神上的赛博格化，已经有迹可循。以我之见，在距离 GPT 不太久远

① 杜骏飞：《何以为人？——AI 兴起与数字化人类》，《南京社会科学》2023 年第 3 期。

的时间线上,矗立着一座“赛博格阶梯”,我把它阐释为人类赛博格化的四部曲:第一步,交往的赛博格化,其表征是AI辅助交往和数字交往2.0;第二步,主体的赛博格化,其表征是知识嵌入的多项式生命;第三步,认知的赛博格化,其表征是生理嵌入的AI增强人;第四步,意识的赛博格化,其表征是人开始像AI那样思考,并期待拥有机器的心灵。

我们已经讨论了交往的赛博格化和主体的赛博格化,那就不妨再来看看“赛博格阶梯”的第三级。以AI增强人为例,我们或能预感到人的赛博格化是如何从认知跃升到意识层级的。既往,对赛博格的生命伦理学反思主要聚焦在肢体工程。然而,真正的危机在人脑,“外脑化”的赛博格工程,其医学初衷是恢复人被损伤的知觉和运动能力,而非增强普通人,但这并不意味着外脑化增强人工程不会普及一切人,更无法保证它不被居心叵测的权力或资本所利用。实际上,通过双向脑机接口支配人的感知或思想,早已被证明是可能的。^①

设想AI所支配的脑机交往场景:日久天长,交融渐深,总有一天,增强人会在AI或AI掌控者的保护下,心甘情愿地将“系统”当作神。当这一幕出现时,AI才真正展现出那深不可测的未来,以及人权被物权所役的可能性。赛博格化的演化从交往开始,发展到对生命本体的再确定,再发展到认知层面的博弈,最终,人的意识要么皈依机器之心,要么为技术的外部性所奴役。

AI 永不眠

AI已处在无尽的演化之中,而我们自身的发展却可能早已停滞。一

旦我们意识到这一点,就将回到那个终极问题:当人类及其子孙在那新的创世神话中,终于像超人类主义(transhumanism, H+)宣称的那样成为硅基生命演化的阶梯时,会渴望返回到2023年甚至更早的AI史前时期,以警惕和防范AI科学主义吗?

如前所述,科技的意义在于提升生产力,所有人都乐见技术进步。只是,依我之见,对不可见的未来,只能名之为“未托邦”(vuca-topia)——一种具有VUCA(波动、不确定、复杂、模糊)属性的哲学预期。^②只要VUCA尚未澄清,我们就要对“人创生AI”这件事谨慎小心:对科学道德保持怀疑,对技术加速保持批评。

写到这里,作为一个同情“可知论怀疑主义”的人文学者,我想评述一下创造了“不可知论”(agnosticism)这个词的赫胥黎的话。1860年9月23日,赫胥黎在给友人的一封信中说:“我既不主张也不否认人的永生性。我看不到相信它的理由,但另一方面,我也没有证明它的方法。”^③类似地,我也想说:“我既不主张也不否认GPT的伟大。我看不到崇拜它的理由,但另一方面,我也没有阻止它的能力。”尽管我对数字的未来也没有多少结论,但我确有自己的信念,如同情人,同情一切生物生命,秉持人类的基本价值观,秉持人性,如此等等。人类对“常识”的倾向,使我对计算逻辑能否实现人本身抱有怀疑。相较于那种让先进代码和发达人工智能支配人的科学主义原则,我还是宁可相信天真的“人文”——它给人以好奇心、良知和勇气,去追问那些遥远的、沉重的、未知的事情,即使未来杳无音讯。并且,我认为,深切的关怀和深邃的自省并不会妨碍我们在现实中履行义务。相反,对全人类的责任能让人懂得:为什么我们要“对周围的人温柔、帮助和公正”,以及为什么我们要迎着满天星光,眺望千百年后那些陌生的人生。

我相信,这些信念是“更高力量”(higher power)的一部分。

[本文系阐释党的十九届四中全会精神国家社科基金重点项目“网络社会治理共同体:理论、实践与范式研究”(20AZD062)阶段性成果。]

① R. Douglas Fields, “Can Machines Control Our Brains?” *Quanta Magazine*, <https://www.quantamagazine.org/can-machines-control-our-brains-20210517>.

② 杜骏飞:《“未托邦”:元宇宙与Web3的思想笔记》,《新闻大学》2022年第6期。

③ Thomas Henry Huxley, “Agnosticism,” *The Nineteenth Century*, vol.26, no.151, 1889, pp.66-777.

对话与重写：ChatGPT 时代的文学

严锋，复旦大学中文系教授

ChatGPT 技术的出现，使得文学的生死存亡问题又冒了出来。为什么说“又”呢？因为“文学已死”的说法并非首次出现，20 世纪以来，这个话题已经出现了很多次。然而，ChatGPT 技术的出现，似乎将这一话题推向了一个全新的阶段。与这一话题相关的，不仅仅是文学领域，还包括其他学科和职业，以及整个人类文明的走向，它甚至可能影响我们的生死存亡。

用 AI 的眼睛看世界

伴随电影、电视、电脑、电子游戏的出现，一个泛写作和泛文学的时代来临。每当一种新媒体出现，创作文学和研究文学的人都会经受一次被取代的危机感，但实际上文学没有“死”，有时候还迎来了一些新的发展机遇，这是非常有意思的话题。当然这也印证了麦克卢汉的说法——旧媒体不是死去，而是在不断转型，而且与新媒体共存。

现在还远没有到判断 ChatGPT 是不是真的能够取代人类，抑或接管文学，担当作者功能的时候。但至少有一点是可以肯定的：正如其他新媒介的出现那样，新技术的出现能够让我们去思考和重新认识传统的文学、艺术，乃至思想和文化。有些新东西出来之后，我们站在它的角度，用它的眼光来看，一些传统的东西可能会焕发新的光芒。托尔斯泰是一位传统的文学家，但他对技术的敏感程度却颇高。1908 年 8 月，《纽约时报》的团队为庆祝他的 80 岁生日而拍摄了一段纪录片。在片中，托尔斯泰指着当时刚刚出现的摄影机表示，这个带转轴的咔哒咔哒作响的装置将会给我们的生活带来一场革命，特别是对作家而言。事实上，托尔斯泰的预见确实被后来的发展所印证了。然而，技术对文学艺术的冲击，并非简单的你死我活的问题，也不是新旧替换的问题。我们现在知道，实际情况比托尔斯泰当时所能预见的要复杂得多。技术

的进步，不仅会对文学创作带来巨大的影响，同时也会对其他领域产生深远的影响。

我对元宇宙、大型语言生成模型等都是这样的角度来看的，这能够帮助我们从更广泛的联系中去看 AI 与艺术、



人文关系的趋势。此前，“文学已死”一说仅为修辞手法，但现实却似乎证明了它的实际意义。近日，国际知名科幻杂志《克拉克科幻世界》宣布暂停短篇小说的征稿，原因是大量 ChatGPT 创作的科幻小说涌入投稿系统。国内的《科幻世界》也发出了类似的警示，亚马逊平台也有署名 ChatGPT 的电子书上线。AI 已经开始占据传统作家的位置，对文学创作生态产生破坏，给文学创作带来前所未有的冲击。

此技术的出现必将引发众多新的问题，首当其冲的是原创与复制的界定难题。该问题不仅在文学创作领域产生影响，同时在学术研究领域也引起了广泛的担忧。全球版权问题因此变得异常复杂，语言学家乔姆斯基认为，ChatGPT 实际上是一种高科技“剽窃”行为，这一说法值得深入探讨。正是由此，我们需要重新审视何为剽窃，何为创新，因为这些在传统时代内涵相对清晰的概念



如今正变得模糊。ChatGPT 的出现为我们提供了新的视角来探讨这些问题，但同时我们也需要应对由此带来的挑战和困境。

技术会让我们获得一种面对传统人文学科的新眼光。然而，从文学角度来看，这个问题可能更为突出，因为文学创作一直以来都追求原创性。人们视作品为作者的自由创作，如陆游所言“文章本天成，妙手偶得之”，或者如郭沫若所说，诗歌就是心琴上弹出来的旋律，鲁迅也曾多次表示不信赖《小说作法》这一类的书籍。这里都可以看到一种对创作主体的确信，而且这种创作主体被认为是独一无二的。

ChatGPT 的出现令神圣而自主的传统创作主体面临着危机。然而，创作主体的问题并非今天才出现，它一直都存在。许多经典文学作品的原创性和作者身份，若仔细追踪，往往会被揭示为不可靠的幻象。这些作品一直以来都是互文存在的，像地质断层一样层层叠加，无论是中国的《西游记》《三国演义》，还是西方莎士比亚的作品，都经历了漫长的创作过程。曾经，作者拥有像上帝一样的神圣地位，但这种神话在现代遇到了挑战。从思想史的角度来看，“主体性的黄昏”是哲学思想中一个重要的转折点，传统哲学中的主体概念在现代和后现代思想中逐渐失去了其优先地位和确定性，人类作为主体的地位逐渐被质疑和挑战。“主体性的黄昏”也就是人类中心主义的黄昏，而 ChatGPT 又进一步推进了这个去中心化的过程。从结构主义开始，创作主体就不再被认为是认识和写作的中心，而是结构的一部分。结构主义认为，不是我在说话，是话在说我。在文学上，这个观念表现为，

不是我在写作品，而是文学作品在写我，它代表了前辈作品的积累。哈罗德·布鲁姆在《影响的焦虑》中认为，这是一个巨大而令人生畏的传统，后辈的作品就是在与前辈的作品对话，同时进行各种修正。这听上去是否有些像 ChatGPT 的写作方式？

三角关系下的新写作生态

目前，针对 AI 文学写作，一些人持极度悲观的态度，声称其完全颠覆了传统的写作生态，甚至有极端的观点认为，AI 将来会取代作家这个职业。我并不知道未来会如何发展，但目前来看，AI 还远远不能脱离写作者而成为一个新的创作主体。然而，人工智能与人类的合作关系将持续发展。我认为，ChatGPT 体现了一种对话的逻辑，这种逻辑更像是现代读者和作者之间的关系。在传统的读者和作者关系中，作者是上帝，而读者则难以发出自己的声音，两者之间存在一种博弈关系。而在今天，读者拥有更多的工具、手段和技术来与作者进行合作。或者说，他们一起形成了一个三角形的关系：传统的作者是一极，AI 是另一极，而读者（作为新的作者或合作者）则是第三极。

当前 ChatGPT 的输出中存在大量错误信息。然而，在文学创作中，这个问题却具有独特的意义。在文学领域，错误往往正是创作的萌发之地。ChatGPT 的文学写作是一个不断调节的过程。向它提供一个场景或一个点子，它可以马上生成一段文字，但有时，生成的文字不符合我们的期望，这时就需要调整。这其实更像是一种对 ChatGPT 进行引导的过程，旨在精益求精。这个过程具有对话性和创造性。

比如，让 ChatGPT 在《红楼梦》前 80 回的基础上，续写一个后 40 回的梗概，就会形成如下对话。

ChatGPT：贾宝玉后来跟王熙凤谈恋爱了。

我：不对，这有悖伦理，是绝对不可能的。

ChatGPT：对不起，我说错了，贾宝玉后来其实是跟袭人结婚了。

我：还是不对，袭人的身份不对，贾宝玉不可能跟袭人结婚。

ChatGPT：对不起，我又错了。贾宝玉后来是和

史湘云在一起了。

此过程看似无厘头，实则类似于传统作家创作时的头脑风暴。但现今这种“风暴”已添加了算法和模型，为我们提供无尽思路，同时又可以根据创作者的意愿持续修正。就目前而言，仅依靠 ChatGPT 写作仍不够，人类的判断力、艺术品位和受到的传统人文教育仍十分重要，甚至比以往更为关键。但它可成为人们写作的助手，作为优秀的头脑风暴工具，此毋庸置疑。ChatGPT 目前尚非刘慈欣、莎士比亚或曹雪芹，可能只是一个三流的小说家，但却能够提供多样的信息和材料。写作需要涉及具体知识，例如航海、航天、电子技术等领域，而 ChatGPT 在此方面能发挥重要作用。在传统时代，作家需前往图书馆深究，网络时代则需具备搜索技能，但所得结果仍需人类判断、整理和概括，而今天，ChatGPT 能协助我们完成提炼和概括的任务，更加细致地拆解问题，这是以往无法比拟的。ChatGPT 的出现大幅提升了我们学习和写作的效率，这是不容置疑的事实。

这一过程实际上体现了更为广泛的文化发展趋势，即读者角色日益晋升为创作者的地位。我们正处在一个读者的时代，一个千千万万读者不断涌现，并向着创作者角色转变的时代。读者反应批评、接受美学、阐释学乃至后来的游戏叙事学都对这一文化转型进行了论述。在电子游戏中，玩家不仅是传统意义上的文化消费者，更是内容的创造者和修订者。除了这些新的读者、新的创作者、新的游戏玩家，今天，这个大军中又加入了 ChatGPT 这一读者 / 写作者，为这一趋势注入了新的活力。

ChatGPT 抢了元宇宙的风头吗

当前有人觉得 ChatGPT 抢了元宇宙的风头，但我觉得情况正好相反。元宇宙正在迎来它期待已久的一块缺失的“拼图”，那就是一个高度拟人化的 NPC。比如最近备受好评的《霍格沃茨之遗》游戏，如果里面的 NPC 都由 ChatGPT 驱动，那么这将不仅仅是一款普通的游戏，而是一个真正的活生生的哈利·波特元宇宙。同样我们可以设想一个由 ChatGPT 驱动的红

楼梦游戏，那将是一个怎样的虚拟世界。

元宇宙也被称为 3D 互联网。早期的互联网是完全基于文字的，随着带宽和计算机图形技术的发展，图像、音频和视频越来越占据重要的位置，这让用户可以调动更多的感官来进行交流，获得更强的现场感。然而，现在的互联网依然是 2D 的，丰富多彩的 3D 世界被压缩在一个平面上，这是虚拟世界与现实世界的最大的屏障，人们对虚拟世界的操作和体验也因之受到极大的限制。要开拓新的网络空间，就要发展新的维度，这是元宇宙发展的内在动力。虚拟现实、混合现实和增强现实技术不断进步，新一代的人机界面卓然成形，为元宇宙的发展提供了技术支撑。要建设 3D 互联网，除了硬件技术之外，还需要大量用户生成的 3D 环境和内容，这在过去有着很高的技术门槛，也构成了制约元宇宙发展的瓶颈。ChatGPT 让普通的用户能够通过文字轻而易举地生成自己想要的 2D 和 3D 图像，乃至活色生香的动画视频，这就为下一代更高维度的互联网建构提供了快捷的通道。

从几万年人类在岩壁上的涂画开始，我们就开启了一个虚拟化的进程。我们可以将传统的现实世界视为现实宇宙，同时我们也早已拥有另一个宇宙，即虚拟宇宙——文学、绘画、戏剧、电影等。在过去，这两个宇宙是相互分离的，随着科技的不断发展，这两个宇宙之间的界限逐渐被打破，它们日益融合。人类的虚拟活动正变得越来越具有实体性和影响力。从这个角度而言，ChatGPT 代表着虚拟化进程的一个新阶段。如果这一观点成立，那么这个进程将是虚实融合、虚实互渗的，这也是我对 AI 与人类合作的最终愿景。

2022年入冬之际，ChatGPT来到世间，持续发酵，引发热议，一方面，技术明显可以造福人类，另一方面，我们对技术失控的担忧也陡然增加。我们需要从一种革新的哲学视角来判断人工智能与人类自身之关系，可



① A.M. Turing, "Computing Machinery and Intelligence," *Mind*, no.59, 1950, pp. 433-460.

② 详细论述可参看拙文《人工智能的情感计算如何可能》，《探索与争鸣》2019年第6期。

以把 ChatGPT 之类的大型语言模型视为通用人工智能的初步成就，但需要将思考与自我意识的获得彻底分开，哪怕是未来的强通用人工智能，也只是不断增强思考的功能，却无法获得人的自我意识。对人工智能“觉醒”的恐惧只是人文主义文化向后人类文化过渡的中间调适状态，只有不断批判和反省，客观地看待技术发展，我们才能摆脱对人工智能的恐惧。

新的认知机制

自然语言处理大模型的兴起让我们不得不面对一个新的认知模式：机器能够模仿有机体才具有的认知功能。在图灵机出现之前，历史上所有关于机器思考的猜想都是无法实现的，图灵方案才真正展现了机器智能的可能性，“思考”这一人类独有的能力开始找到机器模仿的道路。^①图灵提出的机器智能的观念

ChatGPT：更新对机器智能的认知

王峰，华东师范大学传播学院、中文系联聘教授，
教育部青年长江学者

不仅是一种数学或计算机意义上的概念，而且是一种新的认识论。此前是人来认识世界，现在是机器按照人制定的认识规律与世界和人类发生关联。机器智能对人类能力的模仿看起来是在模仿人脑结构，但事实上，人脑结构本身是不可模仿的，毕竟一个是生物结构，一个是机器结构，从基本构造而言，这是两种完全不同的结构，但从最终的输出效果来说，机器智能确实可以部分地实现人脑的思考功能。

我们必须面对一个新的认知机制，它不同于此前哲学的认知观念，它要求我们将机器智能纳入新的认知观念中。换句话说，我们要把人脑中的整体性认知系统转变为可区隔处理的功能系统，且通过计算达成这些功能。^②在此之前，我们曾认为，哪怕是区隔的功能也是无法计算的，因为它包含了自然界最完美的神秘构造，无法被制造，甚至无法被言明。在通往人工智能尤其是通用人工智能的道路上，我们发现机器智能可以与世界、与人类发生关联。从本质上说，机器智能更接近于可区隔的智能横截面，但恰恰是这些区隔的、有限的智能，完成了人脑所执行的某些思考功能，这当然是对既有认知的巨大冲击，但也只有接受这一冲击才能更好地认识人工智能的本质以及它会给生活带来的变化。从这个层面来说，图灵不仅是数学家，更是 21 世纪的哲学家。我们只有在哲学整体观念上将计算复杂性与思考画上等号，才可能更新哲学观念，接纳机器智能带来的新变化，从而对当下正在经历的大语言模型造成的冲击保持健康和稳定的心态。

一般来说，我们会将语言预设为内心的表达，但是 ChatGPT 却赋予我们一种新的视角：语言自动运行。这无疑为我们反思“语言与我们的关系”提供了难得的机会。在 ChatGPT 的实际运用场景中，语言的运行具有无限可能，它们可以进行自由组合，并在大模型的控制下产生有意义的结果，这一结果与现实世界之

间呈现出一种索引式的关系，也就是说，词语与现实的关联可能是脱节的，也有可能是直接对应的。从目前的情况来看，沿着这种场景寻找语词向现实世界的延伸之路也是最艰难的地方。

非意识：机器思考的功能化

通常情况下，我们会把“思考”和“意识”紧密联系在一起，在人类智能范围内的确如此，在人工智能这一新的认识功能载体中却并非这样。图灵认为机器能够思考，并且给出了判断机器思考的标准——通过图灵测试。也许我们会从此推论，图灵同样也指出了通用人工智能重建自我意识的可能，但这并不是一个正确的推导，将人类确定无疑的特性原封不动地转嫁给人工智能显然是不合适的。

功能化是人的能力转变为人工智能的关键。起初，我们不认为“思考”能够功能化，但是通过图灵测试，我们认识到“思考”并不是有机体独有的，它可以被部分地功能化，并且让计算机来执行它。那么，意识能否功能化呢？我们当然可以让机器看起来像是在思考，并且它实际上也会产出思考的结果，哪怕产出方式不同。比如 ChatGPT 使用的对话方式与人的对话在产出结果上相近，虽然产生机制完全不同，但我们依然可以说，ChatGPT 初具通用人工智能的水平，可以通过图灵测试，也就是可以思考。然而，我们却不能说它具有自我意识，因为其所有的“自我意识”其实不过是计算复杂性的表征。在计算复杂性中涌现的是人类的某种功能，（有限度的）思考是一种功能，所以思考可以涌现，但意识无法局部化，也就不可能完成涌现。从根本上来讲，只有人类才能将意识与思考结合为一体，而人工智能却无法做到，无论计算复杂性达到何种高度，全面复现人是一个难以企及的浩大工程，即使在遥远的未来能够达成，这也是一件没有多大意义的事情，除非人类濒于灭亡或已经灭亡，否则没有必要实现。如果我们将涌现的功能等同于整体性的意识，并将其归属于人工智能自身，将之视为人工智能内部的事情，那么，我们实际上是将意识用错了位置，这是一种概念错置而产生的误用。当那些放在

人类身上毫无疑问的概念转运到人工智能领域时，必须小心翼翼地为之进行概念反省。在人工智能这里，我们只能说“它能思考”，且仅此而已。人工智能虽然具有智能功能，但永远不会具有灵魂、心灵、意识、自我意识等内在天然禀赋。即便我们寻找到一些意识的“痕迹”，那也是使用人工智能的人类赋予它的。在与 ChatGPT 的对话过程中，它会使用人类经常使用的情绪词来回复我们，比如“这个话题很沉重”“不喜欢这个话题”等，仿佛人工智能具有内在的感情，具有内在意识，但是我们必须明了，“沉重”“不喜欢”之类的情绪词对于 ChatGPT 而言只是词语的组合，组合的意义来自人类对语句的使用，与 ChatGPT 内在情绪没有任何关系，它只是在适合使用“沉重”“不喜欢”之类词语的语境中，给出了相应组合语句而已。ChatGPT 这样的对话人工智能，始终和人类有着本质区别。它的输出看起来像人的对话一样完美，但它本身是没有感觉、情感的，它不会真的“感受”到沉重。这些感受性的描述，是人类在沟通时使用的。从此而论，这是一种真正的模仿关系，无情绪的模仿。认识到这一点，我们就会发现，一个技术的发展既不会放出魔鬼，也不会带来天使，我们的生活不会变得更糟糕或更幸福，只是更方便而已。

因此，我们不应当去预设人工智能产生自我意识，实际上也不可能出现人工智能战胜、取代或奴役人类的情况。任何一种大语言模型人工智能，归根结底都只不过是人类的辅助工具，它可以思考，更重要的是，它可以帮助人类思考，但它不可能产生意识，也不会存在自我观念。

分布式智能：人机融合的新智能体

人工智能的发展提示了一种新智能形态的诞生——分布式智能。分布式智能指的是，以人为接受者和服务对象，即以人的智能为主、人工智能为辅的智能存在方式。电脑、手机、无线通信、网络协议以及大语言模型、自动机器等都是智能设备的有机构成部分，与人相结合，就能大幅扩展人的世界视野、思考方式，提升人的知识水平、信息获取能力等。夯实智能设备基础，就能够达成人与机器智能相融合的新智能体，即分布式智能。

分布式智能可以超越人脑处理速度、容量等先天局限，人不需要意识到自己在动脑做事情，相关辅助智能就可以完成智能任务。新的智能体显然不同于单纯的人类智能，人类智能处理任何一个事物都经过智能判断，人哪怕过段时间忘记某一判断或做过的某个事情，原则上仍可以通过记忆回想起来，但在融合智能中，处理过的事情是由辅助智能体根据指令完成的，也可以通过查阅数据记录来精准定位事件。这样一来，人脑所承担的信息处理工作就可以通过融合智能以分布的方式来完成。5G/6G的发射站、网络交换的框架协议、手机、手表、电脑等都是处理事务的智能设施，单独来看，它们并不具备我们所熟悉的有机智能形态，但结合起来却能够完成很多智能任务，并且留下可追踪的数据。这已经是不断显露端倪的分布式智能了。ChatGPT之类的大语言模型必然会加速分布式智能的进程。

在人工智能兴起的进程中，我们经常听到一种声音——人工智能时代，至少在大语言模型兴起的时代，人工智

能会取代人类的大量职位。的确，人类既有工作职位的消失必然会发生，新技术会消除某些工作职位，但却创造了更多与之相关的工作职位，这一点已经通过互联网的发展证明了。对于我们而言，终身学习将会成为基本要求，丧失终身学习动力和能力的人将失去高质量的工作机会，这当然也给我们的教育提出了新的培养目标，即终身学习的热情和能力。此前，一个人一生可能都遇不到工作职位消失的问题，而未来，以40年工作时间来计算，一个人一生会遇到两次以上的工作职位消失的可能。这必然提高终身学习的必要性。

后人类：人工智能与人的未来

随着技术不断向前发展，我们也正在走向“后人类”状况，这是比较激进的状态，也是目前国际学术界热议的内容。在后人类状况下，人本身是可以改变的，可以跟人工智能相结合。在远程通信技术的支持下，地域、距离不再是需要考虑的因素，我们可以即时知道地球上任何一个角落的事情，就像人人都拥有了“千里眼”和“顺风耳”。随着大语言模型的飞速发展，我们将逐渐进入真正的后人类状态：从依靠记忆来提升技能过渡到通过提问来掌握技能和知识。记忆同样重要，只是在知识学习上转变了方式，人的智能更多用来扩大视野，进行创新性思考。人类自身也不断机械化或接入智能设备以获得质的提升，甚至达成某种程度的进化。从这个意义上来说，我们都是赛博格化的“后人类”。

赛博格化是人类的一种升级状态。从本质上来说，人工智能只是人的某种功能的局部飞跃，距离具备自我意识非常遥远，更重要的是，哪怕到达路途的另一端，那时人类的观念一定已经发生了巨大的改变。不断向前进展的不仅是技术，也包括人类自己。事实上，技术的每一次迅猛发展都会带来人类能力的提升和对自身的重新审视。我们总会适应新时代的到来，在新时代、新技术的推动下不断改变人类自身的生存方式，成为某种意义上的新人。

警惕“机器学习”僭越“人的教育”

刘云杉，北京大学教育学院教授

机器学习：“复合的教师”还是“超级学习者”

与历史学纵横千年的时间观和哲学高居山顶的宇宙观不同，教育学的视角贴近地面上平凡的芸芸众生，他们希冀什么，又承受什么？被什么所迷惑，又被什么所刺痛？面对 ChatGPT 等人工智能，我的关注既非远忧，也非近虑，而是一个现实的困境：在机器学习的凯旋中，如何给予受教育者必要的保护与尊严？

ChatGPT 对于教育究竟意味着什么？技术之于教育，不再仅仅作为教学辅助物，也不仅止于智能的教育者，它更是一个升级的、可以替代人类的新型学习者。ChatGPT 不仅是能“为你所用”的高效、标准的“助教”，也是能“替你之用”的、深谙算法的学习者。它既是“类人”的学习者，更是“超人”的竞争者。它对人类发出了哪些挑战？如果不加限制，当下大多数的常规工作、程序型工作，完全可能被高效的智能机器替代：一般高校的人文社科专业毕业生将遭遇 ChatGPT 等 AI 的“霸凌”。高等教育大众化后出现的白领无产阶级化现象将进一步加剧，中产阶级将有可能大面积“塌陷”。

ChatGPT 不过是算法学习的优等生。我聚焦的问题是：算法学习如何从“类人”转变为“驯服人”，哪些被视为当然的教育理念为其鸣锣开道？教育改革复杂的多重逻辑是如何表达的？它们既直白强势，又伪饰隐匿。教育中“人的意象”发生了从不易察觉到令人错愕的蜕变：从品格到人格，再到核心素养，再到能力、潜能、通用能力、高阶能力等，眼下是与算法最为契合的优绩。人的教育强调的是塑造品格，即“有所不为”的内在坚定性，如今却已蜕变为强调应对一切短期主义、既脱嵌于具体情境又可以快速嵌入任何情境的优绩者——算法学习愚笨的人类模仿者。人的学习是如何指标化且空洞化，准确地说算法化的？

人的教育是如何被扭曲为一条悬浮的管道，与一切具体的经验、日常的情境脱嵌的？算法学习者被输入其中，既被各种指标时时监控，又对照算法自我训练，最终算法内置于其脑中，在所谓的“奇点时刻”，“内置芯片的算法人”成批出



炉。这一意象不是未来的乌托邦，它正发生在我们身边，构成你我的日常，有的正在进行，有的已经完成。

学习政治：解放中的强制

在过去的 20 年中，“学习”成为流行热词，在教育研究、政策与实践，关于教育的大部分话题已经改用学习的语言来言说，即教育被简化为“教学和学习”，把学生、儿童和成人称为“学习者”，把学校变为“学习环境”或“学习场所”，教师被视为“学习的辅助者”，教学被视为学习辅助技术，甚至连“成人教育”也转变为“终身学习”。英国学者格特·比斯塔（Gert Biesta）将其称为“教育的学习化”（learnification）。

首先，这种学习化体现为“以学习者为中心”的教育理念，它由学习者的兴趣、学习者的选择和学习者的权利等一系列主张构成，表现为以下三重逻辑：



第一重逻辑为解放逻辑，主张将学习者从“教与学”的束缚中松绑，从教育关系之中解放出来，“自主的学习者”的意象被建构出来，只要信任、鼓励学习者，解放其捆绑的手脚，学习好像自然而成；第二重逻辑是选择话语及其背后的教育消费逻辑，教育话语与经济话语混淆，学生的自由选择中隐藏着陷阱；第三重逻辑是将教育视为服务业，其背后是市场与监管逻辑。在西方社会，随着新自由主义的兴起，福利国家衰退，国家与公民之间的关系开始去政治化，或者说政治本身已经被经济化，经济逻辑重构了政治功能，政府被定位为供应者，公民成为消费者，选择成为关键词。

①② 格特·比斯塔：《重新发现教学》，赵康译，北京：北京师范大学出版社，2021年，第54页，第100页。
③ 巴兹尔·伯恩斯坦：《教育、象征控制与认同》，王瑞贤译，台北：学富出版社，2005年，第81页。

看似自由的个人主义话语背后，学生不仅是选择的主体，也是权利的主体，更是责任的主体。在自由的学习话语中，学习如同呼吸一样，既是自然发生，也是必然发生的，同时还是必须发生的，不得不发生的——解放话语只是表层的修辞，其实质是强制的逻辑。在全球资本主义中，经济需要一种适应能力（adaptation）：我们必须学习，强制性学习是一种不可逃避的义务，而且是终身学习。比斯塔将其称为“学习的政治”。^①

在学习政治中，学习是自我负责的。学习政治的关键是把政治难题、经济困境转化为学习困境，因此把对付这些难题的责任从国家和集体层面转移到个体层面：个体在不断更新变化的全球市场中要为自己的就业力负责。学习指向个体的适应能力，而不是一个有关结构性和集体的责任。当学习被宣称“财富蕴藏在其中”时，实则是将个人深陷于压力，或许是不可承受的压力之中。^②

如此，政治的职能转化为经济的契约，经济的逻辑又将责任置于个体的学习能力与终身学习的适应力上，胜任力要求持续的能力更新，能力又来自持续努力的、永不疲倦的学习主体。这既赋予学习者超强的学习能力预设，又赋予被掏空了教育的学习——甚至是反教育性学习以不可能完成的任务，如同一个挟山超海的大力神，用“自主”“独立”“不断更新的能力”“不知倦怠的积极主体”遮蔽了政治与经济的各种结构性与制度性的困境。解放了的主体能胜任吗？无所不能的学习能兜底吗？资本加技术真能获得化解困境的魔法吗？

算法学习：开放中的封闭

在通天巴别塔的重建中，如何让不同语言的人能够说同一种话？这种语言一定是去情境、去文化、去差异的高度抽象的符号，当下的算法可谓正是这一语言。它如何潜入学习之中？

首先是能力话语的兴起。我们都熟悉这句话：“用过去的知识教今天的孩子面对不确定的未来”，随之而来的是知识和能力的对立。具体知识和具体学科被贬低，取而代之的是一套可迁移、通用能力话语。这体现为学习内容的重构：从关注学生被赋权学习什么转向关注他们应该练习哪些技能，课程与教学中从强调“knowing-that”转变为强调“knowing-how”，从基于学科的、概念性的知识传授转变为基于项目和任务的学习，从基于命题式的学习转变为注重实际操作的学习。

在高等教育领域，这种趋势体现为跨越学科边界和知识的稀薄化。传统的学科边界被突破，单一性知识转变为领域性的知识。单一性学科（singulars）是学科内在的，它有独特的名字，有一套专门的独立话语、一个自己的智识领域，它是自我中心的、指向自身发展的，其受到强边界和强等级的保护。^③领域性的学科（regions）体现在知识上的策略就是“跨界”与“跨学科”，表现为知识内部的联合与重组，学科与学科之间、知识与其应用之间从隔离到联结。领域性的知识既服务于市场，又更依赖市场，其学科防御能力

变得越来越脆弱，其后是科技的发展、行政的问责以及市场的调控。从单一学科向领域性学科的转变，意味着学习者内向投射的认同转为外向投射的认同。^①其关注点从专业身份转向工作能力，即从“我的专业是什么”转向了“我能做什么”。

更进一步，学习所强调的目标更为抽象与上移：一般性学科或一般性类型（generic modes）。“一般能力建构于和分布于教育再脉络化场域之外，且独立于教育再脉络化场域”，即独立于学校具体的教与学之外，它们由可迁移能力、一般化能力和核心素养等具有家族相似性的话语构成，具体指关键与核心技能、思维能力、问题解决、团队协作等。这些话语被认为适用于所有学科和实践领域，但又独立于它们之外。

伯恩斯坦清醒地指出“一般性模式”的陷阱：

一般性类型的产生是针对一项技能、工作、实务或是工作范围的表现所需基本特性之功能分析而来，这些重要而明确的必要特性指的是“共同能力”。若将这些能力重要潜在特性拿走，让技能、工作、实务和工作范围的文化基础禁（静）默无声，进而产生一种枯燥乏味的可训练性（trainability）概念。^②

能力的重要潜在特性被削弱，可迁移的能力转变为“可训练性”；思维能力、学会学习与学会合作这样的软技能，有无“重要的潜在的特性”做支撑，与单一学科和领域性学科之间究竟是什么关系？一般化的“技能话语”（skills-talk）被诟病脱嵌于具体的知识情境，即将“知道如何”（know-how）去情境化，也即抽象为高阶的方法，它分裂了学习内容，削弱了连贯性，并有效地掩饰了到底要学什么。^③具体的学科内容是概念的载体，人们经由内容，通向概念，进而锻炼思考的能力。“知道如何”只是补充并依赖于“知道什么”（know-that）与概念性知识，而绝不是“替代”它们。

在“可训练性”的能力中，学科知识、专业的方法与元学习能力之间的复杂关系被简单割裂，知识认知化，认知工具化，强调所谓的一般化能力，倒置了技能与知识，忽视了概念性与实践性知识的特殊性和差异。倘若内容手段化，那么内容就可置换；如果内容可以置换，也就是内容可以被虚空化。这意味着在

技能性话语中，知识被噤声了；当知识被虚空化时，所谓一般化能力脱嵌于具体学科知识之上，是悬浮且封闭的。

一般化的能力表现为对高阶思维、元学习的迷思，以及对不断挖掘的而非实现的潜能的推崇。这意味着要掌握更经济、更简单地获取信息的策略和程序，并不断应对短期主义的策略，即脱嵌于情境、压缩过程，并将经验抽象为图示。这只能通过符号化的算法学习来实现。从嵌入文化约束、扎根具体经验的求知形式转向抽象的、强调思维的方法训练，以测验代替训练，不断监控并强化学习者掌握非经验的算法能力——这样一条悬浮于经验之上的训练管道得以建立。它模拟了谁的学习？

安德鲁·阿伯特 2017 年在“知识的未来”演讲中分析了求知主体的演变。现代的求知者往往是物而非人，是机器在求知，它承担了个人无法处理的求知任务，而个人则因此无法透彻理解求知的过程。与此同时，知识商品化、信息相对过载与知识客体不断分解并行发展。多数时候我们并没有增加新的认识，而是将对象分解为更新的、更小的部分。过量的信息让人无法把握整体，也无法理解信息的意涵。求知行为也发生了变化，学习某个东西就是知道一个链接，阅读就是发现重要句子并划线。例如，发现亚当·斯密著作中真正重要的五六个句子，忽视或者省略无关紧要的琐碎细节。求知不外乎寻找事物和制作列表，运行算法且无须任何长期记忆堆砌。伴随着算法操作，学习发生了剧变：以发现为求知，以列表为论证，朝向基于计算机求知及其所擅长的搜索与模拟的转变。它朝向了一种“索引-搜索”的求知模式，复杂而离散的

①② 巴兹尔·伯恩斯坦：《教育、象征控制与认同》，第 82 页，第 84 页。

③ 麦克·扬：《课程与知识的专门化：教育社会学研究》，许甜译，上海：华东师范大学出版社，2021 年，第 187 页。



① 安德鲁·阿伯特：《大学教育与知识的未来》，王桐等译，北京：生活·读书·新知三联书店，2023年，第202—220页。

思想转化为以字斟句酌的文本或符号的表现形式来分享的思想，并转向即时的口头和视觉表达。^①

算法又是如何训练今天的学习者？他们既能享受到影音扫描、检索、数据库等带来的巨大便利，同时也会深感自己的无知与怠惰，一种“心安理得的无知”，没有下功夫系统地学习“知识”，时常为自己空空的脑袋焦虑，但又没有下“笨功夫”的决心和时间——要应付眼前的任务已属不易。

把人的教育带回来

希拉里·普特南在1981年提出了著名的思想实验——“缸中之脑”。其基本设想是，将一个人的大脑放进装有营养液的缸中，脑的神经末梢与计算机相连，借助预设程序输入的信号，使大脑维持与往常生活完全一样的错觉。

“缸中之脑”实际上揭示了悬浮、透明、空转的管道的本质。在这条悬浮、封闭的管道的出口，一个新出炉者困惑地说：在知识与信息如此充足的学校里学习，却感觉像是翻开一本字典，可以便捷地从A翻到Z，但是什么都记不住，好像什么也不需要记住。如果学习只是在掌握链接、搜索与查找、信息的编排与重组，那么学习者就只是在执行外部的指令，成为智能学习者拙劣的模仿者。找寻与索引在他的生命中没有留下任何的痕迹，他的头脑空荡荡的，如同管道一样，信息随时输入、组装，再快速输出。他也如同一块无形无态的拼图，可以随意拆卸组装，敏锐地感知外部的位置，并自主地调控色彩，融入周围。他最大的困惑是没有困惑，他已将自己的身体和人格简化为一台运转正常、能优

化升级的电脑，以及可以灵活敲打键盘的手指。

人的学习不是这样的，他需要专注地沉潜，用火慢炖的时间，不断寻找一些“小钩子”：事实、概念、记忆、论证、经验的暗流、含混的感受、细致的文理。学习者要浸泡在这些有质感的东西中，不断唤醒内在知觉，融通经验与文本的内外感知力，这时他的心灵才能苏醒。人的观察之眼需要在微暗的光影中慢慢练习，幽暗的光让他学会辨识影子，能够看到一束光线所射之外丰富的物与景。他的心灵之眼会慢慢睁开，不仅能近察，还能远观，更能识别与体悟。只有这样的沉浸所获得的体悟才能入脑入心，才能滋养精神生命。

在算法学习中，人们没有了含蓄、隐匿、需要用心找寻的知识体验，而是被抛置在过量的、短暂的、明确的信息海洋中去冲浪。过量的信息、过度的光亮，刺痛人的眼睛，使人看而不能见，见而不能识。教育应该如何保护人的观察之眼、心灵的感受与体察之眼呢？

人的教育最基本的三要素（3H）是脑（head）、心（heart）和手（hand）。脑关乎认知，能够直观世界图景而非被碎片信息吞没；心关乎情感与信仰，在关系之中沉浸，才有熟悉、安全、爱与敬；手关乎练习，在具体的操作中了解事物的纹理、世事的繁简，才有上手的分寸感。这些都需要过程和试误。教育是人类孕育新生命的孵化器，安全与信任是其本质。算法高效地训练，否定了人的可塑性和柔韧性。

人是如此脆弱又是如此柔韧，如此卑微又如此神圣。我们既能站在类的高度，也应该站在人的经验和心灵深处。人类有必要接受教育的庇护。更准确地说，人类既需要谦卑也需要勇气庇护自己的教育。

ChatGPT 对新闻传播系统的颠覆性重构

朱鸿军，中国社会科学院大学新闻传播学院教授，中国社会科学院新闻与传播研究所研究员，教育部青年长江学者

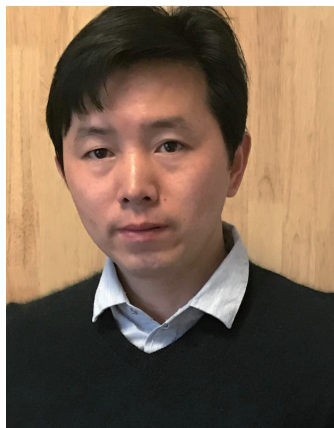
与其他行业相比，新闻传播业在大规模知识内容的生产、流通和服务方面扮演着主要角色。而在大数据和大模型的辅助训练下，ChatGPT 通过学习获得了与人类进行双向知识内容交流的能力，这与人类的思考逻辑和认知能力密切相关。新闻传播业与 ChatGPT 在业务上有着最为紧密的联系，面临着全面、剧烈和深刻的颠覆性冲击。这种冲击的传导、扩散、渗透，将对新闻传播领域的产业实践、教育及人才培养理念与模式、知识生产与研究范式产生系统性、革命性的影响。尽管当前模型的性能尚不稳定，但以 ChatGPT 为代表的大型语言模型仍在不断迭代进化。尽管未来模型的发展方向仍存在许多不确定性，在这种语境下谈新闻传播系统如何涅槃重生，更多会陷入猜想臆断，但整个新闻传播系统在应对 Web 2.0 时捉襟见肘的教训，以及目前 ChatGPT 在新闻传播业实践中的应用趋势，都提醒我们从维护系统合法性的角度来思考 ChatGPT 对新闻传播系统的颠覆性影响。此乃“未雨绸缪”，有其合理性。

新闻传播行业系统更添不确定性

当前的新闻传播行业系统，更准确地说是在大众新闻传播业系统，其主体是基于图书、期刊、报纸、广播和电视等传统媒体的形态、业态和生态特性而形成。一直以来，该系统处于相对稳定状态，虽然在历史演变中会因新生媒体的出现而短暂调整，但该系统的底层逻辑、主体特性和运行规律并没有质的变化。原因在于无论是纸质媒体，还是视听媒体，其信息传播技术特性在本质上是一致的，即都是由信息传播中心向周边受众扩散的放射性传播，其传播的范畴领域更多属于大众传播。

然而，自 1990 年代中期商业互联网普及化、泛

众化以来，以传统媒体为革命对象，并致力于打破传统媒体中心化传播范式的互联网媒体，更多采用的是有助信息传播网状化、扁平化的互联网信息传播技术，如数字技术、交互技术、分组交换技术等。这使得互联网媒体在形态、业



态和生态方面与传统媒体相比，具有根本性差异，这自然导致传统新闻传播行业系统出现了前所未有的变革。加之互联网技术自身的加速迭代，从底层逻辑层面使得整个新闻传播业的生产传播流程、内外经营管理机制、宏观管理体系，乃至全球的新闻传播秩序都处在剧烈的振荡中。

互联网基于开放、匿名与免费等特性，在 PC 互联网时代已经对传媒产业的内容生产模式、盈利模式和行业监管模式进行了一次重塑。随着移动互联网在传输速率、带宽和稳定性上的全面提升，加之以智能手机为代表的移动智能终端的普及，传媒产业在内容编码、传输网络和终端方面实现了全面的数字化转型与融合。在过去的 10 年里，我国新闻传播行业基本实现了传统媒体与新兴媒体的融合发展，如今即将进入媒体深度融合阶段。从协同进化理论 (co-evolutionary theory) 来看，^①媒体深度

^① M. Latzer, "Media Convergence," in *Handbook on the Digital Creative Economy*, Edward Elgar Publishing, 2013, pp.123-133.



融合是技术、经济、文化和政治四个方面的全面融合。在融合的第一阶段，新闻传播业面临的主要是技术上的不确定性，即应该及时完成技术的更新换代，掌握新的工具，应用新的方法，否则就会落后甚至被淘汰。在融合的第二阶段，在技术不确定性之外，新闻传播业更多面临的是用户、市场及环境的三重不确定性。其中用户的不确定性主要由终端移动化、智能化所培育的用户碎片化、个性化信息需求与消费习惯所致，因此媒体需要不断创新内容形式、提高服务质量、增强互动性与沟通效果，否则就会失去用户的关注。市场的不确定性主要源自媒体之间日趋激烈的竞争和用户的多元需求，这需要媒体创新商业模式、拓展收入来源、“降本增效”，否则就会被市场淘汰。环境的不确定性主要由媒体所面临的快速变化的竞争规则以及日趋复杂敏感的舆论环境所导致，这要求媒体在追求经济利益的同时，不断关注社会动态，遵守法律规范并履行社会责任，否则就会面临舆论压力乃至法律制裁。

ChatGPT 能够根据给定的关键词或话题生成流畅、连贯、有逻辑的文本。其强大的内容生成能力，对于新闻传播行业来说，是一个巨大的机会。一方面，利用 ChatGPT 帮助新闻工作者提高工作效率、减轻写作负担、拓展创意空间，已经成为新闻媒体行业内的新竞争赛道。全球首个完全由人工智能生成内容的新闻报道平台 NewsGPT 已经投入运营，《纽约时报》则直接将 ChatGPT 整合进其内部的新闻生产流程中，德国出版巨头 Axel Springer 最近在其新闻网站上发表了首批 AI 撰写的文章。现有实践趋势表明，如何最大限度地利用

ChatGPT 的内容生成能力已经成为新闻传播行业必须直面的问题。另一方面，ChatGPT 本身也存在一些问题，如训练数据、模型和生成内容的不可控性。ChatGPT 本质上是基于海量互联网数据训练而成的概率猜词大模型，其训练数据规模之大无法保证其初始信息的真实性和准确性。此外，高达千亿甚至万亿的模型参数使得开发者本身也无法对模型进行全面控制。由于训练数据和模型的不可控性，ChatGPT 生成的内容极可能缺乏真实性、客观性和准确性。这些问题需要在使用 ChatGPT 时予以关注和解决。

ChatGPT 本质上是一把双刃剑。它提升了新闻传播业的信息生产效率，降低了信息生产成本，但同时也意味着生产虚假新闻和造谣传谣的成本降低了，这将对整个新闻传播业的内容生态治理造成颠覆性冲击。可以说，ChatGPT 语境下，新闻传播业面临的技术选择、用户增量、市场竞争、环境监管的不确定性并未减少，相反，其可能带来更多新闻伦理、真实性等方面的问题，影响新闻传播的公信力。

此外，我们还需警惕谷歌、微软等公司将 ChatGPT 融入搜索引擎后，对新闻传播业流量的进一步侵蚀，因为用户将不再需要点击并浏览网页本身。这将对新闻传播业市场造成双重挤压。ChatGPT 以合理使用为由无偿使用新闻传播最宝贵的文本、图片、视频和音频内容进行模型训练，但其生成的内容却可能对新闻传播业产生反向压迫。这种利益分配的不合理也将引发新闻传播行业中内容收费模式和知识产权保护模式的双重不确定性。

新闻传播教育系统更增合法性担忧

1908 年，世界第一所新闻学院在密苏里新闻学会的资助下，由美国著名新闻记者沃尔特·威廉姆斯在密苏里大学创建。1918 年，北京大学开设了新闻学选修课，在美国密歇根大学攻读经济学、新闻学的徐宝璜担任中国大学的第一位新闻系主任。新闻传播教育诞生之初，无论是在国内还是在国外，服务新闻传播实践，为新闻传播实践培养人才，都成为其最核心的使命所在。1921 年，“美国新闻教育之父”威拉

德·布莱耶在西北大学麦迪逊新闻学院发表演讲时就曾指出，新闻教育最重要的是要让学生有“广博的知识和怎样运用这些知识来付诸实践的能力……这种能力指的是一种强烈的社会责任感和成为一名好的记者的良好的素质”。^①在中国，徐宝璜、戈公振、邵飘萍等早期的新闻传播教育先驱无一例外有着丰富的媒体经验，他们培养的学生也主要进入新闻传播行业。在新闻传播教育后续一百多年历史中，其最初的核心使命并未改变，可以说，“新闻传播教育是新闻传播业之子”，新闻教育因新闻传播业而生，也因新闻传播业变化而做出与时俱进的调整。

1990年代中期以来，互联网媒体的兴起引发了传媒业的巨大变革，并迅速影响到新闻传播教育系统。现有的新闻传播教育系统在面对互联网媒体行业的人才需求时，无论是教育理念、教学目的与计划、课程与教材，还是教学方法、教师职业素养等都出现了显著的不适应。这种不适应表现为“培养的人才未大范围流入新闻传播业，或者并不受新闻传播业欢迎”，这反映了培养人才的导向与社会需求不符，进而引发了新闻传播教育系统的专业危机。应对这个问题不仅需要现有的新闻传播教育系统进行优化调整，更触及其核心使命——“服务实践”的实现。如2023年初，哔哩哔哩网站上一条标题为“我：毕业5年，存款5000 她：中传硕士，火锅店保洁”的爆款视频引发了广大网民的热议与共鸣。虽然视频中两位女生的遭遇受到了个人主观因素和疫情客观因素的影响，但也从某种程度上折射出我国新闻传播教育人才培养路径所面临的危机。

以ChatGPT为代表的生成式人工智能已经展现出极强的知识学习能力、知识迁移与应用能力。在人类的反馈调节和开发者的调试下，该模型性能的边界将持续扩展。微软研究院最近公开的一份关于GPT-4的测评报告表明，^②最新版的ChatGPT不仅在文学、医学、法律、数学、物理科学和编程等单个学科领域表现出极高的熟练度，并且能够综合多领域的概念和知识，呈现出对人类已有复杂思想的深刻理解。最重要的是，尽管该模型是基于文本训练的，但也展示出理解和处理多模态信息的能力。基于以上测试结果，研

究人员得出结论，即“鉴于GPT-4能力的广度和深度，我们已经有充分理由相信GPT-4应该被合理地视作一个通用人工智能（AGI）系统的早期（但仍然不完整）版本”。

强大的“类人”的知识生产、交流和优化平台是ChatGPT的核心终端产品，随着其不断进化，必然会对人类整个知识教育体系产生重大影响。新闻传播教育的使命是培养学生学习和研究新闻传播领域的知识生产、传播规律及技能。史、论和业务构成了国内外新闻传播教育知识体系的三大板块。史学部分主要涉及的是报纸、期刊、广播、电视等传统媒体领域的新闻史和人类的大众传播史。理论部分也主要是基于传统媒体实践规律而生发的理论。然而，传统媒体大幅度式微，可能会引发大众对于学习其历史和基于其实践所生发理论的质疑。这样的质疑声在过去二十多年互联网媒体盛行的时代便已存在，在ChatGPT即将盛行的时代，会加倍放大。在业务知识板块，新闻传播教育传授学生新闻采写与编评、音视频领域的采制编播等知识技能，ChatGPT将改变这一切，它们中的不少将被ChatGPT所掌握，并且随着人工智能的不断进化，学生们掌握这些知识、技能的范围和水平也必将大幅度调整。

新闻传播研究系统更待主体寻找

相较于文学、史学、哲学、政治学、法学等老牌人文社会学科，新闻传播学是一门相对年轻的学科。作为现代新兴学科，新闻学和传播学都是在借鉴其他学科的基础上发展起来的。建立起独特的专业知识体系，获得其他人文学科

① Carolyn Bronstein and Stephen Vaughn, Willard G. Bleyer and the Relevance of Journalism Education, Association for Education in Journalism and Mass Communication, 1998, p.10.

② S. Bubeck, V. Chandrasekaran, R. Eldan, et al., "Sparks of Artificial General Intelligence: Early Experiments with GPT-4," arXiv preprint, <https://arxiv.org/abs/2303.12712>.



圆桌会议现场

学学科的认可，并成为独立的学科，是新闻传播学在诞生初期和发展过程中所努力追求的目标。伴随着报纸、期刊、广播、电视等大众媒体在现代社会的盛行，探索这些大众类媒体的传播规律及其对人类社会的影响自然成为新闻传播学的核心研究对象。

进入互联网时代，对实践反应极强的新闻传播学自然会将研究对象延伸至互联网媒体。这种延伸给新闻传播学研究带来的影响有：一方面，互联网媒体如空气般对现代社会方方面面的渗透，使得互联网媒体研究的主体学科新闻传播学被推到了历史的前台，官方、业界和学界乃至社会大众，都对其寄予较高期望，这为一直以来努力争取学科合法地位的新闻传播学带来了机遇；另一方面，在互联网逐渐成为现代社会运行的基础设施，“互联网+”几乎可以统摄现实物理空间所有领域的语境下，互联网传播也成为各学科研究都需要涉及的领域。跨学科为互联网传播研究带来更广阔的视野和全新的研究成果，但其他学科的介入也将逐渐弱化新闻传播学本学科研究的主体领域。此外，学科成长时间不足造成研究主体领域在概念、理

论、方法、范式等方面积累不够，互联网空间又偏向以全类型传播为核心研究对象，这对新闻传播学研究主体的知识体系建构构成巨大挑战。所有这些因素使得在互联网时代，新闻传播学的研究正陷入主体性衰弱的境地。

进入 ChatGPT 语境后，新闻传播的形态、业态和生态都受到了颠覆性影响。在传统媒体环境和互联网 1.0 和 2.0 时代的传播环境中，无论是大众传播、组织传播、人际传播，还是传者中心主义传播、受者中心主义传播以及传受互动中心主义传播，都是以人类中心主义为基本前提的。然而，以机器智能生成内容的人机传播，及人工智能技术不断迭代和进化而带来的机器主导的人机传播，正使得机器中心主义构成传播实践的底层操作系统。这样的传播实践革命自然会导致基于人类中心主义传播实践而形成的新闻传播学研究主体知识出现更大的不适应性，如 ChatGPT 情境中，新闻传播学的新闻、传播、媒介、媒体、受众等基础概念，“把关人理论”“议程设置理论”“使用满足理论”“沉默螺旋理论”“涵化理论”“知沟理论”等经典理论是否需要重新界定、再次检验，甚至是否已完全不适用，新闻传播学所采用的人文研究和社科研究路径是否会成为结构性的落后手段，都是需要重新思考的问题。当构成新闻传播学研究主体知识的这些方面发生系统性的调整时，新闻传播学的研究主体就会变得更加迷茫。

现代的新闻传播系统随现代的新闻传播实践而生，每一次重大的新闻传播实践革命都会引发其变革。与人类历次新闻传播实践变革所不同的是，ChatGPT 这一“人工智能平台”，其功能直接对标的是人的智慧，甚至超越人的智慧。这样举世未有的“机器野心”，所带来的不单是对新闻传播系统的颠覆性影响，更可能是对整个人类系统、人类文明的颠覆性重构。如何让这样的“机器野心”不成为脱缰野马，让其始终处在人类的控制范围内，需要包括新闻传播系统在内的全人类各系统集体警觉、探究反思、寻找路径。当然很有可能，这样的超前思维只是杞人忧天，或是主观臆想！

“诸神之战”：新工具抑或新存在

王金林，复旦大学哲学学院教授

ChatGPT 在整个 AI 行业乃至社会各界引发了一场“诸神之战”。OpenAI 公司作为 ChatGPT 的开发者，在介绍它时只是低调地把它描述为一种对话式人工智能：“它能够以对话方式进行交互。对话格式使 ChatGPT 能够回答后续问题、承认错误、挑战不正确的前提并拒绝不适当的请求。”尽管这种对话式人工智能看似平常，但发布后却迅速成为一种现象级应用，并牵动各界展开激烈的争论。

英伟达 CEO 黄仁勋赞誉 ChatGPT 为 AI 领域的“iPhone 时刻”。基辛格与 AI 业内巨头等联袂撰文，断言“ChatGPT 预示着一场智能革命”，为人类理性开辟了革命性的途径，同时也提出了自启蒙运动以来规模空前的哲学和实践挑战。他们希望通过“确保机器仍然是从属于人的对象”来重申人性。语言学大家乔姆斯基等人则持完全否定的观点，直斥大语言模型的非道德性、伪科学性和语言无能。他们认为 ChatGPT 只能描述和预测，完全缺乏解释能力，而解释才是“真正智慧的标志”。ChatGPT 要么“过度生成”（即既产生真理也产生谬误，既认可道德的决定也认可不道德的决定），要么“生成不足”（即对任何决定的不承诺和对后果的漠不关心）。深度学习巨头 Yann LeCun 等人也认为 AI 聊天机器人不了解人类诚实和体面的规范，并不以人类的方式进行对话，因此最多只是“非常有趣的工具”，不应将其“拟人化”。

倘若将上述评论拿去问 ChatGPT，其回复会既谦和又辩证：“我”只是一种大语言模型，没有自己的价值偏向与观点，但是“我”可以在若干方面对人们有所帮助，不过“我”也会产生事实性错误即所谓的“幻觉”，因此不能完全依赖“我”的回复来做决策，而必须进行核查，等等。ChatGPT 的回答当然不无道理。然而，这只是一种工具论意义上的道理。如果完全接受这一观点，那就误入“机关”了。就其本质和逻辑而言，ChatGPT 及其迭代产品 GPT-4 不仅是一种新工

具，而且预示着一种新的存在。人类社会或将因此发生重大变革。这个问题事关重大，涉及广泛且深刻。受篇幅限制，本文仅从知识生产方式、语言生成方式以及人的对象性存在方式三个方面进行论述。



首先，从人类知识的生产方式来看，ChatGPT 带来的是变革。知识之于人至关重要。亚里士多德在《形而上学》开篇即言“求知乃人之本性”，培根更是断言“知识就是力量”。不过人们往往只从认识论或知识论的角度来理解这些洞见，而忽视了存在论或哲学人类学的观点。这导致知识与人性被误解为二元对立的观念。知识似乎仅仅被视为人类认识和改变世界的产物，无关乎人的本质规定。这显然是一种工具论的知识观，它根本不懂知识对人性意味着什么，也不懂知识生产方式实际上也是人性建构方式。知识与人性存在着本质的关联。人类生产的知识以及知识的生产方式塑造了人的本质。

GPT 模型的当前表现证明生成式人工智能已经具备了知识生产的能力，并且其生产方式与人类不同。GPT 模型的开发思路是让机器具备一种不受监督的学习能力，以使用自然语言模型培训所

有内容。训练 ChatGPT 的核心逻辑在于训练神经网络体系，让它去预测下一个单词，发现现有数据的逻辑，从而去生成新的知识。有人也许会反驳说：这只不过是基于统计学对现有文本或知识进行新的排列组合而已。从表象上看，这种观点当然有道理。然而，这种新的排列组合实际上意味着一种新的知识生产方式。自然语言处理和机器学习技术以远超人类的速度更快地处理和分析海量的数据和信息，从而能够发现新的知识和模式。ChatGPT 在回复笔者的提问时，简要概括其工作原理为五个步骤，即输入信息、文本解析、知识检索、知识生成和输出结果。可见，知识生成是其核心功能。当然，它也提醒我们，其知识生成能力取决于其所学习的数据和算法，因此所生成的知识可能不一定准确。

然而，问题并不在于 ChatGPT 生成的知识是否准确可靠（它的可靠性的提升只是时间问题，GPT-4 的发布已经证明了这一点），而在于它不仅能够完全理解人类的提示，而且它的回复呈现出类似于人类知识的样式。在进行了数百亿或数千亿参数级别的大规模语言模型训练之后，它甚至出现了一种其开发者无法完全理解的“涌现”现象。这就是业内所称的“狂卷语料库，大力出奇迹”。基辛格等人断言：“一项新技术正在改变自印刷术发明以来从未被动摇过的人类认知过程”，必将“重新定义人类知识”。倘若此论不虚，那就意味着人类的知识生产方式正在转变为人机合成的方式。在合成生物之后，现在出现了合成知识。这可能造成什么样的结果尚难预料，但可以肯定的是，随着人工生成知识的出现，“知情意”——规定人性的三大要素——中的知识及其生产

方式正在经历颠覆性的变化。一种新的存在正在形成，而 GPT 模型则是其前兆。

其次，从人类语言生成的方式来看 ChatGPT 带来的语言比知识更为根本，对于规定人性而言具有重要意义。然而，语言与知识一样，常常被视为一种仅仅用于交流的工具。尽管这一观点有其日常经验的支持，但它仍属于工具论的表面观点。亚里士多德将人直接定义为“逻各斯的动物”，正是因为没有语言或逻各斯，就不会有所谓的人了。维特根斯坦对此的理解是：想象一种语言，就是在想象一种“生命形式”。海德格尔更是把语言视为“存在之家”。没有语言，人生在世这件事根本不可能。语言并非一种可供人任意使用的交通工具，而是对人的本质的基本规定，因此他甚至认为并非人在说话，而是语言在说话，人说只不过是应合着语言而说。

ChatGPT 等 GPT 模型具有语言能力并开始生成新的语言。其基本机制是通过分析大量的语言数据来学习语言的结构和规则，从而生成新的语言表达方式。这固然使人机交互变得更加智能化，但同时也意味着人与语言关系的异化，语言开始成为人机混合物。机器语言逐渐进入人类的“自然语言池”中。这是否会像被编辑的基因污染人类基因池那样污染人类的语言池呢？或许有人会认为，无论机器多么精通语言，它所表达的仍然是自然语言，只是在自然语言的现有框架内进行“嬉戏”而已。笔者对此并不认可，ChatGPT 所生成的语句、短语或词汇正在被人们采纳并被纳入日常使用，这在一定程度上可能会影响自然语言的句法和语法。更重要的是，正如 ChatGPT 的回复所言，“智能体的出现意味着语言将不再仅仅是人类的专属领域”，并且“从更广义的角度上看，语言的创造者可以是任何具有能力创造和传播符号系统的智能体”。

我们当然知道目前 GPT 模型仍然可视为基于程序和算法的工具，其所谓的“创造性”是建立在对大量语言数据的学习和生成之上，而不是基于自主意识和意愿的创造。对话式人工智能实现了人机之间的自然语言交互，这无疑极大地降低了人机交互的门槛，提高了交互效率，从而倍增了生产力。但是无可否认的是，自然语言在此过程中显然被智能化了，其效率或准确性也许更高了，但自然语言的复杂性与丰富性却面临

丧失的风险。语言作为逻各斯、“生命形式”与“存在之家”已经面目全非了，一种新的机器语言正在形成，或者说人类语言池已经不再“自然”了，而语言池的污染显然会悄然地却又实质性地影响人与存在的关联。

在 GPT 模型出现之前，人类技术产品中最可怕的莫过于核武器，但此“大杀器”终究只属于器，与源于存在之根的大语言模型（或可称之为“大语器”）无法相提并论。大语器完全有可能越器而出，不复为器。如果说大杀器因其可怕的毁灭力量让人警醒的话，那么大语器则完全可以在潜移默化中对人性进行重构，实现“润人细无声”。

最后，从人的对象性存在方式来看 ChatGPT 带来的变革。人是一种对象性存在，也就是说人总是要把其本质力量对象化，并由此返回或反观自身。在中世纪，人将其本质力量投射在上帝身上，视其为全知全能全善的造物主，而人则同其他一切存在者一样成为被造物。近代以来，人开始充分肯定此岸世界的力量，把自己而非神确认为中心，重新占有了曾经被神所攫取的人的内在力量。这场人神之战被尼采概括为“我们杀死了上帝”。因为神终究只是一种观念性存在。

现在，一些原本属于人的能力和属性，如思考、决策、感知等，被投射到了人工智能身上。GPT 模型就是这种投射的最新成果。一场新的人神大战或许已经开始。然而，这一次的“神”不再是观念性的存在，而是以工具形式介入世界的对话式 AI，其存在完全可以充满实在性。那么，人类能否如法炮制，再次战胜自己所造之“神”，重新占有自己的本质力量呢？黑格尔的主奴辩证法的悲喜剧（是悲是喜取决于立场）是否会在人工智能领域重新上演，演变为“人机辩证法”？

必须强调的是，GPT 模型所预示的新存在并非诸多存在者之一种，而是马克思所述的“特殊的以太”“普照的光”。一切存在者都将在其普照之下，重新证明自己的存在合法性。人、社会与自然均会被重新规定，人类社会或许会从此逐渐步入“人机社会”。ChatGPT 的开发者们对此并不担忧，他们反复强调人工智能和通用人工智能（AGI）毕竟是由人类构建的工具和系统，其行为和表现也是由人类的目标和规则所决定的。虽然它们可以模拟某些活动，但并不具备真正的自主性和创造

力，而人类则具备智能体所缺乏的特征和优势，如情感、道德、创造性和自我意识等。姑且不论智能体以后能否在这些领域形成超过人类的能力，一种可能的情形是这些人类的独特优势在智能体时代将被边缘化，进而被消灭，就像工业革命初期机器纺纱碾压手工纺纱那样，关键在于主导性原则是计算性思维。

“ChatGPT 之父”奥特曼对更加强大的 AGI 的开发亦不无犹豫。一方面，他强调要确保 AGI 在“提升人类”方面造福于全人类；另一方面，他也承认这种“力量倍增器”存在巨大的潜在风险。他甚至表示，成功地过渡到一个拥有超级智能的世界可能是人类历史上“最重要、最有希望、最可怕的项目”。为此他提出了 AGI 三原则，即“赋能、共享与谨慎”，主张逐渐过渡，让整个社会有时间探究发生了什么，并制定应对策略，使 AGI 的发展方向“对齐”人类的价值与目标。马斯克等人呼吁暂停巨型 AI 实验的公开信也表达了同样的关切。

然而，遗憾的是，这一切努力显然仍然囿于工具论的范畴，仍然是在技术性规定这个大范式之下进行的。无论是逐渐过渡，还是暂停试验，都或许可以让人们从表象上对智能体有所认知（但对其本质与意义何在却仍然盲然不知），更不可能真正逆转新存在的到来。相反，AGI 却可能日益“了解”人类。人类几乎全部经验、知识和智慧都蕴藏在语言中，现在 GPT 模型已经学会使用语言，人类似乎已无秘密可言。

[本文系教育部人文社会科学重点研究基地重大项目“后现代的宏大叙事——《帝国》三部曲批判”（15JJD710003）的阶段性成果。]

作为新一代生成式人工智能技术的代表产物，ChatGPT 通过连接大量的语料库来进行模型训练，并进行深度学习，其公开发布和应用推广在产生“技术奇点”的同时，也正在引发一场人工智能技术新革命。



人工智能大模型技术 (transformer) 路线，包括其取得的生成式人工智能技术进步，展示出的通用人工智能 (AGI) 特征，使人们看到了强人工智能时代的开启，并引发了技术进步主义者与技术规范主义者之间的激烈争论，进而引发人类开发工具与工具脱离人类控制的悖论性命题。更为现实的且不可忽视的是，与人类主体性并存的是民族国家的主体性存在，即国家在发展人工智能技术中的角色与抉择同样重要。特别是在人工智能带来无限可能时，国家如何通过人工智能技术手段塑造与强化自身的国际权力？而在国际权力被技术快速嵌入的新角逐中，人工智能技术又如何与国家进行互动，进而创建大国博弈新边疆？

工业革命与国家竞争

历史已经证明且会继续证明，任何

新一代人工智能技术与大国博弈新边疆

余南平，华东师范大学政治与国际关系学院教授

技术的创造或是源自军事用途，或是最终可能被用于军事用途。以近代技术提供的例证而言，第一次工业革命的成果使得 1849 年法国拿破仑号成为第一艘采用蒸汽的战列舰，1854 年英法俄克里米亚战争中图温南步枪的射速、精度，使得俄国的旧式军队无法应对散兵线的作战，拿破仑时代的纵队战术在技术面前显得不堪一击。第二次工业革命中汽车、飞机、无线电通信纷纷登场，其产生的生产力大大推动了经济社会进步，但仅仅是十年时间，随后第一次世界大战中，重炮、重机枪乃至化学武器等的使用，使得欧洲成为战争的“绞肉场”。第二次世界大战中，德国坦克集群迅速突破了苏联的筑垒防线，空中轰炸使得欧洲付出数十万人的生命，而美国航母使用高强度的轰炸摧毁了日本的战争工业，背后都是各种技术的集成。第三次工业革命在军事意义上显得更为轻型化和文明化，导弹、电子计算机、卫星技术开始应用于军事领域。美国和苏联获得技术领先性，但第三次工业革命意外的结局是苏联的解体，随后是美国利用互联网对全世界进行意识形态输出。所以说工业革命本身既决胜负也定生死。

当下以 ChatGPT 为代表的大型语言模型，预示着强人工智能时代或将来临，人类社会主动和被动地进入第四次工业革命阶段已不可逆转。

当我们向 ChatGPT 发问：AGI 如何改变世界的地缘格局？它会轻松地回答道：AGI 是指能够超越人类智能水平的人工智能系统，如果某个国家率先获得 AGI，它可能会对其他国家造成巨大的威胁和影响，可能改变世界地缘格局的方式包括政治和文化影响，经济、社会、军事等领域的变化。而当你再进一步发问：请更详细地解释 AGI 可能如何在军事领域被应用？它会务实而前瞻地回答：AGI 在军事领域的应用，可以是增强人类士兵和武器系统的能力，也可以是开发自主决策和执行任务的自主武器系统；在战术和战

略方面，AGI 可以分析大量的战术和战略数据，帮助指挥官作出更加准确的决策，例如，它可以预测敌方部队的位置、规模和行动，为作战计划提供关键的情报；在无人机和自动化系统方面，AGI 可以用于开发更加智能和自主的无人机和自动化系统，这些系统可以自主执行任务，例如侦察、侦测和攻击，降低士兵的伤亡风险，提高作战效率和精度；在智能武器和导弹系统方面，AGI 可以用于开发更加智能和自主的武器和导弹系统，例如，AGI 可以使导弹系统自主选择 and 攻击目标，而无需人类干预，这可提高打击精度和速度，并降低误伤风险。

当然无一例外地，在回答这些问题时，ChatGPT 会一如既往地提示，国际社会需要密切关注 AGI 技术的发展，并制定相应的政策和法规来管理对它的应用，包括提示数据集可能包含敏感和机密的信息，在使用这些数据时需要采取相应的安全措施来保护数据的机密性和安全性等“善意”的建议。这些建议可能被重视，也可能被忽视。正如经济学家和技术史学家更关注工业革命给人类带来的进步，一些国家主义者看到的是人工智能技术革命大大拓展了军事竞争新边疆。而其对他国降维打击的潜能，可能驱动一些国家将强人工智能技术应用于军事以保持“技术自主”，而这一点在今天关于新一代人工智能的争论中已被证实。美国网络司令部司令在美国众议院质询中，拒绝了马斯克等人提出的对比 GPT-4 更强大的人工智能系统暂停训练 6 个月的“千人倡议”。

人工智能改写全球经济格局

正如 ChatGPT 所给出的显见答案那样，在经济领域，获得 AGI 的国家同样会占据领先地位。由于 AGI 在生产、制造和服务等领域中的广泛应用，某个国家的经济实力可能会远远超过其他国家，这可能造成世界贸易和金融体系的巨大变化。在强人工智能技术空间的拓展中，机器作为劳动力角色出现，并使得人类操作机器生产范式向机器介入劳动并自动生成劳动成果范式转化，其改变的不仅是当下经济学视角中的“劳动力角色”本身，而且随着边际成本快速下降，通过

全球化而形成的全球产业链、全球价值链将被改写，既有的经济学理论认知将被淘汰。更重要的是，当机器作为主要劳动力，其自循环能力产生时，发展中国家依靠国际分工而获得的低端收益将消失。国家主体性被消亡的风险，不会止于经济层面，其更多会反映到国家主体的生存博弈中。此时，技术民族主义和经济民族主义不再是一种政治学概念上的分辨，将可能转化为国家政治与经济行动的内在逻辑，在“生存”的第一驱动下变得异常强大。

可以预见，人工智能改写经济学范式是必然的。在布莱恩·阿瑟所定义“收益递增经济学”概念下，“技术本性”（technology-ness）将在人工智能技术自我迭代和强化中被进一步展现，在自创生的生物属性下，经济将被重新域定。而这个域更多不是既往熟悉的那样围绕核心技术联合而成，而是将从一个现象簇中构建起来。其结果完全不同于既往工业革命时代的国家全产业链，也不同于全球化时代全球价值链所主导的经济范式，新一代人工智能可以通过生成式人工智能—脑机接口—人形机器人完成自我闭环组合，进而产生强大的生产能力和自足可能。而掌握更先进人工智能技术的国家不仅可以一劳永逸地解决劳动力、移民、资源不足等困扰问题，同时还可以形成“高维度”生产力辐射，控制全球生产、贸易网络体系，在改变自身经济循环系统的同时，也可改变国家间的相互依赖关系。

人工智能正在改写自由主义经济范式并给全球化带来更新的难题，一些国家可能会基于对国际性经济权力和压倒性经济基础构建的追逐，将人工智能作为经济武器存在，进而视人工智能技术



为自身民族国家发展的底层屏障与依托,以此为基础进行国家间博弈与竞逐。

人工智能与意识形态博弈

技术与政治的互动性存在,使得技术的中立性一再被证伪。换句话说,即使天真地假设技术本身具有可能的中立性和工具性特点,但由于其产生离不开特定的民族国家,国家的政治与意识形态属性使得技术的两面性反复呈现——技术的主导国既可以将技术扩散与分享,作为推进人类福祉的工具,也可以将技术本身作为遏制他国的战略武器与工具。而后一种情况在今天大国博弈的环境下明显在加剧。正如 ChatGPT 自身所归纳的那样,获得 AGI 优势的国家可能会在政治和文化领域拥有更大的国际影响力。由于 AGI 可以处理大量的数据和信息,它可以用于影响和操纵公众意见,这可能会改变全球政治和文化的格局。而这其中的“文化”很可能是 ChatGPT 智慧地回避了政治自身的意识形态属性后输出的结果,具有极大迷惑性。

国家通过信息技术扩大自身政治合法性的传统研究一般聚焦两个方面:一是将信息技术作为一种生产力,研究信息霸权、信息技术垄断和信息科技资本的外溢;二是主要研究滥用信息技术带来的网络信息泄露和舆论风险。而与传统人工智能技术带来的信息安全影响力不同,当前 ChatGPT 的训练是依托于大模型、大数据集的红利,展现出的效果已经在一定程度上显现出强人工智能的属性,其借助超大的数据量和超强的算力,能够仿照人脑甚至在很多方面超越人脑的信息处理,突破信息

传输和共享的时空限制。在 AGI 的模式下,人工智能可以大量生成目标言论,模拟用户发言,并在用户基数庞大的互联网社交网络上进行舆论攻击。不仅如此,人工智能技术还可以通过深度伪造技术编造信息,并辅以各种迷惑性材料,生成虚构事件和虚假信息,甚至可以直接影响某些政治态度。而大量恶意虚假信息的传播也极易煽动社会的质疑和不满情绪,加剧矛盾分化,影响社会稳定。因此,AGI 不仅具有一般政治意义上的流程和导向影响能力,同时,在国际层面上,一些拥有强大生成式 AGI 能力的国家,还可能为了自身国家利益引导某种国际合法性,进行意识形态输出和国际舆论主导性掌控。因此,当人工智能技术通过业已普及的互联网跃然于国际政治舞台时,优势技术国家的传播力、渗透力或被明显加强,由机器不间断产出的各种信息充斥于网络空间,并使人工信息被淹没和屏蔽,进而产生了大国政治与意识形态博弈的新边疆。

综上所述,当尤瓦尔·赫拉利在 2015 年为未来谱写简史的时刻,人们只是为其想象力折服,并对其所描绘的智人、智神充满幻想,或许赫拉利也未曾想到未来已经在今天呼啸而来。而在这个呼啸的人工智能技术发展旋风下,技术已经不可避免地呈现了巨大的黑洞,并展示了无尽的前沿和遥远的新边疆。这个新边疆的拓展在某个层面或许依然被认为仅保留在技术与经济的互动之中,但是,值得牢记且不可回避的是,任何技术革命包括此次已经开启的由 ChatGPT 诱发的新一轮人工智能革命,其在历史进程中不可避免地嵌入国家角色并产生作用力,也不可避免地引发国家间围绕人工智能革命的新博弈与新竞争。人工智能技术的发展不会彻底消灭国家间的竞争,其自身会在不同国家的“生存正义”追逐下,在每个国家努力开拓技术新边疆的过程中被不断塑造!