

AI 时代的技术突破与风险反思

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

“每个人都是时代的产儿。哲学也是这样，它是把握思想中的自己的时代。”^①我们正加速进入一个人工智能时代。人工智能是技术和知识创新的母体，它正在对我们生活的世界产生支配性影响。马克思说：“在一切社会形式中都有一种一定的生产决定其他一切生产的地位和影响，因而它的关系也决定其他一切关系的地位和影响。这是一种普照的光，它掩盖了一切其他色彩，改变着它们的特点。这是一种特殊的以太，它决定着它里面显露出来的一切存在的比重。”^②人工智能就是我们时代的“以太”，它照亮了整个世界，决定着事物的价值，改变了社会关系，影响着民族国家内外权力关系的消长。

由于技术的先发优势和资本优势，美国长期支配着 AI 的技术发展路径。杭州深度求索公司发布的 DeepSeekV3 和 DeepSeekR1 大模型，以其高性能、低成本以及免费开源的策略惊艳了全球。在中国，DeepSeek 已被接入一些门户网站、云平台和政务系统，在医疗、教育、商业等领域得到应用。其在各个领域的影响使得全民接受了一次规模空前的 AI 启蒙。冰火两重天，它给美国同行和资本市场带来的却是“惊吓”。在 DeepSeek 的技术表现和免费开源策略面前，美国已有的技术垄断地位和资本赢利前景都似乎蒙上了阴影。由于它在美国主导的 AI 领域的强力介入，引发了全球竞争，AI 成了地缘政治的关键筹码，国际竞争也进入了一个 AI 时代。

进入 AI 时代

如何理解 AI 时代，我们可以从人机协作深度与 AI 自主性两个维度来认识。这一发展大致经历了三个阶段，今天我们正在从第二阶段向第三阶段，即向全功能 AI 阶段过渡。

第一个阶段是 AI 辅助时代 (Age of AI Co-Pilots)。在这一阶段，人类主导决策，AI 作为效率工具，辅助人类执行各项任务。其技术支撑是机器学习 / 深度学习基础模型，它可以完成复杂任务的分解，可以通过标准化流程进行数据分析，可以提



① 黑格尔：《法哲学原理》，范扬、张企泰译，北京：商务印书馆，1995年，第12页。

② 《马克思恩格斯选集》第2卷，北京：人民出版社，2012年，第707页。

供建议和预测最佳实践。AI 辅助技术今天已经广泛地被应用于分析医学影像、生产线异常检测等。这一阶段的特点是 AI 依赖人类的明确指令，在未预先定义的场景和未预先制定明确规则的领域，AI 不能起作用。因此，它的人机协作深度和自主性较低。

第二个阶段是 AI 智能体时代 (Age of AI Agents)。人们普遍认为，2025 年是智能体元年。所谓 AI 智能体，是指能够感知环境并据此做出行动的物体。这里所说的智能体，不仅仅指物理世界中可移动的智能物体，如机器人或机器狗之类的事物，也包括根据一般的指令，自主地分析环境信息、制定规划、调用工具自主完成任务的计算机模型。AI 智能体有环境交互感知和动态优化的能力。以环境保护为例，它能够通过传感器和监测设备收集大量环境数据；利用数据分析技术进行信息处理，实现对环境状况的实时监测和评估；通过智能控制系统对环境治理设施进行自动化管



理和优化控制；还可以通过建立环境模型和预测算法，对未来环境变化进行预测和分析。AI 智能体已被广泛运用于行程规划、产品设计、企业运营、健康管理、城市治理等领域。在这一阶段，AI 智能体实现了人机深度协作，人类确定任务和目标，AI 自主完成任务。

第三个阶段是全功能 AI 时代 (Age of AI Functions)，即通用人工智能阶段。AI 作为社会基础功能单元，具备类人的综合认知能力，能完成人的大部分任务。在技术上，全功能 AI 时代就是通用人工智能 (AGI) 时代，它具备认知泛化、价值对齐、自我迭代的能力。AI 全功能时代正在到来，随着个人智能助手和机器人的发展，AI 将完成人的大部分工作。因此，具身智能体的发展尤其重要。在中国和美国等 AI 技术先进的国家，一方面是计算机智能体的飞速发展，另一方面是机器人的发展，如宇树科技生产的机器人已惊艳全球。一旦具身智能体不仅具有了类人智慧，能完成环境感知、信息处理、任务规划的智力任务，而且像人一样在物理世界独立行走，它就会成为名副其实的“类人”行动者。人类就将真正进入人与 AI 共生、“硅基-碳基”深度交织的新文明时代。

DeepSeek 的“高光时刻”

直接意义上，DeepSeek 有两大明显优势，一是通过算法优化打破了大力出奇迹的规模化法则 (Scaling Law)，降低了 AI 的门槛和使用成本；二是源代码的开放，任何人都可以接入自己的网络，进行本地化布置和优化。这两个特征打破了美国对 AI 技术的垄断和霸权，以至于让美国人惊呼，又一次“斯普尼克时刻”来了。虽然业界一般认为 DeepSeek 是技术创新，不是科

学创新。但是，它的技术特点、成本优势和开源模式影响的不仅是中美技术实力的相对优势，还影响了全球资本的市场价值。毫不夸张地说，历史上甚少有像 DeepSeek 这样有着立竿见影效果的科技成果。不论该模型以后的命运如何，可以毫不夸张地说，AI 时代迎来了 DeepSeek 时刻，我们已处于它的影响之下。

第一，DeepSeek 高性能模型开源，极大地降低了 AI 技术的使用门槛，给全球打开了一扇机会平等的平权之窗。开源并非 DeepSeek 公司首创，OpenAI 公司初创时理念就是如此，Meta 公司的 Llama 模型也是开源的。但是，在与闭源模型的竞争中，开源模型一直处于不利地位。DeepSeek 彻底颠覆了这一格局，迫使 OpenAI 首席执行官奥特曼承认 ChatGPT 从开源到闭源是站在了错误的一边，犯了历史性错误。中国工程院院士王坚在世界互联网大会人工智能专委会的发言中指出，AI 不是一次工具的革命，而是一次科学革命的工具。可以说，AI 具有元工具的性质，是生产技术的技术，生产知识的知识。AI 的运用大大推动了科技的发展，大幅度地提高了生产效率。而开源策略又有助于推动 AI 技术本身的发展。

第二，通过程序优化，DeepSeek 有效地降低了 AI 的使用成本，提高了 AI 的可及性。AI 不是单一的工具，而是综合赋能的工具，它的可及性特别重要。可及性既可指没有人被阻止获得某物，也指人们经济上负担得起某物。旧时王谢堂前燕，飞入寻常百姓家。DeepSeek 向普通客户免费开放，让商业用户低成本使用，使 AI 技术为普通人所用，大大地促进了 AI 的普及。

第三，DeepSeek 以其高超的训练模式大大地降低了对芯片和电力的消耗，使 AI 产业可持续发展。DeepSeek 也许不是史上性能最好的模型，却在经济和环保方面迈出了重要的一步。马斯克旗下 xAI 刚刚发布了号称地表最强大的 Grok3 大模型。其性能比 DeepSeek 有提高，但它使用了 20 万块英伟达 H100 卡，要耗费高于 R1 模型 260 倍的电力。显然，这一模型对环境是不友好的。我们生活的环境是十分脆弱的，电力和资源节约十分重要，在这方面 DeepSeek 同样起到表率 and 示范作用。

第四, DeepSeek 的“深度思考”具有类人思维的优点。用户不仅得到答案,还直接看到模型的思考过程。这无疑提高了 AI 系统的透明性,增加了客户对它的亲近感。ChatGPT 问世时最为人诟病的就是 AI 像是一个“黑盒子”,充满着神秘和不可解释性。DeepSeek 展现的思维链已经部分地消除了对它的神秘感。DeepSeek 的思维链,就像一个思考者的自言自语,这种互动形式不仅增加了对话的乐趣,对人的思维也有反哺作用。

第五, DeepSeek 目前取得的成功体现了人工智能中人的作用。美国 AI 发展奉行大力出奇迹,实际上是把 AI 视作机器智能,或算力和数据的作用。DeepSeek 在技术方面取得的突破告诉我们,软件和算法与硬件和数据一样重要。人工智能应该像人一样聪明,像人那样思考,因为人的生物大脑是效能最高的智能体。实际上, DeepSeek 的多项技术,如大大提高 AI 推理效能的混合专家架构 (MoE)、用于长文本阅读的“原生稀疏注意力机制”(natively sparse attention) 等,都具有较明显的“人化”特征,充分体现了人的聪明才智。最近中国在 AI 领域取得的各项技术成就表明,中国不仅是制造大国,而且可以成为创造大国。

AI 竞赛及其风险

遗憾的是, AI 在全球的飞速发展并没有伴随着相应的国际治理体系和安全体系的建立。过去美国对 AI 的主导权相当于“罗马治下的和平”或寡头制。中国科技企业的发力使得 AI 领域由“帝国时代”进入“战国时代”。2025 年 2 月召开的巴黎 AI 行动峰会上, AI 安全主题已经被边缘化了,竞争和发展成为各国的优先选项。AI 全球竞争有积极意义,但如何确保其安全和向善是人类面临的严峻问题。

人工智能或许是人类有史以来最伟大的成就,而且有可能是最后一项伟大成就。然而,在笔者看来, AI 生长在一个有毒环境中。目前,人类面临着气候危机加剧、地缘政治紧张、民粹主义兴起、国际秩序重构等一系列问题,在这样的环境中, AI 为善的潜能

必然受到限制,而其为恶的风险则进一步放大。

可预见的风险有许多:劳动替代效应和大规模失业;被掌权者滥用损害个人自由;幻觉和假消息传播引发社会动荡;等等。最令人担忧的是 AI 掌握核武器,可能引发的毁灭性战争。从 ChatGPT 发布始,人们就在谈论价值对齐,但是,如果人类无法就 AI 的发展做到价值对齐, AI 价值对齐即使在技术上是可能的,在政治上也是难有保障的。

AI 教父、诺贝尔奖得主杰弗里·辛顿 (Geoffrey Hinton) 认为,我们生活在一个由“旧石器大脑、中世纪制度与类神技术”构成的世界中, AI 的力量与人类控制能力之间存在着严重的不对称性。固然,人脑是生物进化的最高成就。人脑在思维的灵敏性、适应性和能量消耗上有着 AI 难以企及的优势。但是, AI 大模型的规模集成所产生的智能可以远远超过人类。更重要的是,自从智人大脑形成后,上万年几乎不变,而 AI 由于算力堆叠和模型优化却在不断地迭代和进化。就此而言,如果与 AI 进行智力竞赛,人类注定会输。AI 正在成为一项“类神”技术,向着全知全能方向发展。“中世纪制度”当然是一个比喻,泛指人类制度的落后。即使文明发展到今天,人类在许多方面还受偏见和迷信纠缠,被种族主义、民族主义和各种其他极端主义意识形态支配。在全球范围内,我们正在向中世纪倒退,旧的野蛮死灰复燃,新的野蛮层出不穷,就像特朗普的政策和施政方式这样。

AI 带来的一类危险是滥用风险。这里所说的不是一般意义上 AI 的经济和社会冲击,而是指它有可能被政治家和技术寡头用来实施更加野蛮和残酷的集权统治。固然,我们的世界从来没有完美过。但是,

只要人的智力还受自然的随机分配，只要人类生命及其繁衍还受生物遗传机制的限制，只要统治者的权力还受到他的自然生命周期限制，只要人类还在通过语言相互交流，再野蛮的状态都有希望改变。

AI 滥用带来的另一威胁是人类毁灭或彻底被奴役的风险。辛顿在许多次访谈中都提到，虽然 AI 是人类的发明，但是，迄今为止，AI 智能是如何产生的，我们并未真的了解。我们不知道 AI 未来是否会自行修改人类的指令，按照“回形针最大化”方式行动，如果那天到来，那么人类有可能遭遇毁灭性风险。他还认为，即使 AI 不消灭人类，当其能力全面超越了人类之后，似乎难以想象它还会接受人类的控制。也许到那时候，人类的处境就像希腊神话中的人一样，依赖于一个比自己强大得多的事物。而神是无比强大的，无论在绝对意义上还是相对意义上，它都可以不需要人而存在。人类之所以被保存下来，只是为满足诸神自己的消遣。这提醒我们：我们正在养育一只老虎幼崽，必须确保它长大后无法杀死我们。

在当今科技与伦理领域，有效加速主义与有效利他主义是两大思潮。有效加速主义认为，技术快速突破是解决当前及未来社会问题的关键所在，人类应该尽可能减少对技术创新施加限制。加速主义指出，通过持续的技术创新和社会结构的系统性重构能够催生出更加理想的社会形态，有关 AI 统治人类或毁灭人类的担心是杞人忧天，技术的内在风险是可以通过试错和自我调节得到有效缓解的。这一思潮倾向于由少数技术精英来主导技术发展进程，避免受到不必要的干扰，而资本与技术的结合可以形成一个“正反馈螺旋”，快速推进 AI 发展达到其极限。

有效利他主义强调，在追求技术进步

的同时，必须优先考量其社会影响和安全责任。它甚至呼吁，在更为完善的安全保障措施出台前，应该暂停训练比现有最先进水平（如 GPT-4 或更高级别）更为强大的 AI 系统。有效利他主义支持全球合作和对 AI 的更严格监管，要求技术发展应当惠及全人类，注重技术应用的公平性和安全性，反对单纯以利润为导向的商业化模式。

在某种意义上，有效加速主义和有效利他主义的竞争是 AI 发展的资本主义与社会主义、技术主义与人道主义之间的竞争。今天我们应该更多地完善 AI 的应用，而非一味追求 AGI 或 ASI（超级人工智能）的发展。甚至 AI 技术的应用也需要与社会防护网共同发展。伊格尔顿曾说：“我们那些更乐观的先人有时会把历史想象为一辆列车，把我们从黑暗的山谷拉到光明的山顶。但是，如果历史是一列火车，那么我们需要记住那些从来不能到达目的地的人，那些因翻车而倒毙，或因绝望而跳车撞死在轨道上的人。”^①

不仅如此，AGI 或全功能 AI 是否值得追求也需要反思。本雅明曾指出，真正的历史唯物主义不是沿着无限的直线时间去追求连续的进步，而是准备在任何时刻使时间停止。如果上帝没有为 AI 的进化设定极限，人类就必须为它设定界限，如果它包含着危险的话。技术进步的相对成就与人类生存的绝对要求之间的关系是不对称的，保护人类生存和人类的尊严必须成为 AI 发展的绝对命令。在技术的毁灭潜能越来越明显的今天，我们不能把人类道德建立在对美好明天的盲目向往上，而应把它建立对人类毁灭的恐惧上。齐泽克指出，“悖论地说，唯一克服灾难的办法是接受灾难是不可避免的”，^②只有这样才能赶在爆炸前切断导火索。

这是我们在为 DeepSeek 所取得的成就而欢欣鼓舞时，需要思考的另一层问题。

[本文系国家社科基金重大项目“复杂现代性与中国发展之道”（15ZDB013）阶段性成果。]

① Terry Eagleton, "Waking the Dead," *New Statesman*, 2009-11-12.

② Slavoj Žižek, "First as Tragedy, Then as Farce," *Verso*, 2009, p.16.

ABSTRACT

China's Moment in Artificial Intelligence

Chen Xiaoping & Du Junfei & Shao Yilei & Ding Gang & Wang Tianfu & Cai Hengjin & Ren Jiantao
& Wu Guanjun & Wang Xingfu

Abstract: The advent of DeepSeek has caused a global sensation. In China, it is regarded as a “technological innovation of national significance”, while in the United States, it is seen as a “Sputnik moment” challenging its AI hegemony. As an important representative in the field of artificial intelligence, DeepSeek has not only promoted technological innovation but also triggered people’s rethinking about the development direction of AI and reminded people to deeply consider the future trend of human civilization. In view of this, on February 21, 2025, we held a interdisciplinary seminar titled “DeepSeek: China’s Moment in Artificial Intelligence?”. The seminar aimed to explore the profound impact of artificial intelligence technology on the development of human civilization, as well as the potential economic, social, and cultural opportunities and challenges it may bring.

Professor Chen Xiaoping pointed out that artificial intelligence is a human activity that studies and applies machine intelligence, and it should adhere to some basic principles, namely, adhering to the distinctiveness of machine intelligence, giving priority to human well-being, and ensuring the controllability of artificial intelligence. Professor Du Junfei put forward the concept and concerns of the “fantastic society”. He believes that AI continuously generates dialogues and continuously calls itself, forming an infinite recursive content stream. Professor Shao Yilei pointed out that the price war of AI models is one of the most important bloodless wars in this century and even in human history. DeepSeek has broken the US monopoly through technological innovation and reconstructed the global power pattern with an ultra-low price strategy. From the perspective of the subjectivity of generative AI, Professor Ding Gang discussed the intrinsic value and significance of AI in reconstructing Chinese education. He further called for the construction of a warm, profound, and inclusive future education system which combines subjectivity and humanistic care. Professor Wang Tianfu believes that digital intelligence technology has not only reshaped the way information is disseminated and interacted with, breaking the traditional limitations of time and space, but also redefined the interaction models among individuals, organizations, and society, forming a new pattern of social cooperation in the era of digital intelligence. Professor Cai Hengjin believes that the essence of DeepSeek’s technological innovation lies in decoupling. In the future, we still need to strengthen the research on human consciousness, which will help us better understand and develop machine consciousness, and thus promote the realization of superintelligence. Professor Ren Jiantao pointed out that the competition between China and the United States in the field of artificial intelligence should not be regarded as a zero-sum game, but rather global scientific and technological progress should be jointly promoted through cooperation and healthy competition. Professor Wu Guanjun pointed out that in the situation where artificial intelligence gradually possesses autonomy and creativity, we need to rethink human subjectivity and roles. Professor Wang Xingfu pointed out that DeepSeek has promoted the development and diffusion of AI technology and effectively challenged the US hegemony. An even more severe problem is that we are facing a structural dilemma of “paleolithic brains, medieval institutions, and god-like technologies”. Therefore, we need to think critically when it comes to AI fever.

New Labor: Artificial Intelligence, Technological Substitution and the Future of Employment

Yin Jiankun & Yuan Zhigang & Du Yang & Zhang Chenggang & Yang Li & Yao Jianhua
& Han Keqing

Abstract: As a crucial factor closely related to people’s livelihood, labor is facing the vigorous development of a new round of scientific and technological revolution and industrial transformation. New labor forms, new labor relations, and new labor structures are constantly emerging. In particular, the rapid development of generative artificial intelligence and humanoid robot technology is profoundly reshaping the appearance of labor employment and real life in the era of digital intelligence. In September 2024, we held a thematic seminar on “New Labor”, focusing on topics such as generative artificial intelligence and new labor, the innovation and transformation of new labor, and the risks and governance of new labor. It is hoped that we can comprehensively examine the evolution of “new labor”, properly handle the dilemmas of “new labor” in a forward-looking manner, and stimulate the potential of “new labor” in the process of Chinese-style modernization.

Mr. Yin Jiankun pointed out that the new employment forms catalyzed by new technologies are integrated employment forms. It is urgent to fully promote the high-quality development of new labor forms through