

# 当 AI 学会行动：技术哲学视域下的智能体崛起—— 评佩塔尔·拉丹利耶夫《AI 智能体的崛起》

夏凯捷

(广东金融学院 广东 广州 510521)

**摘要：**2026 年佩塔尔·拉丹利耶夫《AI 智能体的崛起》中文版的引进，与人工智能从对话机器人向智能协作体演进的技术趋势形成呼应，标志着 AI 正迎来从“工具”向“准主体”的历史性跃迁。本文以技术哲学为视域对该著作展开深度解读，指出其突破了“技术孤立论”误区，构建了 AI 智能体、区块链与量子计算相互赋能的系统性演化框架；借助唐·伊德的后现象学理论，剖析了 AI 智能体从人技诠释关系向他者关系的跃迁，探讨了“技术建构的主体”诞生的哲学意涵；结合技术批判理论揭示了三大技术融合背后的量子威胁、信任危机等系统性风险，并客观指出著作存在低估制度惯性与社会复杂性的缺憾。研究认为，该著作不仅为理解前沿技术融合趋势提供了系统框架，更启发学界与社会思考技术时代的人类定位，以及技术文明未来发展的深层命题。

**关键词：**AI 智能体；技术哲学；技术跃迁

2026 年 1 月，在 CES 展会的主舞台上，英伟达 CEO 黄仁勋宣告人工智能正从“一次性对话的聊天机器人”演进为“真正的智能协作体”。几乎同一时刻，牛津大学导师佩塔尔·拉丹利耶夫的新书《AI 智能体的崛起》由中信出版集团引进中文版。这不仅仅是一次出版上的巧合，更标志着一个技术时代的转向——我们正在见证 AI 从“工具”向“主体”的历史性跃迁。本文尝试从技术哲学的理论视域出发，对这部著作进行深入解读。

## 1 跳出“技术孤立论”的系统性视野

在人工智能已然泛滥的科普书市场上，多数作品仍陷于“技术孤立论”的误区：要么狂热追捧大语言模型的“涌现能力”，要么专注于区块链的“去中心化信仰”，要么将量子计算描绘成无所不能的“算力神话”。拉丹利耶夫这部著作的最大贡献，在于它系统性地打破了这种思维定式，将 AI 智能体、区块链与量子计算置于同一个演化框架下加以审视。

这种视野的开拓，首先得益于作者独特的学术背景。作为牛津大学计算机科学系导师、艾伦·图灵研究所研究员，拉丹利耶夫曾在剑桥大学和麻省理工学院从事博士后研究，拥有十年苏格兰皇家银行网络安全管理经验，并曾在英国国防部任职。学界的前沿洞察、业界的落地经验、国防领域的安全视角——这三重身份的交叠，使他有能力驾驭这一横跨三大前沿技术的宏大议题，而不至于沦为某一领域的“技术布道者”。

在拉丹利耶夫的论述中，AI 智能体、区块链与量子计算构成了一个相互赋能的“不可能三角”：量子

---

**作者简介：**夏凯捷（2004—），女，广东金融学院国际教育学院本科生，金融学专业。

计算是引擎，为 AI 智能体注入超凡算力，突破经典计算的瓶颈；AI 智能体是大脑，实现技术的自主落地与价值释放，推动 AI 从“Chat”走向“Act”；区块链是信任基石，确保决策透明、可追溯、不可篡改，解决 AI “黑箱”带来的信任危机。三者相互赋能、相互制约，共同推动人类迈向通用人工智能（AGI）乃至超级人工智能的门槛。

## 2 从“工具”到“主体”：伊德技术现象学视野下的智能体跃迁

本书的核心洞察之一，在于揭示 AI 智能体从“工具”到“主体”的演变过程。传统的人工智能系统大多是被动执行任务的工具，而智能体则具有目标导向性、环境感知能力和自主决策能力。当这种智能体与区块链的去中心化、不可篡改特性结合，便创造出能够自主交互、建立信任关系且无需中央权威协调的分布式智能系统。

这种转变可以从美国技术哲学家唐·伊德（Don Ihde）的后现象学理论中获得深刻阐释。伊德将人与技术的关系归纳为四种基本模式：具身关系（技术融入使用者身体经验，如眼镜）、诠释关系（技术呈现世界的文本化表象，如温度计）、他者关系（技术成为准他者，如机器人）和背景关系（技术退居背景，如暖气系统）。传统 AI 系统多处于“诠释关系”或“背景关系”中，而 AI 智能体的崛起，意味着技术正日益逼近“他者关系”——它们不再是纯粹的工具，而成为与我们交互的“准主体”。

伊德的理论提示我们，技术在协调人与世界关系的过程中，不仅改变着人类活动的习惯，更使得技术形成一种“技术文化”，进而改变着人类的思维模式。人类越来越相信和依赖技术构建出来的“真实世界”，这正是拉丹利耶夫在书中反复警示的风险。当 AI 智能体具备自主决策能力，能够在区块链网络上完成复杂交易、在医疗领域做出诊断建议时，我们不得不面对伊德提出的核心追问：技术究竟是在扩展人类经验，还是在塑造一种新的“技术化的知觉”？

这种转变具有深远的哲学意涵。长期以来，西方哲学传统中“主体”概念始终与人类意识绑定：笛卡尔的“我思故我在”将主体性奠基于理性思维，康德将道德主体的核心特征归结为“自律”。当 AI 智能体开始具备自主决策能力时，我们不得不面对一个根本性的追问：“主体”的边界究竟在哪里？伊德曾区分“身体一”（现象学的活身体）与“身体二”（社会文化建构的身体），并指出技术构成贯穿两者的“第三维度”。那么，当智能体开始行动，我们是否正在见证一种“技术建构的主体”的诞生？

## 3 技术的双刃剑：量子威胁与信任危机

值得注意的是，拉丹利耶夫并未陷入技术乐观主义的迷思。相反，他以网络安全专家的敏锐，深刻揭示了技术融合背后的系统性风险。量子计算的双刃剑效应是全书的精彩章节之一。一方面，量子计算能够以前所未有的速度解决复杂问题：传统计算机需要数年才能完成的计算，量子计算机可能只需几分钟。这种算力飞跃不仅能加速 AI 智能体的训练效率，更能让其处理此前无法承载的复杂任务——从模拟新药研发到预测复杂金融市场。但另一方面，量子计算的强大算力可能破解当前主流的加密系统，对 AI 智能体的安全、区块链网络的完整性构成颠覆性威胁。Shor 算法能够破解 RSA 加密，Grover 算法可以加速密钥搜索——量子计算机一旦成熟，现有数字安全体系将面临崩塌。

这一担忧可以借助法兰克福学派的技术批判理论加以深化。马尔库塞在《单向度的人》中指出，现代技术社会的“单向度化”使技术控制无所不在，社会丧失批判性。当 AI 智能体与量子计算融合，这种控制可能达到前所未有的程度。当代技术批判理论的重要代表安德鲁·芬伯格（Andrew Feenberg）则进一步指出，技术的发展是一个社会斗争的舞台，它由技术标准和社会标准共同决定。这意味着，我们不能简单地将技术发展视为线性进步，而必须追问：究竟谁在设计技术？谁的利益被嵌入技术代码？

芬伯格提出“技术的代码”概念，强调技术设计中已然包含着特定的社会价值和权力关系。从这一视角审视拉丹利耶夫笔下的 AI 智能体，我们必须追问：这些即将崛起的“主体”，承载的是谁的价值观？区块链的“去中心化”是否真正意味着权力分散，还是可能造就新的中心？量子计算的“算力解放”是否会加剧数字鸿沟？

#### 4 制度之维：被低估的社会复杂性

尽管本书是一部具有深刻洞见的著作，读者仍需保持批判性思考。一个值得商榷之处在于，作者的技术乐观主义倾向可能低估了制度惯性和社会接受度的制约。

拉丹利耶夫描绘的图景令人神往：量子计算驱动的 AI 智能体在区块链信任框架内自主运行，药物研发加速、金融风控优化、供应链透明化、气候模型精确化。然而，现实中的技术发展往往比理论构想更加复杂曲折。正如已有评论者指出的，目前三大技术的成熟度差异极大——AI 相对成熟，量子计算仍处早期，其深度融合面临诸多工程挑战。更重要的是，技术的演进从来不是单纯的技术问题。

技术哲学家汉斯·阿切特胡斯倡导的“经验转向”提醒我们，技术哲学研究应当回归具体技术的经验描述，客观理解技术与社会的共同进化。制度惯性、利益格局、伦理争议、法律框架、公众接受度——这些“非技术因素”往往构成技术落地的真正瓶颈。荷兰技术哲学家彼得-保罗·维尔贝克（Peter-Paul Verbeek）将后现象学与伦理学相结合，提出“技术的道德中介作用”概念。他通过产科超声技术的案例分析表明，技术不仅调节我们对现实的感知，还深刻影响道德决策和价值判断。

以医疗领域为例，拉丹利耶夫引用 IBM 沃森系统的案例，展望量子计算加入后能够以前所未有的规模模拟分子相互作用，助力攻克疑难疾病。然而，维尔贝克的警示值得铭记：任何技术产品的存在都以一系列有形或无形的人类文化设置为前提。医疗 AI 落地面临的真正挑战往往不是技术能力，而是监管审批、责任界定、数据隐私、医生接受度等一系列社会制度问题。即便技术证明某种 AI 诊断准确率超越人类医生，谁来为误诊负责？患者是否愿意接受机器的诊断？医保体系如何纳入 AI 服务？这些问题同样复杂，却不在于本书的论述范围之内。

#### 5 结语：在技术的时代思考人性

拉丹利耶夫在书中引入了一个独特的人文元素：虚构人物“约万”从 2024 年到 2084 年的成长故事。这个跨越六十年的叙事，试图生动描绘技术演进对人类生活的深远影响。这种“硬核技术+人文叙事”的结合，恰恰揭示了本书更深层的关怀：在技术的时代，我们终究无法回避人的位置。

当代技术哲学研究已呈现多元拓展的态势，从技术本质论、技术价值论到技术责任论，学者们日益关

注技术时代的人文关怀。正如拉丹利耶夫在书中所言，应对人工智能的社会影响需要驾驭复杂的监管环境，有效的治理必须在创新与问责之间取得平衡。我们需要管理 AI 的决策过程，确保区块链应用的透明度，为量子计算的使用设定标准以防止滥用。这些都不是单纯的技术问题，而是需要社会共同参与的公共议题。

《AI 智能体的崛起》的价值，不仅在于它为读者提供了一个理解技术融合趋势的系统框架，更在于它启发我们思考技术文明未来的深层命题。当 AI 学会行动，当量子计算突破算力边界，当区块链重构信任机制——我们究竟是在构建一个更美好的世界，还是在制造更复杂的系统性风险？这本书没有给出最终的答案，但它为我们提供了追问的起点。而技术哲学的诸多理论资源，恰恰可以成为我们继续追问的得力工具。

#### 参考文献:

- [1] Ihde D. *Technology and the Lifeworld: From Garden to Earth* [M]. Bloomington: Indiana University Press, 1990.
- [2] Feenberg A. *Questioning Technology* [M]. London: Routledge, 1999.
- [3] 拉丹利耶夫 P. *AI 智能体的崛起* [M]. 董世敏, 许盛辉, 译. 北京: 中信出版集团, 2026.

# When AI Learns to Act: The Rise of Intelligent Agents from the Perspective of Technological Philosophy

*XIA Kaijie*

(Guangdong University of Finance, Guangzhou 510521, Guangdong, China)

**Abstract:** In 2026, the launch of the Chinese version of Petar Radanliev's *The Rise of AI Intelligent Agents* coincides with the technological evolution of artificial intelligence from conversational robots to intelligent collaborators, marking a historic leap of AI from a "tool" to a "quasi-subject". This paper interprets Radanliev's work in depth from the perspective of technological philosophy. It first points out that the book breaks the misunderstanding of "technological isolationism" and constructs a systematic evolution framework of AI intelligent agents, blockchain and quantum computing. Then, it uses Don Ihde's post-phenomenology to analyze the transition of AI intelligent agents from human-technical interpretation relations to otherness relations, and discusses the philosophical connotation of the emergence of "technologically constructed subjects". Furthermore, it reveals the systemic risks behind the integration of the three technologies combined with the critical theory of technology, and objectively points out the book's deficiency of underestimating institutional inertia and social complexity. Finally, it holds that the book not only provides a systematic framework for understanding the trend of technological integration, but also inspires people to reflect on the human position in the technological age and the deep propositions of the future of technological civilization.

**Keywords:** AI Intelligent Agents; Technological Philosophy; Technological Evolution