

从巨机器到超级智能：AI 政治的新展望

顾超，系北京大学政府管理学院助理教授

聚焦于“巨机器”理论，原因在于当前关于人工智能未来的两种主流论述范式及其潜在局限。当前关于人工智能发展的讨论主要呈现两种倾向：一是技术中心主义范式，主张技术所引发的问题应通过进一步的技术手段予以解决。然而，该进路在某种程度上规避了更深层次的政治性议题——许多所谓的技术挑战，实为权力结构与政治选择的体现。以价值对齐问题为例，关键不仅在于如何实现对齐，更在于谁有权定义应对齐的价值体系，以及何种对齐标准能够获得主导地位。二是伦理中心主义范式，该范式将人工智能治理的核心议题界定为算法偏见、决策透明度与问责机制等伦理失范问题。

一、超级智能：作为终极“巨机器”的 AI

在探讨超级智能的潜在风险时，公众常聚焦于其超越人类的智能水平可能带来的生存性危机。然而，从政治学视角审视，超级智能所带来的更为现实且严峻的挑战，或许在于其可能催生一种前所未有的“超级权力”——即一部分行为体借助超级智能的系统性能力，实现对绝大多数人的新型支配。这一判断可从历史社会学家刘易斯·芒福德所提出的“巨机器”理论中获得启发。

芒福德以古埃及与美索不达米亚文明中建造金字塔的社会组织为例，指出“巨机器”本质上是一种高度刚性、由人构成的社会技术系统。其核心特征包括两个方面：其一，一个高度集权、进行单向命令发布的中枢指令系统；其二，大量被标准化、可替换、可计算且丧失自主性的“人类零件”。芒福德进一步指出，二十世纪中期的核武器系统可视为巨机器的最后一个工业时代原型。而在当代，超级智能则极有可能演化为巨机器的终极形态：超级智能算法有望成为前所未有的高效、理性指令中心，而其依赖深度学习所固有的“黑箱”特性，则加剧了系统的不透明性与问责困境。在此情境下，人类面临双重意义上的“零件化”：既作为数据原料的供给者，供养超级智能系统的运行；又作为算法模型的塑造对象，其行为与认知被系统性地引导与规训。其后果是，当超级智能以其超凡效率执着于某一单一目标时，极易与人类社会的多元价值、整体利益及长远福祉产生根本性冲突。无论最终呈现为“机器统治人”抑或“人通过机器统治人”，其本质皆是工具理性压倒价值理性，人类被迫适应并服从于一种非人性的、冰冷计算的逻辑。

二、权力五边形的当代特征

芒福德曾以权利（Rights）、生产力（Productivity）、利润（Profit）、政体（Polity）与公信力（Publicity），即“5P”要素来描绘现代巨机器的动力机制。在此基础上，本文尝试提出一个由控制（Control）、计算（Computation）、资本（Capital）、共识（Consensus）与文明（Civilization），即“5C”要素来构成的框架，用以解析超级智能系统作为新型巨机器的运作逻辑：

控制 (Control)：权力形态从传统的显性暴力与命令，转向通过超级智能实现的精准预测与自动化控制。这种权力机制高度弥散、渗透至社会微观层面，与福柯所描述的微观物理学权力具有内在契合性。

计算 (Computation)：计算能力成为核心的权力基础与生产方式。世界万物被持续地数据化，并服从于单一的“优化”逻辑，导致社会发展的多元目标被简化为对效率的极致追求。

资本 (Capital)：资本与计算深度绑定，形成赢家通吃的格局。用户数据成为关键生产资料，训练尖端 AI 模型所需的巨大算力构成难以逾越的资本壁垒。资本的终极目的不仅在于剩余价值，更在于垄断未来的决策权与控制权本身。

共识 (Consensus)：社会共识的形成机制发生转变。共识不再主要源于公共领域的理性交往与辩论，而是日益被算法被动塑造甚至主动生成。信息茧房与定制化叙事的精准传播，使得公众舆论与集体情感面临被系统性操控的风险。

文明 (Civilization)：人类文明面临根本性抉择：是走向芒福德所警示的、由超级智能主导的“巨机器文明”，即一个物质丰裕但社会结构僵化、人类精神生活被算法编排的境地；还是迈向真正人机共生、以人为本的文明形态，使技术服务于人类文明的全面繁荣与精神充盈。无论何种未来图景将成为现实，人的价值、尊严及其独特性如何在此新语境下被界定，已成为一个无法回避的紧迫议题。

三、制衡之道：警惕超级权力，而非超级智能

为应对上述挑战，单纯依靠技术方案或伦理指南显然不足，核心在于构建一套能够有效分解、监督与制约此种超级权力的政治与法律体系。可能的治理路径应包括以下四个层面：

技术层：推动关键 AI 模型向独立公共机构有限度开源，并建立强制性的常态化算法审计制度，以增强系统透明度与可问责性。

数据层：通过明确个人数据产权与收益权，探索“数据信托”等治理模式，强化个体对其数据的控制权，从而动摇数据垄断的权力基础。

治理层：建立融入民主参与机制的多元共治监管框架，例如引入公民陪审团等民主审议程序，确保社会多元价值判断能够被嵌入技术系统的设计与规制过程。

经济层：推行利益共享与所有权改革，例如开征 AI 特别税、设立“数据红利基金”，并探索关键 AI 基础设施的公共所有權模式，以瓦解权力过度集中的经济基础。

四、结语

综上所述，巨机器理论为理解超级智能的社会政治内涵提供了富有启发的视角。将政治学维度引入人工智能治理的核心讨论至关重要。未来的走向并非由技术本身预先决定，技术进步在本质上亦是权利与权力结构的重塑过程。当前的选择正考验人类社会的政治智慧与行动勇气，在此过程中，人文社会科学家无疑承担着关键性的批判、阐释与引导责任。最终目标应是超越技术治理的工具性思维，迈向一个真正捍卫人的主体性与尊严的人机共生文明。