

关于数据要素学科建设的深度思考

周傲英

华东师范大学数据科学与工程学院，上海 200062

摘要

数据作为第五生产要素是重大的理论突破。数据生产要素与土地、资本等传统生产要素相比，数据具有非稀缺性，数据越用越多、越用越有价值。当前的人工智能是数据智能，是联结主义的胜利，本质上是数据驱动的智能的自动化。数据智能热意味着整体论作为科学研究方法论的复兴，新的科学发现将不仅仅局限于还原论范式。数据是互联网的底层逻辑，党管互联网落实在操作层面就是党管数据。国家数据局成立两年来，在数据基础制度和数据基础设施等方面开展了全面的探索，迫切需要学术界和科技界提供理论和技术支持。开展数据要素学科自主知识体系构建的探索是当前我们义不容辞的责任。

关键词

数据生产要素；人工智能；科学研究方法论；学科知识体系；数字化转型；人才培养

中图分类号：

doi:10.11959/j.issn.2096-0271.2026007

Deep thinking on the discipline of data as productive factor

ZHOU Aoying

School of Data Science and Engineering, East China Normal University, Shanghai 200062

Abstract

The recognition of data as the fifth factor of production represents a significant theoretical breakthrough. Unlike traditional factors of production such as land and capital, data is non-scarce; the more it is used, the more it accumulates and the greater its value becomes. Current artificial intelligence is essentially data intelligence, marking the triumph of connectionism, and is fundamentally the automation of intelligence powered by data. The fervor around data intelligence signals a revival of holism as a scientific research methodology, indicating that new scientific discoveries will no longer be confined to the paradigm of reductionism. Data underpins the logic of the internet, and the CPC's oversight of the internet, when implemented at the operational level, translates to the CPC's oversight of data. Over the two years since its establishment, the National Data Administration has conducted comprehensive explorations in areas such as fundamental data systems and data infrastructure, creating an urgent need for theoretical and technical support from academia and the scientific community. It is an unequivocal responsibility for us to develop a novel knowledge system for the study of data as a factor of production.

Key words

data productive factor, artificial intelligence, Scientific research methodology, disciplinary knowledge system, digital transformation, talent cultivation

又到金秋，又到 926。十二年前的今天，也就是 2013 年的 9 月 26 日，华东师范大学成立数据科学与工程研究院，正式开始了数据科学与工程作为独立学科方向的研究生人才培养。时光如白驹过隙，蓦然回首，离华东师范大学开展数据学科建设艰难探索已经过去了十二年，明年的今天将是数据学院成立十周年纪念日。每年的这个日子，我都会感慨万千，今年尤其，不仅是因为高速运转的年轮，更是因为我们所处的时代。在学校和社会的大力支持下，数据学院走过了“筚路蓝缕，以启山林”的艰难旅程。有幸的是，我们探索数据学科建设的十二年恰好是人们对数据的重要性认识突破的十二年。当年人们只是用“大数据”^[1]这个喜闻乐见的说法来说明数据很重要，后来数据成了第五生产要素。再加上，我们一直把“Data is Power”作为学院的口号，认为数据是和电一样的动力。

在今天这个具有特殊意义的日子里，回望走过的路，展望未来的路，很有必要。我一直认为“大数据”是说数据很重要^[2]，数据作为第五生产要素，成立国家数据局，数据的重要性不言而喻。继大数据热之后出现的人工智能热成为新的关注点，各行各业的数字化转型和“人工智能+”行动正在全面铺开，用“如火如荼”来形容也不为过。人工智能和数字化转型，还有新质生产力、高质量发展等和数据有关系吗？是什么样的关系？立足我们的实践，结合学习去年召开的全国教育大会和今年中央教育工作领导小组关于《高等学校学科专业设置调整优化行动方案（2025-2027 年）》的精神，我想利用这次机会，对数据要素学科自主知识体系的构建进行一次

深度思考。

1 数据是第五生产要素

“数据要素”指的是将数据视为一种生产要素。数据被确立为第五生产要素，在我看来，这是我国对人类的贡献，是中国共产党对这个世界时代性的贡献。数据本身并非新生事物，但为何我国会将其明确列为第五生产要素？生产要素本质上是一个政治经济学概念，在马克思主义政治经济学中尤为常见。因此，我认为将数据确立为生产要素，是马克思主义中国化在当代的一项标志性的成果，其意义深远，必将长远影响中国的发展。我们需要深入解读数据作为生产要素的内涵与特征。数据与资本、土地、劳动等传统要素不同，我们必须突破传统的认知，重新认识数据，并构建数据要素学科的自主知识体系^[3-4]。

2020 年 4 月 9 日，中共中央、国务院发布《关于构建更加完善的要素市场化配置体制机制的意见》，首次明确将数据列为第五生产要素。在传统政治经济学理论中，生产要素长期限于三类——劳动力、土地和资本。这一理论框架可追溯至 1860 年代马克思在《资本论》中的系统阐述。从那时起，直至 2020 年，在整整 160 年的时间里，要素体系基本未有结构性扩充。也正因如此，从第三生产要素“资本”到第五生产要素“数据”的跨越，不仅经历了漫长的一个多世纪，更代表了一次重大的理论突破。这一突破，为我国数字化转型与数字经济发展提供了重要的理论支撑与指导思想。基于这样的认识，可以对不同经济形态下的生产要素作了一个简要划分：

农业时代的主要生产要素是土地与劳动；第一次工业革命后，马克思主义政治经济学将资本确立为第三大生产要素；而随着2020年数据被列为第五大生产要素，“企业家才能”也被追溯为第四大生产要素。那么，为什么在第二次工业革命时期，“企业家才能”显得尤为重要？支撑第一次工业革命的科学主要建立在牛顿力学与欧几里得几何等经典理论之上，而第二次工业革命则以麦克斯韦电磁理论等为核心科学突破。电力的出现与应用超出了当时人们的普遍认知范畴，正是在这样的背景下，企业家的想象力、远见与坚韧毅力成为推动技术落地与产业变革的关键力量，“企业家才能”也因此被“追认”为第四生产要素。

2 对生产要素的解读

何为生产要素？在我看来，它必须能够创造价值，并参与价值分配。从历史现象观察，越是后期出现的生产要素，其创造的价值体量往往越大，在分配中所占的份额也越高。在农业时代，土地是核心要素，占有重要地位，因而产生了地主阶级；进入工业时代，资本成为关键要素，资本家居于主导地位，分配占比高，进而形成资产阶级与无产阶级的对立。1860年马克思的《资本论》深刻地剖析了资本这一生产要素，成为“千年第一思想家”，为“全人类的解放提供了理论武器”。我们是在马克思主义指导下建党建国的国家，实践表明，围绕生产要素所建立的制度及相应的基础设施至关重要，而且是一个艰难的探索过程。

以马克思主义为指导思想的中国共产党对资本这一第三生产要素有着深刻的认

识，在新中国社会主义发展实践中，对如何发挥资本的积极作用也有一个艰难的过程。2022年4月29日，时任中国人民大学副校长、现任上海财经大学校长的刘元春同志在第十九届中央政治局第38次集体学习中，专门就“依法规范和引导我国资本健康发展，发挥资本作为重要生产要素的积极作用”的发言。这一话题意味着，资本会带来消极作用，也就是说，在如何有效“驾驭资本”这一问题上，我们仍面临诸多挑战。2021年建党百年之际，吉林省原省委副书记林炎志同志撰文《不代表任何利益集团的中国共产党，如何驾驭资本？》，深刻提出一个现实命题：在资本主义国家，资本背后是财阀与资本家集团，执政者代表其利益，驾驭资本顺理成章；而中国共产党不代表任何利益集团，又该如何驾驭资本？由此引申至数据领域：我们尚未形成系统的“数据论”，又该如何驾驭数据？我们党不代表任何数据公司或互联网企业，又该如何引领数据要素的发展？

在我国，最高级别的文件已经明确把数据确立为第五生产要素，这是认知和理论上的重大突破。有了指导思想层面的突破，接下来的重要的任务就是实践层面的落实，因此，“驾驭数据”就成为中国共产党人必须面对的时代命题。正如我们在历史上逐步认识和驾驭资本一样——从共和国早期的“割资本主义尾巴”，到改革开放时期的释放资本活力，再到2022年刘元春同志强调发挥资本的积极作用，说明我们始终在探索中前行。今天，百度、阿里、腾讯这样的互联网企业之所以能较好驾驭数据，源于其明确的利益驱动。而国家数据局、上海数据集团这样的国家行政和国有机构，它们不代表任何利益集团，如何才能履行好数据要素的管理、运营与治理职责？这正是我们这个时代必须回答的重

大命题。

3 数据生产要素的特点

数据作为生产要素和传统的生产要素有共同的特点，也有很大的差异。对数据最朴素的理解，就是“以数有据”——作为决策或执行的依据。在传统认知中，数据往往是业务活动的“副产品”：银行留存用户信息是为了结算，运营商记录通话数据是为了计费。作为依据的数据本身不是目的，它是为目的服务的，更像生产过程的“排放”，所以，这样的数据应该是“越少越好”，过度留痕可能被视为形式主义甚至面临问责。然而，当数据被确立为第五生产要素时，其逻辑发生了根本性转变。数据如同土地和资本那样的生产要素，生产要素当然是“越多越好”。作为依据的数据和作为生产要素的数据，尽管“数据”二字未变，但对于“数据”的理解有云泥之差、天壤之别。这是为什么我一直要求我们的团队，必须谦卑地对待“数据”，要谦卑到尘埃之中，因为我们不理解“数据”。

我们所讨论的数据，是存储在计算机中的，是“人对世界认知的结果在计算机中的表示”。这一定义虽长，却揭示了一个基本事实：数据具有人民属性，数据来源于人，数据为了人（Data from Human, Data for Human）。在大数据兴起之初，也就是十二年前我们开始建设数据学科的时候，我们的基本观点是：“大数据”是说数据很重要，就像“大调研”是说调查研究很重要一样，主要不是指规模的大小。当时对大数据的另外一个解读：大数据是指数据的采集、汇聚、关联和使用，涉及数据的生命周期。到了现在，我们才理解，

数据的应用是这个周期里的最后一个环节，但却是最重要的。先有应用，然后才有数据，不是，先有数据再有应用。人对世界的认知是全面的，如果要把人的认知投射到计算机中去，一定得先问问目的：为什么要采集这些数据？因为数据需要成本——数据采集、数据存储需要设备，会产生能耗。因此，必定是先有应用需求，而后才产生有目的的数据积累，而非相反。这正是我们认知上的一个根本转变：应用先于数据，而非数据先于应用。在现实中，通常的情况是已经有了大量的数据，希望我们这些研究数据技术的人帮他们把数据用起来，这给了我们超越了技术本身的挑战，那就是做“逆向工程”，去弄明白收集这些数据的目的，然后才能用起来。

数据来源于人，数据为了人，这是数据具有人民属性最根本的原因。数据是人对世界认知的结果在计算机中的表示，没有人就没有数据。五大生产要素中，劳动是人与生俱来的，土地和资本是身外之物，企业家才能和数据这两个新晋的生产要素也是与人密切相关的。数据的人民属性回答了为什么世界上只有中国共产党、只是在中国，数据才会被明确为第五生产要素。此外，数据作为第五生产要素还有一个区别于土地、资本等传统要素的突出特点：非竞争性。数据越用越多，数据越用越有价值。这和传统生产要素的排他性和独占性截然不同，没有了稀缺性，传统的与经济学和市场相关的理论就不再适用。正因为如此，建立数据产权制度等基础制度显得尤为紧迫，我们必须明确数据的权属问题，必须理解数据的人民属性。

4 数据是电力一样的动力

数据是当前人工智能的前提条件。没有数据，就没有当下的人工智能浪潮。早期的人们耳熟能详的人工智能，像波士顿动力的机器人和IBM的百科知识抢答赛冠军 Watson，都是基于逻辑或控制论等数学基础研发的系统。目前当红的人工智能，如 AlphaGo、ChatGPT 等，它们和以前的人工智能迥然不同，它们属于联结主义的流派，是数据驱动的人工智能。关于“数字化”的另一种说法是现在流行的“数智化”，在我看来“数智化”指的就是数据驱动的智能的自动化。数字化转型的核心要义就是自动化。正如 John McCarthy 在 1966 年所言：“人工智能就是智能的自动化”。数据是人的经验，深度神经网络是人造大脑，机器学习则是让机器学习人的经验。所以说，当前的人工智能本质上深度学习的胜利，深度学习的本质是“机器经验主义”。

数据驱动的人工智能其实可追溯至李飞飞发起的 ImageNet。她作为理论物理出生的学者，从追求极致数学模型的“爱因斯坦范式”，转向了大规模收集数据、通过竞赛推动科学发现的“开普勒范式”。这看似是科学方法的一次“倒退”，实则是一种螺旋式上升^[5]。图灵奖得主 Jim Gray 提出的科学研究“第四范式”——数据范式，标志着科学正重新拥抱整体论。数据不再是简单的物理实验记录，它是人对世界的认知，当我们的信息系统有能力收集大量数据的时候，科学发现就进入了一个崭新的阶段，数据密集型的科学发现。第四范式是互联网和大数据时代的开普勒范式，直觉、怀疑、想象力与勇气变得至关重要。李飞飞的直觉让她走向了理论物理的反面，利用数据成就了自己。她利用互联网收集数据，靠“众包”成就了她自己当代第谷的角色；又利用互联网发起竞赛激发全球

智慧，靠“众智”成就了她“半个开普勒”的角色；而 Hinton 等人的算法突破，则补全了另外半个“开普勒”的角色。回顾科学史，现代科学的诞生正源于牛顿为解释开普勒从大量天文数据中总结的行星定律，从而发现了万有引力并发明了微积分。前牛顿时代实质是开普勒的数据范式时代。如今，我们仿佛正回归那个数据驱动的发现时代。这一轮科技革命呈现出一个鲜明特征：技术开始倒逼科学。人们已经做出了 AlphaGo，但对围棋的深层原理并未完全理解；人们已经做出了 ChatGPT，可认知科学并未取得同等突破。这不再是“科学理论引领技术进步”的线性模式（如爱因斯坦的 $E=MC^2$ 指引奥本海默造出原子弹），而是技术在倒逼科学。数据，正是驱动这场变革的关键所在。

而我们今天所见的这一切技术成果是如何诞生的？AlphaGo 源于人类下棋的过程数据，是无数对弈经验训练了卷积神经网络 CNN；ChatGPT 则源自人类写作的结果数据，是互联网上海量的文本数据训练了 GPT，从而塑造了它的能力。它们的本质是基于人类经验的，而经验主义有其不牢靠之处，正因如此，我们更需要呼唤理性主义的回归——呼唤一种新的科学理性主义，那将是一个全新的科学时代。这不仅仅是“AI for Science”，更大程度上是 AI 本身预示着一个崭新科学时代的到来。自成功呼唤来“赛先生”以来，我们也曾一直以为是科学引领技术。仔细回顾一下人类科学技术发展的历史，我们可以发现第一次工业革命的五位代表性人物没有一个接受过正规的学校教育，也就是说，他们都没有系统地学习过牛顿一百多年前已经发现的科学理论。他们在各自领域做出巨大贡献，功成名就获封爵位后，走进大学时才发现牛顿的科学已经在大学教授

了一个世纪，而这些科学能够解释技术发明的本质，并为技术发展指明方向。人类真正进入科技文明，并非在17世纪牛顿开启现代科学之时，而是在工业革命蓬勃兴起的18世纪末至19世纪初。这说明，早期并非科学单方面引领技术，直到后来，科学与技术才逐渐形成相辅相成相互加持的关系。

数据带来人工智能，数据催生新的科学，数据成为第五生产要素。数据如此重要，远远超越我们技术层面的理解。那我们到底该如何理解“数据”这一重要概念？理解一个新概念或新事物的最好办法就是类比或对比，就像找同义词和反义词一样。我们成立数据学院时就确定的标语 Data is Power，就是一种类比：数据是像电一样的动力。我们说Power，如果不加任何说明，一般指的就是电，电力出现以前就有汽力、马力，英文表述中都有Power (Steam Power, Horse Power)。数据和汽力和电力一样，是一种新型的“动力(Power)”。人类文明史上，每一次“动力”形态的更迭，都催生了新的生产力。如今我们强调“新质生产力”，在我看来，其本质正是这种动力革命带来的数量级跃升：一匹马的力量可能是人的十倍，一台手扶拖拉机的牵引力又远超十匹马，一列高铁的速度可以是传统火车的十倍；而我们借助数据，通过海康威视、旷视的摄像头，在追捕逃犯等方面可能比上百名警察更高效。正因为动力更替带来生产力数量级的提升，世界的中心也随之转移：农耕文明时代，中国在世界中央一千多年；蒸汽时代，英国成为日不落帝国；电气时代，美国崛起为世界中心。如今，为何中国正迈向世界舞台中央？为何美国显得焦虑失态？我想，是因为他们感受到某种历史的宿命——历史正在重演。

从参与其中的“新工科”建设实践中，我们也能体会数据是新的动力，记得是在2017年大年初六，我参加一个座谈会，讨论我国新工科的发展，那一次就深切地感受到信息技术和数据在新工科建设中的重要地位。2019年2月我访问斯坦福大学计算机系就坚定了这个认知^[6]。在斯坦福计算机系的大楼底层，它们的展览更清晰地说明：人类的工程学科始于土木工程，依赖的是马力；机械工程依托蒸汽动力；电气工程则建立在电力之上。而在人工智能时代的今天，我们正迈入“数据工程”的新纪元。发展数据学科，建设数据学院，恰如当年创办电机系，我们正站在新动力的起点，塑造着下一代生产力的核心。

5 数据要素学科的知识体系

数据远非我们迄今为止理解的那样，我们必须重新理解并构建关于数据的知识体系。今天，我们正处在一个颠覆和创新的年代，不能等着被颠覆式技术来颠覆，而是要创新性地发展出颠覆式的技术。颠覆式技术来自于颠覆式研究，颠覆式研究来自于颠覆式实践，而颠覆式实践又来自于颠覆式思想，正印证了孙中山先生说的“知难行易”。这意味着，我们必须从根本上改变认知，必须构建出关于数据要素的知识体系。这是我近年来愈发深刻的体会：唯有认知和思想上的率先突破，才有可能催生真正的颠覆性技术。

“Data is Power”，这是我们在2013年启动数据学科建设时提出的口号。数据是电，那么人工智能就像是这个时代的“马达”或“电机”。数据之于数字化，正如电力之于电气化。农业经济时代，我们依赖马力；工业经济时代，我们依靠蒸汽

与电力；而数字经济的根本标志，就是把数据作为新的动力。从这个意义上来说，数字化并非“弯道超车”，更不是“遥遥领先”，而是“变轨换道”和“开辟新赛道”。“弯道超车”和“遥遥领先”，本质上是中国在长期落后环境中形成的思维定式，不客气地说，就是以行动上的勤奋掩盖思想上的懒惰。

唯有具备颠覆性思想，才可能孕育颠覆性实践，进而催生颠覆性技术。数据是这一变革的根本动力，对数据重要性认识的源头在于互联网。互联网彻底改变了人，进而改变了世界。没有计算机，就没有互联网；没有互联网，就没有大数据（我们就无法认识到数据的重要性）；没有大数据，就没有今天的人工智能。尽管人工智能的出现只比计算机晚了十年，但，却是互联网让数据的重要性真正凸显出来，数据让人工智能这一美好的愿望变成现实。不仅如此，互联网直接助力当前人工智能的诞生，李飞飞借助互联网的“众包”和“众智”开辟了人工智能的新方向。当前的“人工智能+”，正如十多年前的“互联网+”行动计划，这里的“+”不是简单的物理组合，“+”表示融合，它引发的是化学反应和生物反应，最大的特点是不确定性。“人工智能+”是“互联网+”的持续探索，理解“互联网+”才能理解当前的形势和未来的方向。“互联网+零售”颠覆了传统的零售行业，成就了阿里巴巴；“互联网+图书馆”，革新了传统的图书馆，造就了百度。

数据是互联网的底层逻辑，数据是颠覆式技术的源头。数据及其相关技术从上至下穿透整个社会结构，在上层，改变商业与经济形态；在中层，冲击伦理与法律秩序；在底层，催生新科学，改变科技发展路径。因此，构建数据要素学科的知识

体系，必须贯穿多个层面：在哲学与政治层面，要认识到数据是人与生俱来的产物——“Data from human, Data for human”，必须体现其人民属性；在法学、经济学与管理学层面，要确立“党管数据”的原则。党管互联网的根本就是党管数据，其根本目的是造福人民；在数学、统计学与计算机科学层面，则要深入理解数据本身。正如只有理解交流电，才能建设电网；只有建好电网，才能发展电动机，进而推动AI的真正落地。

6 构建数据要素学科自主知识体系的必要性

构建数据要素学科的知识体系是我国进入新的发展时期的迫切需求，也是我国提出数据要素这一重要论断和国家数据局成立之后，学术界应该主动作为的重要方向。难度在于“一体化”和“自主”两个方面。

当前我们常提到“一体化”，其本质是做一件事，不是做几件事，然后整合在一起。典型案例如Hadoop和OpenAI，它们都是“应用场景 + 科技创新 + 产业发展”一体化的成功样本。它们都是把解决具体的应用场景问题，研发创新的科学技术、发展全新的商业模式放在一起完成的。OpenAI是如此：它从解决智能问答问题出发，研发成功ChatGPT，使OpenAI自身也成长为一具有巨大商业价值的企业。同样，Google是为了解决自身搜索引擎PageRank的问题而发展出成就Hadoop的核心技术和关键理论。2014年国家提出“创新驱动发展”战略，当时Hadoop如日中天，而它正是为了应对实际的应用需求而诞生的，这让我们深切地认识到“应用驱动创新”。如今，随着ChatGPT和

AlphaGo的成功，我们更加清晰地认识到：“场景驱动应用”。从围棋对弈到智能问答，都要从一个典型的、具有代表性的应用场景切入。场景驱动应用，应用驱动创新，创新驱动发展，这样形成一个完整的驱动链条，这也是一种“一体化”的体现。

根据当前的人工智能实践可以总结出人工智能的方法学，人工智能方法学就是“一体化”的方法学。做人工智能科学研究正确的方法是，同时完成以下四件事情：一、选定一个常规方法难以实现自动化的应用场景，深入理解它；二、对理解的应用进行精妙的建模，精妙之处在于选择一种深度神经网络来模拟人的大脑；三、为训练人造大脑而把人类的经验整理成数据；四、搭建支撑训练人造大脑的计算系统和算力环境。这四件事情不是分开的，是一个团队同一批人一起做的。一体化是指围绕同一目标，在不同层面协同推进，它不是简单地将几件事拼凑在一起，真正的一体化，是做一件事。只是从不同侧面去看这一件事会呈现出不同的内容。我们常说人工智能有三大要素：数据、算法以及算力。那么数据是什么？——是人的经验；算法是什么？——是人脑的模拟器，它不是传统意义上解决应用问题的算法，而是对人脑认知机制的深度模拟，是深度神经网络的实现算法。因此，我们不应轻易抱怨人工智能公司不开放其核心代码——那是模拟人脑的代码，其中蕴含了极其精妙的设计。而所有这一切的最终目的，是实现智能的自动化。因此，在选择应用场景时，必须重点考虑哪些任务适合自动化——比如那些对人来说费力、危险或容易出错的环节。

毋庸置疑，当前的人工智能的发展意味着科技发展进入崭新的时代。正如我们钱旭红校长曾说，科学本质是“分科之

学”。人工智能发展到今天，我们看到的更多是“合”，是“一体化”。不应是只当“拿着锤子到处找钉子”的榔头派，而应该在解决问题（“钉子”）过程中发明“榔头”的“钉子派”。这或许是我们这个时代最需要完成的思维转变。从这个意义上而言，AlphaGo与ChatGPT更像是一把“示范性的榔头”——它们只是开端，而我们真正需要的，是无限种形态的“榔头”，正如我们需要千万种不同用途的电动机一样。因此，我们不仅要学会使用工具，更要理解工具背后的原理，实现对工具的深层驾驭。这也是提升我们对“君子不器”的理解，君子不器指的是要做到道器合一。

从科学方法论的角度来看，“整体论”指中国思想的底色。英国历史学家汤因比曾预言“21世纪是中国的世纪”，正是基于他对现代科学主流方法论“还原论”的悲观看法，整体论的复兴，是当前人工智能昭示的一个事实。基于还原论的现代科学创造了很多奇迹，大数据背景下的整体论正在散发越来越耀眼的光芒。从中国古代的倍九论、节气论、穴位论等古三论到1940年代的系统论、控制论、信息论等旧三论，再到1980年代耗散结构论、混沌论与突变论等新三论，整体论思维一直起着重要的作用。当然，我们并非要抛弃还原论，而是要将整体论与还原论融合使用，形成更完整的认知框架。除此之外，我们必须重新重视“直觉”的力量。爱因斯坦凭借直觉提出相对论，李飞飞依靠直觉推动ImageNet，OpenAI同样凭借直觉确定了技术方向。科学发展到今天，已越来越多地依赖直觉而非纯粹理性来指引方向，想象力变得至关重要。

辛顿近期来访中国时表达了他的忧虑：在资本主义制度下，人工智能带来的生产力飞跃可能导致大公司与富人获取绝大部

分收益，加剧贫富分化。在这一问题上，中国的社会主义制度提供了根本不同的路径。“共同富裕”与“一个都不能少”是发展的根本目标，这也为数据使用、人工智能开发与新质生产力的发展提供了制度保障，使其不至于陷入“富者愈富、贫者愈贫”的恶性循环。正因如此，西方难以构建起以人民为中心的数据要素知识体系，而中国却能够将数据确立为第五大生产要素。这并不是西方缺乏认知，而是其制度难以实现这样的目标。中国在顶层设计上清晰的、正确的。然而在操作层面，我们仍面临挑战：许多人尚未真正理解什么是数据，数据作为生产要素的根本意义，以及它将如何深刻影响科技、法律、伦理与经济发展。目前多数讨论仍停留在概念层次的讨论，未能实现跨学科的贯通。因此，我们当前最迫切的任务，就是构建起属于我们自己的、真正贯通理论与实践的数据知识体系。

站在巨人的肩膀上，当今中国年轻一代正稳步走向时代的中央。我们应当号召更多年轻人，向梁文锋与杨宇看齐——梁文锋曾在总理座谈会上掷地有声地说：“中国的人工智能不可能永远跟随，必然有人站在技术的前沿”；而杨宇说得更加彻底：“很多高山只是想象出来的，死磕到底就能翻越”。这样的信念，正是我们今天亟需的精神力量。在这个数据驱动变革的时代，构建中国自主的数据知识体系，已成为我们这一代人的使命。若将数据比作“电”，那么 AI 就是“电动机”，AI Agent 则是“电气化设备”。这一比喻启发我们回归“第一性原理”的思维方式。以家电为例：电风扇是电动机的一种具象应用，它模拟人扇扇子的动作；但空调的出现，却打破了这一思维定式——它不再模仿“扇风”，而是基于空气冷热交换的原理，直接将热

空气转化为冷空气，实现真正的高效制冷。洗衣机的发展同样如此。传统洗衣方式依赖物理揉搓，但当我们从化学层面理解污渍脱落的本质——即肥皂分子与油污发生反应，使其脱离衣物纤维——便诞生了无需搓洗的滚筒洗衣机。这不是对人工洗衣的简单模仿，而是基于分子作用原理的根本性创新。“第一性原理”最早源于亚里士多德的哲学思想，指“每一个系统中存在一个最基本的命题，它不能被违背或删除”。而今天我们所强调的第一性原理思维，更多融合了笛卡尔的理性方法论与马斯克的商业实践——即抛开现有框架与惯有假设，回归事物本质，从基本原理出发思考问题，而非盲目依赖传统经验或既定路径。这不仅是技术的进步，更是一种思维范式的跃迁——从“经验驱动”转向“原理驱动”，也正是我们构建自主数据知识体系所需要的思想根基。

7 结束语

我们要“相信相信的力量，相信数据就是动力（Believe the power of the belief that data is power）”。这是一个“因为相信，所以看见”的时代，我们不应再让传统的智慧与过往的经验束缚我们的思考。我们正站在人工智能所开启的新启蒙时代，也是人类文明史上一个新的轴心时代——各类知识都将在此被重新创造、重新定义。而这一切的根本动因，正是数据。数据来自于人，数据为了人，国家对数据的重视，彰显了数据的人民属性，这也正是中国发展数据事业所独具的优势。构建数据要素学科自主知识体系，需要我们实现认识论、方法论和实践上的多重超越与系统创新。

参考文献:

- [1] 周傲英. 感悟大数据——从数据管理和分析说起. 大数据[J], 2017, 3(2): 3-18.
ZHOU A Y. Understanding on the big data: beyond the data management and analytics. Big data research[J], 2017, 3(2): 3-18.
- [2] 李战怀, 王国仁, 周傲英. 从数据库视角解读大数据的研究进展与趋势[J]. 计算机工程与科学, 2013, 35(10): 1-11.
LI Z H, WANG G R, ZHOU A Y. Research progress and trends of big data from a database perspective[J]. Computer Engineering & Science, 2013, 35(10): 1-11.
- [3] STONEBRAKER M, CETINTEMEL U. One size fits all: 10 years later[C]// 31st IEEE International Conference on Data Engineering, April 13-17, 2015, Seoul, Korea. [S.l.:s.n.], 2015.
- [4] 周傲英, 钱卫宁, 王长波. 数据科学与工程: 大数据时代的新兴交叉学科[J]. 大数据, 2015022.
- ZHOU A Y, QIAN W N, WANG C B. Data sciences and engineering: an emerging interdisciplinary in the big data era[J]. Big Data Research, 2015022.
- [5] 美国国家科学院国家研究委员会. 海量数据分析前沿[M]. 华东师范大学数据科学与工程研究院, 译. 北京: 清华大学出版社, 2015.
National Research Council of the National Academies. Frontiers in massive data analysis[M]. Translated by Data Science and Engineering Research Institute of East China Normal University. Beijing: Tsinghua University Press-Press, 2015.
- [6] 奚春雁. 加快推进计算机科学教育 做数据科学教育的探路人——华东师范大学副校长周傲英教授专访[J]. 计算机教育, 2020(1):7.
XI C Y. Accelerating the advancement of computer science education: leading the way in data science education — Exclusive interview with professor Zhou Aoying, Vice President of East China Normal University [J]Computer Education, 2020(1):7.

作者简介



周傲英 (1965—), 男, 博士, 华东师范大学数据科学与工程学院教授, 主要研究方向为数据库、数据管理、区块链、数字化转型、金融科技、教育科技等。

收稿日期: XXXX-XX-XX

通信作者: 周傲英, ayzhou@dase.ecnu.edu.cn

基金项目: 国家自然科学基金国际(地区)合作与交流项目(No.62461146205)

Foundation Items: