

# 透视人工智能的三大基石

## 数据：共谱数智未来

林俊杰 罗钦芳

作为数字经济时代关键的新型生产要素和创新要素，数据已融入生产、消费、流通、分配和社会服务管理等各个环节，是人工智能发展的重要保障与动力。近日，国家数据局等十七部门印发了《“数据要素×”三年行动计划（2024—2026年）》，旨在充分发挥数据要素乘数效应，赋能经济社会发展。

### 数据是人工智能发展的基石

数据要素是数字化、网络化、智能化的基础，是人工智能发展的基石。近年来，大数据相关技术、产品、应用和标准的快速发展，为人类识别复杂系统提供了全新的思维方式，探知手段和决策范式，进一步推动了人工智能的创新。

首先，人工智能的发展，依赖于高质量数据的供给和对海量数据的标注与学习。大数据为深度学习等人工智能算法提供了庞大的样本空间，人工智能在不同场景的大规模应用，都需要基于海量数据进行标注、学习和训练，从中发现规律、获取信息、作出决策。

其次，数据为人工智能的发展提供了更加广泛多样的信息来源。一方面，数字技术的发展使得数据多样性成为常态，呈现出多源异构和富媒体特点；另一方面，大数据环境下，支撑人工智能进行管理决策的信息从领域内部扩展到跨域环境。结构化、半结构化以及非结构化的跨领域数据使得人工智能技术和应用可以从多方面多角度获取多模态信息，提升其对复杂世界的认知能力。

最后，数据不仅是人工智能模型的输入原料，同时也为模型的持续创新提供推动力，发挥着创新要素的作用。一方面，基于多源异构的大数据，模型开发者与应用者能够不断对现有模型进行迭代优化，创新算法模型；另一方面，丰富的数据集有助于提高模型的泛化能力，使其在面对新信息

时也能够进行有效的学习与识别。

### 我国大数据的发展现状

党的十八大以来，在以习近平总书记为核心的党中央坚强领导下，我国高度重视大数据等数字技术的发展。通过全社会的共同努力，大数据相关技术与实践应用取得了重要进展。

数字基础设施规模级大幅跃升。根据最新数据，截至2023年底，我国累计建成5G基站337.7万个。截至2022年底，我国移动互联网终端用户数达18.45亿户，是全球首个实现“物超人”的主要经济体；在用数据中心机架总规模超过650万标准机架，近5年年均增速超过30%；在用数据中心算力总规模超180EFLOPS，位居世界第二。

数据资源规模持续增长。2022年，全国数据产量达8.12ZB，全球占比10.5%，位居世界第二；累积数据存储量达724.5EB，同比增长21.2%；全国一体化政务数据共享枢纽发布各类数据资源1.5万类，累计支撑共享调用超过5000亿次。据测算，到2025年全球数据规模将达到175ZB，其中我国数据规模将达到48.6ZB，成为全球最大的数据圈。

数据交易快速增长。2022年，我国数据交易行业市场规模达876.8亿元，同比增长42%，占全球数据交易市场交易规模的13.4%，占亚洲数据交易市场规模的66.5%。预计到2025年，中国数据交易行业市场规模有望达到2046亿元，到2030年有望达到5155.9亿元。

### 数据赋能人工智能的问题与对策

“十四五”时期，我国数字经济转向深化应用、规范发展、普惠共享的新阶段。近年来，人工智能与大数据技术和产业体系日渐成熟，然而，数据赋能人工智能的过程中，依然面临一些问题和挑战，

亟需突破。

一是数据供给质量不高。一方面，尽管我国数据资源丰富，但真正开放、共享和有效使用的数据量仍然偏低。数据“只生产不外流”的现状导致了大量“数据孤岛”的产生，制约了数据的高质量供给；另一方面，数据资源在不断加速累积的同时，价值数据的密度也不断降低，影响了数据的高质量供给。

要加快公共服务领域数据集中和共享，推进同企业积累的社会数据进行平台对接，优化数据供给结构；要秉持公共数据取之于民、用之于民的原则，加速推进公共数据分类分级授权使用，打破“数据孤岛”，强化数据要素的高质量供给，发挥公共数据在数据要素开发利用中的基础性、引领性、示范性作用；要加速探索开展数据质量标准化体系建设，推动数据要素供给调整优化，提高数据要素供给数量和质量。

二是数据产权界定存在困难。数据作为新型生产要素，具有无形性、非消耗性、易复制等特点，对传统产权制度提出了新的挑战。在数据生产、流通、使用等过程中，不同主体对数据有着不同利益诉求，且呈现复杂共生、相互依存、动态变化等特点，传统权利制度框架难以突破数据产权的困境。

要以解决市场主体遇到的实际问题为导向，通过建立数据产权管理法律法规，细化数据资源持有、数据加工使用权和数据产品经营权的“三权分置”框架，创新数据产权观念，淡化数据所有权，强调数据使用权，加速推动构建完备的数据产权制度体系，充分释放数据要素价值。

三是数据流通交易机制不畅。首先，当前数据交易缺少统一的定价和评估机制，数据流通依靠点对点单独交易，信息不对称致使数据流通的透明度较低；其次，对于不同行业、组织、设备而言，数据标准与接口难以统一，数据流通整合的可操作性较弱；再次，尽管我国已逐渐形成具有中国

特色的数据保障制度体系，但仍缺乏更加明确、具有针对性的政策法规。

要加速统筹构建数据交易场所，坚持场内集中交易与场外分散交易相结合，形成多层次、多元化的市场交易体系；要加快推进数据采集和接口标准化，加强异构数据互联互通容器技术，为不同主体设备之间的数据流通提供更加可信可靠的技术支持；要聚焦业务需求，结合具体场景需求，鼓励行业内企业与地方探索创新模式，制定更加详实的数据流通规则与标准，促进数据整合互通和互操作，推动人工智能在跨行业多领域的应用。

四是数据治理体系有待进一步完善。首先，条块分割的行业和传统的属地治理模式难以适应数据要素跨地区、跨行业、跨层级流通交易的治理需求；其次，数据的生成与使用通常涉及多方主体（例如买方、卖方、平台），数据治理责任与流程确认变得更加复杂；再次，数据安全与隐私保护问题随着数据规模的增大以及人工智能技术的发展变得更加突出；最后，庞大的数据量与多样的数据类型对支撑数据治理的技术提出了更高的要求。

要加强数据治理前瞻布局，站在数据强国的战略高度，逐步健全政策顶层设计，打破地区、行业、层级的壁垒，要引导基层政府、市场、社会组织与民众等数据相关主体，通过互动、协商、合作，实现对数据要素的共治治理，构建政府、企业、社会多方协同的治理模式；要贯彻总体国家安全观，加速健全完善数据分类分级、重要数据保护、风险评估、应急管理，发展好数据安全产业，为国家数据安全保障提供有力支撑；鼓励产学研用多方主体创新，加快推进数据可信流通、安全保障等核心技术攻关，推动隐私计算、量子计算、区块链等与数据治理相关的数字技术创新。

（作者分别系清华大学经济管理学院长聘副教授、清华大学经济管理学院博士后）

## 算法：挑战与治理

曹逸轩

在人工智能迅速发展的今天，算法已成为我们生活和工作中不可或缺的一部分。无论是社交媒体上的个性化推送、电商平台的商品推荐，还是智能家居、自动驾驶等“智慧场景”，算法无处不在，为我们的思考和决策提供着隐形的助力，在极大提升效率的同时，也带来了挑战。

### 算法的内涵及演变

算法是用于解决问题的一系列明确指令，使得复杂的问题可通过计算方法得到解决。目前，人工智能算法主要是指能够从数据中进行归纳总结、学习规律，进而进行预测或决策的一类算法。

数据、算法、算力是人工智能的三大支柱，是围成人工智能智力水平水桶的三块木板。形象一点来说，如果将人工智能的发展看作一个人的智力成长，数据、算法和算力三者可分别对应书本、学习方法和学习时长，只有书本质量高数量足、学习方法得当、学习时间充分，才能获得更高的智力水平。算法是人工智能研究的主要部分，其设计目的就是教会机器如何学习。

无论你是否了解，算法现已融入生活的方方面面。例如，短视频平台利用算法，通过分析你的历史行为推荐可能感兴趣的内容，金融机构利用算法进行风险评估和信贷评分，自动驾驶汽车则依赖算法处理和解析路况信息，以及刷脸支付、刷脸支付、拍照识物、自动翻译、地图导航……在

这些习以为常的生活场景中，从信息获取到加工再到应用，涉及算法的任务数不胜数。可以说，算法正在重塑我们对世界的理解方式。

人工智能概念自1956年被提出以来，经历了多次重大演化。2015年以来，人工智能研究向以神经网络为核心的深度学习方向飞速发展，支撑了目前的大部分智能应用场景。2022年以来，以OpenAI发布ChatGPT为标志，人工智能进入大模型时代，这其中的关键是通过算法设计，能够高效利用海量的无标注数据，通过大模型、大算力实现一定程度的通用智能。目前，算法技术正处于一个快速发展的阶段。在国际比较中，中国的算法研究和应用正在迅速发展，并在某些领域与国际水平保持同步或领先。但是在在大模型时代，由于国内芯片技术受限，算力可能会成为目前国内算法发展的瓶颈。

算法的价值远超越简单的技术进步，它为我们提供了理解和处理信息的全新视角，从而改善决策过程。在医疗领域，算法能够通过分析患者的历史数据来预测疾病风险，为医生提供治疗建议；在制造业中，算法通过优化生产流程，提升效率和质量，从而降低成本并提高产量；在金融领域，算法不仅用于信贷评分，还能在股市交易中进行模式识别，帮助投资者作出更明智的决策；在社会治理、科学研究、安全等领域算法都展现出巨大的应用潜力。未来，人工智能将是美国竞争的重要领域，随着技术的发展，我们将看到越

来越多的国家投资于人工智能技术，以期在全球范围内获得竞争优势。这不仅仅是一场技术竞赛，更是关乎经济、政治和社会未来走向的战略布局。因此，算法的发展、应用与治理，在科技竞争愈发激烈的今天，显得愈发重要。

### 算法的挑战与治理

尽管算法为我们带来了许多便利，但也造成了一定的社会问题。信息茧房效应就是典型例子。正如前面所提到，短视频、新闻平台会根据用户的观看历史推荐用户可能感兴趣的内容，这在方便用户获取信息的同时，也降低了用户主动进行信息搜寻、寻找新视角的欲望。用户不断观看这些推荐内容，又刺激算法继续推荐相似内容。以此循环往复，最终使用户观看的都是同质化的内容、类似的观点，无形中剥夺了用户了解更广阔世界的机会，导致个人偏见的放大，个人被束缚在自己的信息茧房里，难以全面地认识外部世界。在宏观层面，信息茧房会降低人与人之间的同理心和认同感，加剧社会的分裂。再如“大数据杀熟”，根据用户的购物历史，算法知道用户能承受更高的价格，可能会提供歧视性定价。在旅游、网约车等领域，近年来都曾有过“大数据杀熟”的相关报告。此外，国外有研究发现，由真实数据训练的算法可能存在种族偏见和歧视，包括对不同种族不同的犯罪倾向预测，对不同种族不同的人脸

识别精度等。这类问题目前在国内的研究尚不充分，但已经引起相关部门的重视。

面对算法可能带来的风险，我们既不能因噎废食，也不能视而不见。大众、算法提供方和政府三方面应该共同努力，形成应对算法挑战的合力。对于大众而言，需要了解基本的算法理念，明确自己的需求以及与算法的关系，合理利用算法为自己服务。对于算法提供方而言，需要从更长远的角度出发，提高社会责任感，将具有社会责任感的算法作为企业更高层次的价值理念。对于政府而言，需要建立相应的伦理指导原则和法律法规，要求企业加强算法的透明度，明确设立监管红线，并制定安全测试方法，以保证市场的有序发展。当然，目前的监管制度难以完全管控算法技术快速迭代的特点，因此如何制定算法监管规范，使之既不滞后于技术发展又能确保公众利益，是摆在监管部门面前的新课题。目前，我国已出台《互联网信息服务算法推荐管理规定》《生成式人工智能服务管理暂行办法》等法规加强对算法的监管，日后，还需根据技术发展的新形势，不断调整完善监管方式方法。

总之，展望未来，我们有理由相信，算法必将为世界带来更多惊喜。与此同时，只有树立算法伦理与价值导向，让其在合理、公平的轨道上健康发展，算法才能真正造福人类社会。这需要政府、企业与公众的共同努力。（作者系中国科学院计算技术研究所副研究员）

## 算力：克服AI“算力焦虑”

张培培

人工智能时代的竞争不仅是算法和应用的竞争，更是算力基础设施的竞争。在万物互联的大背景下，数据量爆炸式增长，对算力的需求达到了空前的高度。2023年2月，中共中央、国务院印发《数字中国建设整体布局规划》强调“打通数字基础设施大动脉，系统优化算力基础设施布局”；2023年10月，工信部等六部门印发《算力基础设施高质量发展行动计划》。这一系列政策都表明，数字经济已进入以算力为支撑的新阶段，加快构建算力竞争新优势势在必行。

### 算力成为新的生产力

简单说，算力是指传输、存储、处理信息数据的能力。每一次技术革命都伴随着生产要素和生产力的跨越式发展。随着人们进入数字经济时代，数据成为新的生产资料，算力也成为新的生产力。过去，发电量被视为现代经济的硬指标，衡量某个国家经济发展和文明程度的标准。如今，代表着数字化信息处理能力的算力成为支撑数字经济向纵深发展的新动能，成为衡量一个国家整体实力以及经济社会发展水平的重要指标。

今天，数字社会的运行发展，离不开强大算力的支撑，从城市到家庭，从政府到企业，算力已成为社会发展的重要动能。云计算、区块链、人工智能、量子计算等重要创新赛道，都离不开算力的推动。近年来不断成熟的新能源电动车，也已不单纯是一种动力设备，更是一种算力驱动的信息产品。未来，随着智能化时代的到来，更多智能产品诞生，大量的终端设备会产生海量数据，每

个生产生活场景都离不开算力对信息的处理，算力将无处不在，而算力服务也将会像水、电一样，成为一种社会基础设施。

### 人工智能时代的“算力焦虑”

过去几十年，在摩尔定律推动下，芯片算力以每18个月增加一倍的速度快速前进。即便如此，近年来，由于人工智能技术不断突破并在更大范围普及，催生了算力需求的爆发式增长，而算力供给短期难以跟上需求节奏，产生了“算力焦虑”。

人工智能发展消耗算力的根本原因在于，其改变了解决问题的基本范式。计算机不能自主产生知识，但可以通过大量的计算来获得数据背后的统计规律。机器对已有数据、知识进行学习，通过大量训练样本，不断修正模型，利用模型在相似情景下作出决定或预测。像ChatGPT这样的工具，都是以大数据+大计算量为基础。也就是说，只要算力足够强大，很多问题都可以转化为计算问题。以前我们认为机器难以解决的很多问题，比如自动驾驶、语音图像识别、内容创作等，现在靠人工智能、依托强大算力，都可以实现。

根据OpenAI发布的报告，自2012年以来AI训练任务所运用的算力每3—4个月就会翻一番；2012—2018年，AI算力需求增长了30万倍。另有研究预测，从2018年到2030年，智能出行对算力的需求将增长390倍，智慧工厂的算力需求将增长110倍。如今，人工智能的发展日新月异，每一次技术突破都需要充足的算力资源作为能量补给，大模型的快速迭代引起算力需求激增，算力供给难以追上算力需求的增长速度，“算力焦虑”成为眼下人工智能发展面临的一大难题。

个生产生活场景都离不开算力对信息的处理，算力将无处不在，而算力服务也将会像水、电一样，成为一种社会基础设施。

### 加快构筑国家算力竞争的新优势

未来，一个国家的算力安全，将与粮食安全、能源安全同等重要。由于算力建设需要持续投入巨大的人力物力财力，算力竞争很大程度上在国家之间、资本科技巨头之间展开，算力竞争已成为大国博弈的焦点。据工信部数据，截至2022年底，中国坐落的总算力已高达180EFLOPS，算力核心产业规模高达1.8万亿元，算力总规模稳居全球第二。然而，中国芯片产业也存在诸如硬件瓶颈等诸多不足，面对新形势，我们必须加快构筑国家算力竞争的新优势。

第一，加快构建高水平的算力基础设施。算力基础设施是算力稳步提升的关键，将算力真正服务于社会经济的各行各业，使其进入千家万户，需要算力供给模式的创新。2020年，国家明确提出新型基础设施建设的范围包含以数据中心、智能计算中心为代表的算力基础设施，算力供给基建化已经成为趋势。智能计算中心将成为未来算力的主要生产中心，供应中心，加快推动算力新型基础设施建设，加速高质量、高性能的智能计算中心的落地。算力的核心是服务器，而服务器的核心是芯片。芯片无疑是中国算力产业的短板。近年来，国内芯片产业迎来了资金、人才的持续投入和政策扶持，国内高科技公司对芯片核心技术也进行了深入探索，并逐步打破外部封锁，未来，中国芯片有望取得更大突破。

第二，实施“东数西算”工程，构建国家算力网络体系。正如电力发展离不开电网，算力发展也离不开“算力网络”。为了让用户享受随时随地的算力服务，形成继水网、电网之后国家新型基础设施，打造“一点接入、即取即用”的社会服

务，构建更大范围的算力网络成为正在探索中的可行道路。2021年5月，国家发改委等四部门联合发布《全国一体化大数据中心协同创新体系算力枢纽实施方案》，提出建设全国一体化国家算力网络体系。我国东部地区算力需求高，西部地区能源便宜，算力成本低，把东部的数据送到西部进行存储和计算，不仅可实现全国统一资源调配，实现更好的算力性价比，更可引导东部数据中心建设集约化发展和西部数据中心建设跨越式发展，改善数字基础设施不平衡的局面，推动算力、网络、数据、能源等协同联动。通过构建全国一体化的算力网络体系，反哺自主算力产业链，可以借鉴当年高铁、通信网络从劣势地位追赶、弯道超车的历史经验，经过统一调度和高强度研发，形成优势突破。

第三，构建商业可持续、市场可受益的算力生态圈。构筑国家算力竞争新优势，关键是把算力真正用起来，形成强大的用户集群，进入正向的技术、商业循环。因此，要在算力基础设施上提供足够的开发能力，以场景为牵引、应用为导向，开放更多典型场景，促进算力产业化应用，打造算力新业务、新模式、新业态；促进算力厂商形成成熟的商业模式和稳定的利润回报，进入商业拉动技术的正循环，促进更高强度的研发投入，获得更大的技术优势；在应用赋能重点领域打造应用标杆，提升工业、金融等领域的算力渗透率，使医疗、交通等领域应用实现规模化复制推广，进一步扩大能源、教育等领域应用范围，构建技术可信、资源可共享、商业可持续、市场可受益的算力生态圈，促进算力产业的整体崛起。

（作者单位：山东社会科学院哲学研究所）

齐鲁策论 人工智能(AI)是新一轮科技革命和产业变革的重要驱动力量，数据、算法和算力被公认为人工智能的三大核心要素。随着ChatGPT、Sora等“百模大战”的到来，人工智能越来越真切地走进人们的生活，人工智能显然已经成为国际间科技竞争的重要战场。我国正在人工智能领域加速布局，能否抓住这三大核心要素，事关我国能否发展好人工智能产业、能否抓住新一轮科技革命和产业变革机遇。就此，大众日报理论版推出专题策划，透视人工智能的三大基石，展望我国人工智能发展的前景与趋势。



扫码关注山东深观察 扫码看理论周刊专题