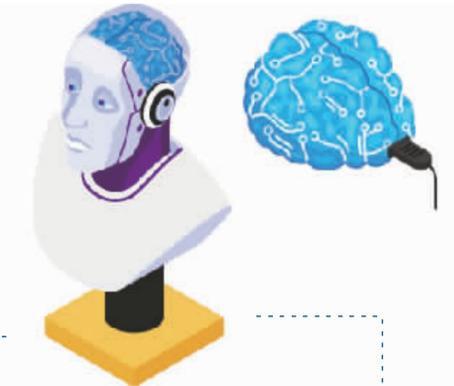


左手算力右手应用 北京冲刺“大模型”

由ChatGPT掀起的大模型浪潮仍在继续。被誉为“AI第一城”的北京，率先将重点落在了政策上。《北京市促进通用人工智能创新发展的若干措施（2023-2025年）（征求意见稿）》（以下简称《若干措施》）5月12日起向社会公开征求意见，5方面21条措施涵盖了算力、数据要素、大模型技术、应用场景等诸多受到广泛关注的领域。而在AI“新基建”领域，《若干措施》也画出了一条涵盖算力、算法以及数据训练的清晰图谱。作为“人工智能实现的基础之一”，教育、自动驾驶和医疗等领域都是大模型的应用场景，未来大模型或将改变短期内“碎片化的协助作用”，带来全流程的个性、精准化服务。



离生活渐行渐近

“为抢抓大模型发展机遇，重视通用人工智能发展，充分发挥政府引导作用和创新平台催化作用……”在《若干措施》中，首句即是这样一句话。

其中提到的“大模型”是什么？科技部国家科技专家库专家周迪向北京商报记者介绍，大模型是指具有数亿或数十亿个参数的深度神经网络模型，这种模型经过专门的训练过程，能够对大规模数据进行复杂的处理和任务处理，是人工智能实现的基础之一。

关于大模型的作用，中国民营科技实业家协会元宇宙工作委员会创办人、秘书长吴高斌告诉北京商报记者，其作用在于通过对大量数据的分析和学习，建立一个庞大的模型，使得机器能够更好地理解和处理人类语言、图像、声音等信息。“这种技术的应用范围非常广泛，包括自然语言处理、计算机视觉、语音识别等领域”。

目前，“大模型”这个词听起来似乎离消费者还很遥远，但事实上，它已在现实生活中展现过自己的“能力”。“大模型基于大数据，背后需要极大算力支持。”张孝荣向北京商报记者介绍，大模型是大规模预训练模型的简称，其中的大模型又主要指参数量大。“比如ChatGPT模型参数高达1750亿，而与之类似的模型，参数也从数百亿到数千亿不等。”张孝荣表示，“从ChatGPT开始，大模型对文字生成、文本理解、自动问答等任务有显著的改善，现在正在向图片、视频和声音等应用领域延伸。”

算力、算法与数据

数据、算力和算法，号称人工智能的“三驾马车”，缺一不可。而眼下的重点在于，这场始于AI领域的全球竞赛，却先在算力上引起了一波“焦虑”。上个月，OpenAI停止ChatGPT Plus的销售，便让人猜测是因为需求量过大，计算资源供不应求，随着国内大模型相继入局，对于算力紧张的猜测也自然蔓延至国内。

《若干措施》的第一条正发力在这一领域，在组织商业算力定向满足本市紧迫需求、高效推动新增算力基础设施建设以及高效推动新增算力基础设施建设三个方面提出了具体的举措。

具体而言，《若干措施》提到，将加强与头部公有云厂商等市场主体合作，实施算力伙伴计划，并确定首批伙伴计划成员，新增算力建设项目纳入算力伙伴计划，加快推动海淀区“北京人工智能公共算力平台”，朝阳区“北京数字经济算力中心”等项目建设，快速形成规模化先进算力供给能力，支撑千亿级参数量的大型语言模型、多模态大模型、大规模精细神经网络模拟仿真模型、脑启发神经网络等研发。

在算法方面，《若干措施》提到，要开展大模型创新算法及关键技术研究，在数据方面，则更聚焦在“高质量”这一关键词。例如针对目前大模型训练高质量中文语料占比过少，不利于中文语境表达及产业应用的问题，《若干措施》便提到，整合现有开源中文预训练数据集和高质量互联网中文数据并进行合规清洗等。

高质量的数据需要“国家队”的护航，在这一领域，《若干措施》表示，加快推动数据要素高水平开放的“国家数据基础制度先行先试示范区”建设，争创国家级数据训练基地，提升北京人工智能数据标注库规模和质等等。

“《若干措施》相当于一个非常明确的专项政策，用以统筹全市资源，更好地赋能，促进人工智能的发展，这种调集全市之力制定专门的政策规划推动一件事的情况，在全国范围内都是相对比较少见的。”北京社科院研究院王鹏如此评价道。

深度科技研究院院长张孝荣也认为，《若干措施》的亮点在于对人工智能产业三要素以及应用与监管方面都做了比较全面的部署和指导。“不过，在算法研究和基础软硬件研发方面，缺少可执行措施，有较大提升空间；在对于数据质量提升和建立统一算力调度平台方面，有些措施可以商榷，有的可以进一步完善。”张孝荣说。

立足首都带动全国

早在ChatGPT走红之前，人工智能便已成为世界科技竞争的焦点。2017年7月，国务院发布的《新一代人工智能发展规划》将人工智能提升至国家战略层面，其中提到，2030年人工智能核心产业规模要超过1万亿元，并将带动相关产业规模超过10万亿元。

数据显示，截至2022年，我国人工智能产业规模已达到5080亿元，相较于2018年的2288亿元，增长了122.05%，连续五年保持稳定增长。

首都北京的科技底色更浓。5月15日的2023中关村论坛新闻发布会介绍，2022年北京研发投入经费占比在6.5%以上，达到全国平均水平的2.7倍。在京国家实验室高质量入轨运行，全国重点实验室目前达到了80家，在全国排在第一位。2022年，北京平均每天新创办的科技型企业293家，不到5分钟就诞生一家科技型企业，独角兽企业102家，居全球第三。

中国数智融合50人论坛智库专家洪勇认为，北京人工智能产业发展拥有优质的科研与教育资源、浓厚的创新创业氛围、丰富的产业生态系统以及完善的政策支持等优势。

在他看来，《若干措施》对北京发展人工智能产业具有重要意义，它将促进创新发展，提升北京在算力资源和高质量数据方面的供给能力，推动创新技术的涌现。同时，它还将强化北京的核心竞争力，巩固其作为人工智能产业核心城市的地位。

此外，《若干措施》也将促进产业生态发展，形成完善的人工智能产业生态系统，推动合作与交流。最重要的是，它将带动经济发展，成为北京经济增长的重要引擎，促进北京在人工智能领域的引领地位。

王鹏也提到，在一系列优势的基础上，未来北京人工智能产业链布局可能会朝三个主要方向发展，首先是基础设施；其次是与数字化、城市转型升级等方面的融合；最后，北京的人工智能不仅是服务北京，还要立足北京，辐射京津冀，服务全中国甚至辐射全世界，这才是北京作为首都，拥有如此多的积累后，应该达到的目标。

“换句话说，北京的人工智能一定处于我国人工智能的第一梯队，那么北京的目标就不单单是服务中国，而是要更好地走出去。”王鹏总结称。



2030年人工智能核心产业规模要超过1万亿元，并将带动相关产业规模超过10万亿元。

北京人工智能产业发展拥有优势



浓厚的创新创业氛围

优质的科研与教育资源

丰富的产业生态系统

完善的政策支持



场景创新已在路上

作为“人工智能实现的基础之一”，大模型的场景应用创新也在不断推进。《若干措施》中提到，将推动通用人工智能技术创新场景应用，其中包括探索在政务服务、科学研究、金融、城市治理、自动驾驶和医疗等领域的示范应用。

其中，在自动驾驶领域，将支持自动驾驶企业研发多模态自动驾驶技术，发挥大型语言模型高维语义理解和泛化优势，基于车路协同数据和车辆行驶多传感器融合数据，提高自动驾驶模型多维感知和预测性能，有效解决复杂场景长尾问题，辅助提高车载自动驾驶模型泛化能力。同时，支持在北京市高级别自动驾驶示范区3.0建设中，构建车路协同数据库，引导企业开展基于真实场景的自动驾驶模型训练迭代。探索基于低时延通讯的云控自动驾驶模型测试，发展自动驾驶新技术路径。

在医疗领域，将支持北京有条件的研究型医疗机构提炼智能导诊、辅助诊断、智能治疗等场景需求，充分挖掘医学文献、医学知识图谱、医学影像等多模态医疗数据，构建基于医疗领域通用数据与专业数据的智能应用，实现对各种疾病和症状的准确识别和预测，辅助医疗机构提高疾病诊断、治疗和预防的决策水平。

“智能医疗和自动驾驶等技术的应用已经有好多年的探索了。”周迪指出，其普及将对普通群众带来重大影响。“以医疗为例，目前医生少患者多，患者排不上队挂不上号是常态。依托人工智能，医疗诊断、影像分析、治疗方案选择等都可以借助人工智能的协助，大大降低人力成本，改善患者看病难的问题。”周迪说。

动态发展中求解难题

事实上，自4月11日国家互联网信息办公室起草《生成式人工智能服务管理办法（征求意见稿）》以来，包括《若干措施》在内的各类文件接连出台，充分体现国家对推进通用人工智能大模型发展的高度重视，有望加速大模型走向千家万户。

而在大模型“飞入寻常百姓家”的过程中，推动大模型的场景应用创新仅是一个开端。由此开端向外延伸，便可看到大模型对于普通群众的实际效用。例如，吴高斌指出，在自动驾驶领域，一个显而易见的效用便是提高车辆的自动驾驶能力，减少交通事故的发生。

此外，周迪介绍，在医疗领域，大模型带来的前述各项作用仅是相对短期的效用，长期看来，大模型可能使生活中的各类服务更趋全流程的个性化。“比如要去看医生的话，中间存在很多流程，但是大模型训练好之后，就可以给出一整套个性化的就医服务方案。相对而言，目前比较着眼于在各种碎片化功能上的协助作用，未来各类服务则可能达到更加整体更加个性化的程度。”周迪表示。

“全流程的个性化”不止于医疗领域。“在教育方面，大模型可以用于个性化教育，根据学生的学习情况和需求，提供个性化的学习资源和指导。”全联并购公会信用管理委员会专家安光勇向北京商报记者表示，“它还可以用于智能辅助教育，例如智能导师或智能助教系统，提供实时的问题解答和学习建议。”

吴高斌认为，总体来说，使用大数据训练得到的大模型对于实际应用、提高工业生产力和促进社会发展具有巨大的前景。“大模型的出现为人工智能的发展提供了新的思路和方法。它的应用将会为我们的生活带来更多的便利和改变。”

不过，在张孝荣看来，教育、医疗、自动驾驶等领域虽然对AI有广阔的潜在需求，但行业数据源封闭，孤岛化严重，难以提供训练大模型所需的大数据，“因此大模型若要满足这些行业的AI需求，仍需要经过很长的时间去探索，逐步地解决在动态发展中遇到的各种行业难题”。

北京商报记者 杨月涵 冉黎黎



2022年北京研发投入经费占比在6.5%以上，达到全国平均水平的2.7倍。在京国家实验室高质量入轨运行，全国重点实验室目前达到了80家，在全国排在第一位。



6.5%

