

AI 赋能之下，上海科创如何乘风破浪

“人工智能越来越重要了，但除了大模型的日常聊天对话，科学范式的变革也正在发生。”上海市政协常委、九三学社市委副秘书长、参政议政部部长张磊说，推动人工智能驱动的科学研究的（以下简称 AI4S）在业内基本已经形成共识，我们认为还要更进一步推动人工智能驱动的工程应用（以下简称 AI4E），使人工智能能够更好为产业赋能。

“比如生物医药领域的新药研发，传统方式下，因为需要不断试验、不停试错，整个过程其实是漫长而低效的。在新的人工智能驱动下，整个研发效率将有望获得数倍的提升。”张磊说，不仅是新药研发，新材料研发、物理化学等领域都可能因为人工智能的加入，从而形成科研“第五范式”。

2023 年以来，九三学社市委进行了大量的调研，在很多论坛里也一直关注人工智能驱动的科学研究的（AI for Science）等。

2024 年上海两会期间，九三学社市委提交了一份名为“关于加快人工智能驱动的科学研究和工程应用发展的提案”，里面提到上海作为我国人工智能（以下简称 AI）发展高地，要加快推进 AI4S（科学研究）和 AI4E（工程应用）发展，是以科技创新推动产业创新的有效路径。

去年 3 月，国家科技部会同自然科学基金委启动了“人工智能驱动的科学研究的（AI for Science）”专项部署工作，紧密结合数学、物理、化学、天文等基础学科关键问题，围绕药物研发、基因研究、生物育种、新材料研发等重点领域科研需求展开，布局“人工智能驱动的科学研究的”前沿科技研发体系。

确实，过去都说上海人精明，今天如何让“AI for Science”的上海 AI，更加“精明”，很多委员代表们，也都从不同角度将目光聚焦在这一点上。

I 打造科研“第五范式”

在九三学社市委提交的这份“关于加快人工智能驱动的科学研究和工程应用发展的提案”中，提案中提到上海推进 AI4S（科学研究）和 AI4E（工程应用）发展还存在一些短板，比如前瞻布局有待加强，核心技术亟待突破；算法算力数据存在瓶颈；学科交叉融合人才稀缺。

“在人才方面，以前提到的是复合型人才，但如果要加快人工智能驱动的科学研究和工程应用发展，那需要的是 AI 和领域知识兼备的融合型人才，这是很迫切的。”张磊接受采访时说，对此，上海不仅要推进科技、教育、人才协同发展，也要强化前瞻布局和行业融合，加快关键核心技术攻关，加强数据算力等要素支撑等。

“这个需要我们一边做，一边建立生态。”张磊说，比如在强化前瞻布局和行业融合的方面，不仅需要重点部署化学与过程工程、生命科学与工程、材料与原子制造工程等；也需要支持企业开展 AI 大模型示范应用；引导企业通过 AI 赋能，构建集研发设计、原料采购、资源配置、生产制造、绿色低碳等要素为一体，需求快速感知和供需精准匹配的“智慧大脑”；同时引导头部企业开放应用场景，为 AI 训练提供高质量的行业语料库，为行业应用提供验证和优化的试验场。

“其实，政府等各方也都意识到这个问题，上海夯实基础研究，打造科研第五范式，虽然还有很多挑战，但同时也有更大的机会和空间。”张磊说。

I 探索建设智能型科学数据库

“上海近年来，花大力气建设了大科学装置、重点实验室和公共实验平台等重大科技基础“硬”设施，极大地提升了上海科创中心的影响力。然而，在支撑国家重大科技基础设施的“软”设施方面，如专业软件、科学数据库等方面进展不够，与西方发达国家相比还存在相当差距。”上海交通大学分析测试中心研究员、民进上海交通大学委员会副主委朱邦尚说。

朱邦尚平时接触的先进设备仪器比较多，通过设备仪器做一些数据分析和测试，“在工作中，我们发现，其实先进的设备仪器，背后是需要一套科学的数据库才能支撑得起来。”朱邦尚说，大数据是新资源、新技术和新理念的信息混合体，已经成为企业、社会乃至国家关注的重要战略资源，“甚至当下发展人工智能、AIGC，底层也需要科学的数据库，如此才能让数据分析的结果可靠可信。”

按朱邦尚的调研，西方发达国家已经建立了大批系统完备的科学数据库，比如总部在美国的国际衍射数据中心（ICDD），是全球唯一经过 ISO 认证的晶体学数据库，是材料学、物理学、化学、地质学、药理学、生物学、检验检疫、司法鉴定等科学研究及工业生产等领域不可缺少的数据库；比如美国国家标准与技术研究院的 NIST 谱库（National Institute of Standards and Technology）、美国斯克里普斯研究院开发的 METLIN 质谱数据库等，都在世界科技发展前沿和创新发展领域具有举足轻重的影响。

“海外的数据库，多数都是有版权的，需要付费，而且很多是有选择性的，即使花钱也不一定买得到。”朱邦尚说，上海在有些领域走在全国前列，像人工智能，但这些还不够，尤其是一些基础性的东西，比如数据库，上海要打造全球有影响力的科创中心，必须有自己的产权和架构。

“如果没有自己的科学智能数据库，只是用别人的数据库，那就好比租了别人的房子，钥匙其实属于别人的，人家随时都可以进来，这中间又涉及到信息安全等问题。”朱邦尚说。

此次上海两会期间，民进市委的集体提案“关于探索建设智能型科学数据库的建议”提到，上海探索建设智能型科学数据库，不仅需要政府层面的清晰引导和周密规划，

设立基本建设资金投入专项，也应该允许各种社会资本进入，为科学数据库注入市场经济的活力。同时，也应该立足本市高校、科研机构、现有学术数据库基础，遴选若干数据库资源基础较好的机构，率先探索智能型数据库的提升途径。而且承接率先探索任务的机构要组织，必须立足国际视野，从起步开始就要有超越现有世界水平的判断力与敏感性。

“基础性技术差距需要通过多元化途径解决，不可企望一蹴而就。”朱邦尚说。

I 打造创新“核爆点”

“早在 2019 年 1 月两会时，我们提出了上海要打造技术之都的建议；之后几年就一直在跟踪。”上海市政协副主席、民进上海市委专职副主委杨蓉说，2024 年两会，我们民进上海市委提出了关于在“浦江第一湾”打造上海技术转移核心区的建议。

“上海在推进国际科技创新中心建设的过程中，我们希望不管是发表的科技学术论文，还是一些项目成果，后续能够在技术转移核心区获得进一步转化，这不仅关系到从 0 到 1，也是从 1 到 100，让‘找技术、到第一湾’成为全球创新链中不可或缺的重要一环。”

在杨蓉看来，黄浦江在闵行区、奉贤区交界处拐了一个“L”型大弯，即提案中“浦江第一湾”的区位优势，不仅是地理层面的，还汇集了上海交通大学、华东师范大学、“大零号湾”科技创新策源功能区、紫竹高新区等独特资源，一旦上海技术转移核心区建成，“串成一条珍珠项链”，这“浦江第一湾”就可以更好承载上海南部科创中心建设的重要使命，甚至可能形成区域创新创业的“核爆点”。

确实，多区域、多维度、多领域打造“核爆点”，助力上海科创中心建设，也成为各方共同努力的方向。

2024 年 1 月 17 日，上海人工智能实验室的书生·浦语 2.0 正式对外发布，它可以根据《流浪地球 1》和《流浪地球 2》的剧情，创作出《流浪地球 3》的后续剧情脉络，也在超长上下文、数理能力和数据分析功能、工具多轮调用等方面，较几个月前有了非常大的提升。

其中关于超长上下文的处理问题，上海人工智能实验室领军科学家林达华说，实验室研究人员进行过一场试验：将关键信息随机插入一段长文本的不同位置并设置问题，测试书生·浦语的 InternLM2 模型能否从中提取出关键信息。实验结果显示，InternLM2 在上下文长度延展到 200K 时依旧保持了近乎完美的召回成功率，验证了 InternLM2 对于超长上下文坚实的支持能力，能够实现长文本中“大海捞针”。

对上海人工智能实验室而言，诞生于 2023 年上半年的书生·浦语，并不仅仅是一个大模型，背后是全球首个贯穿数据、训练、评测等环节的全链条大模型开源体系，它不仅开源了基座模型，还开源了研发大模型的整条工具链，在“授人以鱼”的基础上“授人以渔”。

“未来的发展，应该是以大模型这种以规模增长为主线的，同时又要充分吸纳其他技术路线的发展需求。”上海人工智能实验室主任助理、领军科学家乔宇认为，核心的技术路线一定在大厂，从财力、算力和数据等资源方面，大厂有优势；传统技术路线的创新，要做大量的学术内容。上海人工智能实验室，作为一个大平台，有牵引作用，可以把这两个技术路线很好的结合起来，“如果我们能实现这两个技术路线之间的资源的高效互动，我们就能创新更高效的科研范式。”

对上海而言，不管是向着“AI for Science”的科研第五范式努力，还是未来可能在建设智能型科学数据库的探索，以及人工智能全链条大模型开源体系的打造，以及未来可能的技术转移核心区的建设等等，在迈向全球具有影响力的科创中心过程中，更原始的、更基础的、更产业的、更集聚的效应，正在上海这片土地上慢慢发生，在可见的未来，我们静待繁花一路盛开。



制图 / 潘文健