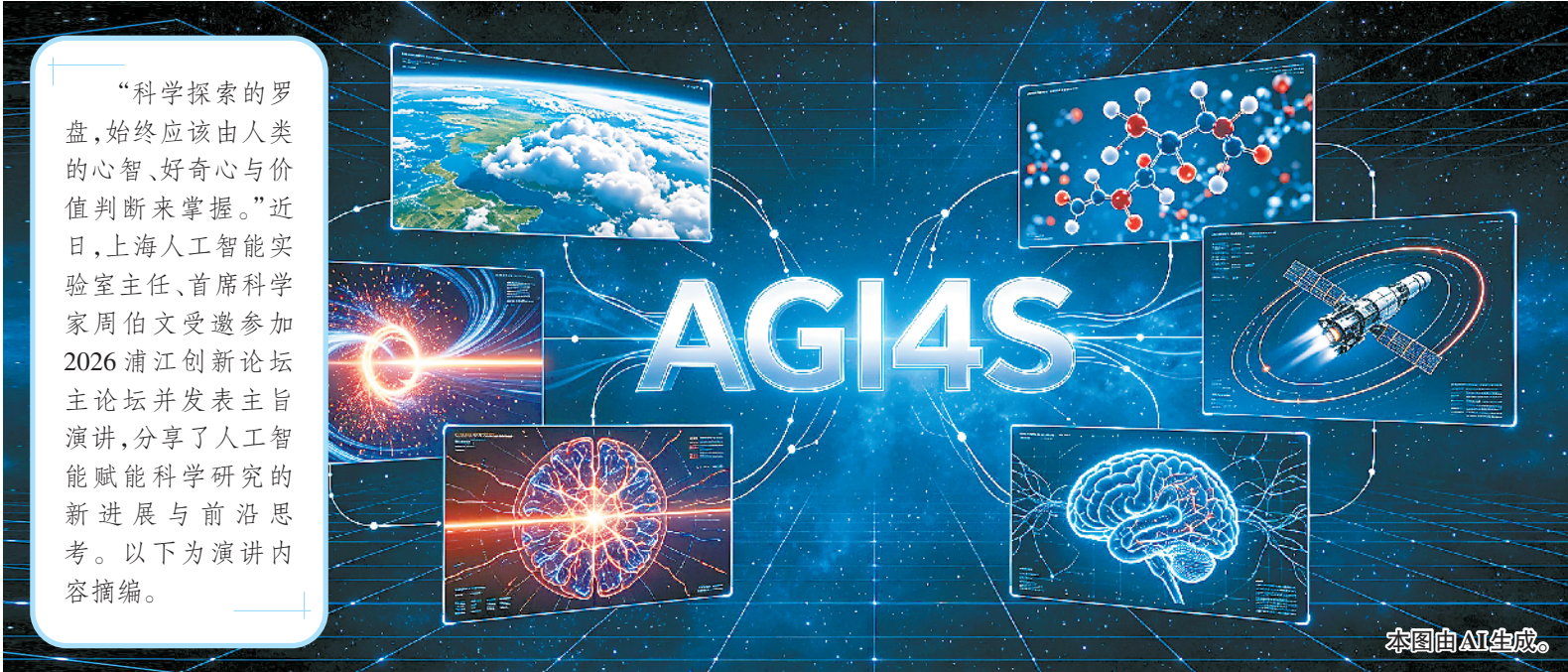


浦江创新论坛上,关于AGI4S的前沿思考

科学研究是下一个编程



科学难题越古老基础, AI提效空间越大

如果人类思考一道难题花100年, AI需要多久? 请大家看3个时间坐标。

第一个, 单克隆抗体。1975年问世, 从患者体内找到对症抗体, 传统流程要走4年。2020年, AI与生物学家协作, 从机器筛选到给药患者只用了90天, 时间压缩到6%。

第二个, 蛋白质折叠。1969年该问题被提出, 人类历经51年求索仍难以破解。AlphaFold仅用2年基本解决这一问题, 约为人类时间的4%。

第三个, 纳维-斯托克斯方程。该方程于1827年被提出, 是千禧年七大数学难题之一。不久前, 一万个AI智能体连续工作88小时, 给出了存在性与光滑性的证明。虽然克雷数学研究所尚未最终确认, 但这意味着, 人类思考了近200年的问题, AI用万分之零点五的时间在逼近。

连在一起看, 一个规律清晰可见: 问题越古老, 越基础, AI提效的空间越大。但我要强调的是: 我们关注的不应只是效率, 而是AI与科学相互深化的关系本身。

从科技创新到产业创新全链条贯通

我们之所以高度重视AI对科学研究的赋能, 核心在于AI能够全链路重塑科技创新范式, 实现从科技创新到产业创新的全链条贯通、一体落地。

在“从0到1”的原始创新阶段, AGI4S能够将科研中大量隐性、散落的碎片化思考, 转化为清晰可落地的科学假设, 让大范围跨领域交叉创新成为常态。传统学科壁垒, 在很大程度上遮蔽了问题本质、降低了科研判断的准确度; 而在AGI4S的赋能下, 学科边界持续消融, 科学家可

以跳出细分领域的固有视角, 更纯粹、更精准地看清问题本源。

在“从1到10”的技术迭代阶段, 创新更多是对已有知识、技术路径的检索、排列与重组。AGI4S让这一过程走向规模化、体系化, 将过去偶然性、个体化的发明探索, 升级为标准化、可迭代的创新工程。

谈到这里, 想与诸位分享我的一个洞察: 现代创新工程化的鼻祖是爱迪生, 他真正划时代的贡献, 并不是灯泡本身, 而是创造了创新的组织形式——他搭建的门洛帕克实验室, 汇聚物理、化学、机械等多学科专家, 通过紧密协作、并行实验、系统化探索, 把零散的科研尝试固化为可复制、可持续的创新机制, 灯泡只是这套先进组织体系的副产品。AGI4S正是以智能范式, 对这套经典创新组织模式进行全方位重构与升级。

在“从10到100”的产业规模化阶段, AGI4S能够大幅加速科技成果量产落地, 缩短传统小试、中试的冗长流程, 催生爆发式成果转化, 真正实现科技创新与产业创新双链融合、深度共創。

借助AI让科学研究像编程一样开展

科研属于极高智力密度的工作, 我们预计AI还需要经历两到三轮能力跃升, 才能更好地辅助科学家开展前沿探索。

基于对上述问题的思考, 我在2026年初提了一个大胆的判断: 科学研究是下一个编程 (Research is the Next Coding)。

关注AI的同仁都清楚, 编程是当前AI落地规模最大、生产力提升最为显著的应用赛道。究其根本, 代码任务具备信号密集、反馈即时、评判标准清晰明确, 反馈即时, 验证可自动化执行, 失败经验能够沉淀复用, 整个过程全程可追溯。

科学研究与编程具备很强的同构性: 二者都以问题为导向、由认知驱动。但二者之间同样存在显著差异——科学研究难度更高、边界更加开放, 价值也更为深远。科研的验证依托真实实验室, 检验来自数字仿真或是物理世界; 试错过程产生的失败经验, 很难转化为可复用资产。同时, 科研的推理链条更长, 有效反馈信号本身十分稀缺,

甚至还会收到存在偏差的反馈。

科研领域普遍存在反馈缺失、反馈失真的难题。那么我们能否借助AI, 让科学研究像编程一样开展? 这就意味着, 未来科研人员可以立足待解决的科学问题, 清晰表达自身的科研意图; AI协同调度文献、数据集、科学知识、计算工具以及自主实验室, 贯通从问题界定、假设生成、方案设计, 到仿真预测、湿实验验证、结果迭代的完整流程。倘若能够实现这一模式, 我们是否就能在科学研究中复刻AI在编程里所取得的巨大成功?

AI和科学形成双轮驱动

我们该如何构建全新的科研范式与方法论, 让这两个轮子相互带动、循环迭代、协同进阶, 持续助力人类形成全新认知, 催生原创科学突破? 在实践中, 我们发现实现双轮驱动有3个关键环节。

第一, 自主进化飞轮。AI必须能够在提出假设、执行验证、获得反馈中持续自我改进, 就跟科学家一样——从一个博士生受训练走向博士后、独立PI, 再到参与有组织科研, 直至成为战略科学家。整个过程中, 很多场景并不存在外部给出的明确反馈信号, 如何实现自主精进、开展更深层次的科学探索, 构建越用越强的正向循环, 是核心难题。

第二, 真实科学验证闭环。科学假设要转化成可执行的实验, 并用真实结果验证推理、修正假设, 以此推动下一轮探索。

第三, 环境即数据。这是上海人工智能实验室新近提出的理念。当前科研不能只依靠文献、分子式、反应式这类静态信息, 这些条件必要但并不充分。真正充分的条件, 是完整留存科研全流程中的高价值信息, 记录任务环境、操作行为、结果变化以及纠错过程, 让科研里所有成功经验与失败教训, 都沉淀为可积累、可复用的学习数据。

这3个关键环节连在一起, 共同构筑起AI与科学持续互促共进的根基。

毫无疑问, AGI会不断拓展人类的知识边界、加速科学发现, 但科学探索的罗盘, 始终应该由人类的心智、好奇心和价值判断来掌握。

EASA举办AI奥运会

航空正从实验室走向实际运行场景

本报讯 日前, 欧洲航空安全局 (EASA) 在德国科隆举办2026年“EASA AI Days”, 并首次推出“AI Olympics (AI奥运会)”, 从全球提交的50个航空AI应用案例中选出10个人围项目, 在现场进行展示和评选。最终, 参会者从10个项目中投票选出金、银、铜奖。

此次入围项目覆盖驾驶舱辅助决策、航空安全分析、适航认证等多个领域。其中, 西班牙 Kairós Aerospace 的 FOR-DEC AI 是一套面向飞行机组的辅助决策系统, 可帮助飞行员整理运行信息、比较不同选项、评估风险, 并识别信息中的不确定性。值得注意的是, 该系统并不试图替代飞行员作出决定, 而是强调让AI的分析过程可解释、可追溯, 最终决策权仍由飞行员掌握。

另一项入围技术则将AI用于航空安全报告分析, 尝试从大量安全报告中提取信息、辅助安全分析。此外, 瑞士 Xcert AI 入围的项目则聚焦航空工程师的适航认证和合规工作, 可利用AI协助处理标准、需求、测试程序以及认证文件等工作。Xcert 表示, 其AI助手将重点服务于航空系统工程师, 以提高认证和合规工作的效率。

相比单纯展示“AI能做什么”, 此次“AI奥运会”更像是一场航空AI应用的现场考试。50个案例先经过评审, 最终10个项目获得现场展示机会, 再由参会者投票产生前三名。EASA 此前公布的活动规则显示, 前三名项目还将获得在大会全体会议上进行专题展示的机会。

与此同时, EASA 正在同步推进航空领域AI监管框架。其 RMT.0742 规则制定任务旨在建立AI可信性监管框架, 并进一步将通用要求适配至不同航空领域, 包括机场运行等。EASA 于2025年11月发布首份AI监管提案, 内容涉及AI保障、人因和伦理等问题, 公众咨询于2026年3月结束。后续工作将进一步针对具体航空领域制定相应要求。

EASA 今年6月发布的最新AI概念文件还进一步强化学习、符号AI以及Level 3 AI系统纳入讨论范围, 其中Level 3被EASA归类为高级自动化。随着AI逐渐从辅助分析走向驾驶舱决策支持、自动化运行和适航认证等关键环节, 监管机构也开始从“AI能不能用于航空”转向研究“AI如何安全、可解释、可追溯地用于航空”。

此次“AI奥运会”释放出一个明确信号, 航空AI正在从实验室走向实际运行场景, 而监管机构也开始同步建立与之匹配的评估和认证路径。

AI下一站, 或趋向本地模型

参数量更小、数据更安全、离线也可用

□ 徐晶卉

日前, 工业和信息化部中国信息通信研究院公布的一份检测报告显示, 上海原点星辉科技有限公司 (StartLux) 研发的本地模型 StartLux-V1.0-27B-Preview, 在可信AI大模型基准测试——MCP专项测试中, 以27B的参数量取得综合性能排名第二的成绩。第一名的DeepSeek-V4-Pro, 模型规模则高达1.6万亿参数。

几乎同期, 国内外AI公司在本地模型研发上动作频频。8月, Meta发布30B本地模型 Muse Glimmer, 云端巨头 Perplexity AI 宣布进入本地模型市场; 9月初, 北京大学“出品”的元空AI联合惠普在上海发布端侧模型产品, 让35B的模型在可断网的电脑上跑。

“AI圈”正在重新审视一条曾被轻视的技术路线: 更轻量的本地模型是否到了崛起的时候?

AI模型并非“越大越好”

过去3年, AI行业笃信“越大越好”, 最直接的竞争方式是扩大参数、增加算力和训练数据。不过, StartLux 首席技术官、语生科学首席科学家郭权玮却发现, 当基础模型能力不断提升之后, 能进入用户工作流的应用却没有同步增加。

郭权玮一直在研究如何提高模型的智能密度, 通过架构、训练和系统协同, 让有限规模的模型能释放更强能力, 并在真实任务执行上逼近更大规模的云端模型。中国信通院的测试结果显示, StartLux 以 Qwen3.6-27B 为基座, 通过自主设计的后训练优化技术, 把一个27B的本地模型推到了与1.6T旗舰同级的位置。“我们采用AI训练AI的方法自主开展训练实验, 并通过反馈持续优化训练策略。”他表示, 这一方法也代表了当前全球模型研发的前沿方向。

有业内专家判断, 模型能力的竞争正在从比参数量转向比智能密度, 而本地模型的进步速度远超硬件。2023年, 最佳本地模型对前沿模型的胜率持平局率仅为23.2%, 2025年就攀升至71.3%, 每年提高约20个百分点。

“3条曲线在今年交汇, 打开了端侧AI的窗口。”北京大学AI联合实验室主任、元空AI创始人兼CEO逢大宽认为, 一方面, 智能密度快速提升, 今天35B参数的模型已达到一两年前云端大模型的能力, 另一方面, 智能体的脚手架能力走向成熟, 同时, 端侧芯片算力持续增长, 本地化浪潮是模型能力、任务执行框架和底层算力三重力量同步成熟的结果。就在近期, 元空AI在上海发布Boxer端侧模型, 涵盖27B稠密和35B-A3B稀疏版本, 其智能密度能追平一年半前的云端标

杆, 这意味着本地设备已具备承载高质量模型并高效执行任务的能力。

盛大网络联合创始人、StartLux 创始人兼CEO陈大年做了一个生动的类比: “人的大脑就这么点容量、这么点功耗, 却产生了人类辉煌文明; 同样水平的能力, 现在的大模型却要靠巨大的参数量去堆, 这应该是策略上出了问题。”他判断, 在95%的场景下, 未来本地模型的能力一定可以打平云端模型, 3年内本地模型会占据80%的市场。

企业定制需求正在上升

无独有偶, 全球巨头也开始关注本地模型的可能性。8月10日, Meta发布30B本地模型 Muse Glimmer, 主张AI应走向本地。8月25日, 头部云端AI公司 Perplexity AI 宣布进入本地模型市场, 推出多模型编排智能体数字员工“Perplexity Computer”的本地化版本。

在专家看来, 本地模型定制需求的上升与企业需求有关。逢大宽认为, 云端大模型在落地时会面临几个现实问题, 首先是数据安全与隐私考量, “今天你把企业的数据全部丢到云上跑一圈, 包括个人的隐私数据、企业的财务数据、制造业的敏感数据, 这真的安全吗?” IDC报告指出, 欧盟《AI法案》与中国《个人信息保护法》配套标准已明确要求



涉及敏感信息的AI处理“原则上应在本地完成”。

逢大宽认为, 成本骤升和稳定性问题也是企业使用云上模型时的痛点。目前, 云端Token费用随使用量呈线性增长, 已进入企业财务报表, 而且部分长任务跑到一半会出现中断, 无法保证持续可预期的输出。据此, 他判断: “AI的下一站, 不在云端, 在本地。”

源自北京大学科学智能学院的元空AI最近就与惠普一起, 将开源大模型和智能体框架, 以及面向办公、金融、科研等不同场景的专有知识, 全部塞进本地设备。元空AI联合创始人宁鲲鹏向记者演示在惠普电脑上“断网整理53张发票”的工作流。记者看到, 在Linux环境下, 元空AI-

Work 调用Boxer本地模型, 自动调取底层的脚本工具, 逐张扫描本地文件夹里的53张发票, 完成格式解析与分类归档, 并在本地生成一张结构完整的Excel报表, 整个过程仅需几分钟。

“金融、医疗、政务等行业对数据出域零容忍, ‘离线可用’的能力为端侧模型提供了爆发的可能。”逢大宽表示, 元空AI Work 可持续24小时不间断工作, 科研场景下可连续执行超过100组复杂实验并完整保留每一组的参数与轨迹。StartLux 也表示, 计划于年内推出第一代本地智能解决方案, 团队正在稳步推进模型训练, 也正对扩散式语言模型等新架构进行深入研究和优化。

据《文汇报》