

中国有望在具身智能赛道率先迎来“AlphaGo时刻”

▶ 本报记者 罗晓燕

在6月12日举行的2026北京智源大会大模型产业论坛上,与会嘉宾围绕“跨越技术成熟曲线:如何定义大模型时代的长期价值”议题展开圆桌讨论,多位专家一致认为,大模型及具身智能的演进远未触及“天花板”,中国有望在具身智能赛道率先迎来“AlphaGo时刻”。

规模化法则远未失效

“大模型整体性能迭代还未到达‘瓶颈’,评测榜单并不完全可信。”北京智源研究院院长王仲远坦言,当前各类大模型评测榜单让人“眼花缭乱”,很多结果无法完全验证,但敢于展示真机、敢于现场演示、敢于进入实际场景的公司,才是真正有底气。

针对市场上“开源与闭源模型差距仅3-6个月”“AI(人工智能)大模型终将沦为卖水卖电”等观点,与会嘉宾均表示不认同。

北京面壁智能科技有限责任公司首席执行官李大海以美国人工智能公司Anthropic为例,指出大模型必须是“T形人才”——

大模型不能只有“通用”这一横线,更需在特定能力上形成纵向长板。“Anthropic正是凭借独步天下的编程能力获得高估值和商业成功。”他表示,大模型必须与场景深度协同,做极致优化。

王仲远进一步指出,Scaling Law(规模化法则)的曲线仍在延续,甚至呈现指数倾向。去年关于Scaling Law失效的讨论,源于大语言模型所用互联网数据“用完了”,但通过后训练、推理优化、Agent(智能体)递归式自我进化等新路径,AI系统能力仍在持续提升。他说,北京智源研究院在多模态领域的探索也证明了这一点:其多模态模型“悟界Emu3”仅用了不到1%的数据、百亿参数,就已展现出明显的性能提升。下一步,北京智源研究院将向世界模型方向探索,寻找物理世界的Scaling(规模化)范式。

世界模型不是具身智能的前提

针对业内关于“世界模型是

否必须先做到完美全场景仿真,才能发展出具身智能”的争论,北京银河通用机器人股份有限公司创始人兼首席技术官王鹤给出了明确的否定答案。“人类也没办法精准模拟、预判世间万物的物理变化,可人们依然能自如地与万事万物互动。完美仿真并非打造具身智能的必要前提。”

据王鹤介绍,银河通用于2025年3月发表全球首篇WAM(World Action Model,世界动作模型)论文,该范式不需要动作标签,可以通过人类视频学习行为的大致动态,从而极大拓展数据来源。

“具身智能正迎来非常光明的Scaling up(模型扩容)时间点。”王鹤认为,未来两年具身智能将全面到达GPT-3.5向ChatGPT转变的关键的预练里程碑。但这需要千万小时高质量数据和百亿以上投入的“入场券”。

当被追问“这是否意味着外界说的世界模型都不靠谱”时,王鹤回应称,WAM本身属于世

界模型范畴,但当前不少世界模型核心能力是作为仿真器训练机器人完成强化学习任务。“人们确实在利用世界模型搭建可交互的可微分物理仿真器,但先让世界模型做到完美仿真再训练具身智能,这条路很难走通。”

在端侧模型方面,李大海透露,端侧模型已从去年的1B(10亿参数)规模升级到4B(40亿参数)规模,明年大概率达到几十B。“具身智能的‘大脑’本质上也是端侧模型,端侧设备资源受限,但模型能力还有极大挖掘空间,真正的‘瓶颈’在于硬件物理条件。”

具身智能或是中国AI的“AlphaGo时刻”

关于中国与欧美AI发展路径的差异,与会嘉宾一致认为具身智能是中国AI有望领先的领域。

王仲远表示,中国的供应链、制造业和场景优势巨大,叠加本土市场容量,足以孵化和催化技术落地。“具身智能、世界模

型很有可能成为中国具有独特性甚至领先的领域。”

“我坚信具身智能是中国的机会,具身智能的AlphaGo和ChatGPT时刻会在中国实现。”王鹤认为,如果具身智能“从0-1”在中国完成,那么“从1-100”必定在中国成熟。

李大海将根本优势归结为人才。“中国拥有最聪明的青年才俊,且数量全球最多,这是中国具身智能有望领先的最底层最重要的因素。”他说。

在AI安全与责任划分上,李大海直言,人类社会一直遵循“吃一堑、长一智”的信条。“AI时代,人们发现和修补安全漏洞的效率大幅提升,代价有望降低,但安全风险往往从意想不到的维度浮现,用教训推动规则完善,这是需要正视的现实。”

王鹤补充说,在工业场景中,具身智能机器人首先要做到和人类一样出色地完成工作,并能够承担相应责任,更长远的人机权责划分将在行业探索中逐步建立。

AI正在叩响“触达实体世界”大门

▶ 本报记者 张伟

在6月12-13日举行的第八届北京智源大会上,北京智源研究院(以下简称“智源”)发布“悟界”全栈技术体系,标志着人工智能(AI)正在完成从“理解语言”到“读懂真实物理世界”的范式跃迁,通用智能体、具身智能、生命科学等前沿成果同步落地,AI正在叩响“触达实体世界”的大门。

智源全面亮出“悟界”全栈技术体系,包括悟界·Brainμ1.0、悟界·OpenComplex2.5、悟界·Physis-v0.1等一系列创新成果。其中,悟界·Physis-v0.1是全球首个通用世界基座模型,以统一物理状态学习实现物理正确、动作因果可溯、长程一致、通用泛化,最终达到全垂类场景应用。

预测下一个物理状态

北京智源研究院院长王仲远表示,人工智能正经历一场重大范式变革,但火爆的世界模型仍处在早期形态,甚至现在世界模型领域夺冠的模型都还不是真正的世界模型。从大语言模型到登上《自然》正刊的多模态大模型“悟界·Emu3”,智源试图再次定义AI的下一站。

世界模型不仅能感知、理解、推理真实物理世界的时间、空间、物理规律和物理常识,还能涵盖文本、视频、深度、力觉、感知等全模态数据,具备主动交互能力,能够支撑各种物理世界的下游应用。



在第八届北京智源大会乒乓球人机对战体验区,观众与人形机器人切磋球技。
本报记者 罗晓燕/摄

“人工智能的范式迭代,正从‘预测下一个词元’迈向‘预测下一个物理状态’。”王仲远表示,这是大模型发展至今水到渠成的必然方向。从早年的大语言模型处理文本,到多模态模型融合图像、音频,再到如今需要解决真实物理空间的时空规律、长时序理解等问题,人工智能正在跨越数字与物理的边界。

王仲远将现有技术路线划分为四大类:以语言为中心的世界模型、以像素为中心的世界模型、以三维结构为中心的世界模型和以视觉表征为中心的世界模型。

“世界模型是通往物理通用人工智能的必由之路。”王仲远认为,四大类模型距离真正面向物理世界的基座模型还有很大距离。他说,智源正在尝试一条可能被视为“第五类”的路径,即以语言为中心的分类和以视觉表征为中心的分类融合,也叫作潜空间表征。

“将来统一的潜空间建模不仅仅是视觉空间,而是全模态潜空间,很有可能是世界模型真正下一个可能的路径。”王仲远说。

与会人士表示,2026年,人工智能领域正经历一场深刻的

范式转移。随着大语言模型在数字文本处理上趋于成熟,行业的共识已清晰指向下一个核心赛场——物理世界。AI不再局限于屏幕内的对话与生成,而是开始通过具身智能、世界模型和自主智能体,真正“触达”并融入实体世界。

需要海量数据训练

北京智源研究院理事长黄铁军在“AI×神经科学”分论坛上阐述了“结构决定功能”与“功能塑造结构”的辩证关系:生命科学中的基本信条是“结构决定功能”,即DNA(脱氧核糖核酸)决定生命个体的身体、大脑和功能,但在生命进化过程中,环境塑造身体、DNA与大脑。个体出生后,后天环境训练神经突触,则属于“功能塑造结构”。当前人工智能以Transformer神经网络为结构,采用数据驱动训练神经连接权重,模仿了生命先天的结构决定功能和后天的功能塑造结构,未来的人工智能需要进

一步优化乃至变革结构,并通过更真实的环境感知训练更高水平的智能。

“智源一直围绕结构决定功能与功能塑造结构两个角度探索人工智能前沿。大模型就是采用Transformer结构,采用海量数据训练而成的。”黄铁军表示,从第一性原理出发,世界大模型必须“要有光”,突破图像视频的低速瓶颈。借鉴灵长类视网膜结构研制的脉冲视觉芯片和相机,实现单像素光强测量,能够在光发生变化的时空点进行必要计算,为世界大模型训练推理打开高速高效窗口提供高保真的物理观测数据。

据悉,自2018年成立以来,智源构建全栈大模型开源技术体系。截至目前,智源开源模型超200个。

从“悟道”到“悟界”跨越,智源正在推动人工智能、物理世界和生命科学“三体互动”,构建AGI(通用人工智能)的大脑、眼睛和身体闭环,让智能真正嵌入物理世界。

新闻背景

世界模型思想源自1943年认知心理学,1990年正式定名,2018年相关论文使其成为通用AI概念。后续形成Meta JEPa认知、Sora视频生成两大路线,前者侧重物理因果推理,后者擅长画面仿真。2026年智源发布悟界·Physis-v0.1,推出国产通用物理世界基座,补齐实体智能底层能力。