

“有关世界模型,理性的声音越来越多了”

北京商报:应该怎样理解多模态对世界模型的重要性?

王仲远:物理世界不只有语言、文字,物理世界是多模态的乃至全模态的,有图像、声音、触觉。比如当机器人在家庭里服务,要知道哪些杯子是易碎的,要知道水热不热。当它进入工厂组装产品、搬运货品,要感受到真实的重力、精细化的操作,比如我们去诊断疾病,不光靠文字描述,还需要去做一些检查,这些数据都是多模态的。

北京商报:不久前,智源将世界模型分成了四条路线,这对行业产生了怎样的影响?

王仲远:智源大会上,我们提出了自己的世界模型分类框架,共四条,第一类是以语言为中心的世界模型,像VLM(视觉—语言模型)、VLA(视觉—语言—动作模型);第二类是以像素为中心的世界模型,像视频生成类的模型

Seedance、Sora;第三类是以三维结构为中心的世界模型,像World Labs联合创始人兼CEO李飞飞团队的Marble模型;第四类是以视觉表征为中心的世界模型,比如2018年图灵奖得主杨立昆的JEPA系列模型。

这四类模型确实都跟世界模型相关,但没法完全达到我们所期待的能力。智源正在尝试一条可能被视为“第五类”的路径,以语言为中心的分类和以视觉表征为中心的分类的融合,也叫作潜空间表征。之前世界模型的概念常被误用,在我们提出分类后,虽然没有完全扭转夸大表述的风气,但现在理性的声音越来越多了。

北京商报:智源计划推出世界模型评测体系,请问该体系会考虑哪些维度?

王仲远:当前世界模型缺乏完备的评测方法和框架。很多评测都是以视频生成为重点,这不代表世界模型作为基座模型的核心能力。

现在世界模型领域夺冠的模型,都不一定是未来真正的世界模型。智源曾提出世界基座模型需要预测下一个物理状态,2026年下半年我们会推出自己的评测框架,目前已经有了一些评测的数据集,例如开冰箱、开洗衣机、开微波炉(旋转式、按压式、触碰式)等任务,会评估现代人工智能模型能否很好地理解并进行相应的预测和识别。

北京商报:接下来会对世界模型有哪种期待,可能会在什么时候出现让大家比较惊艳的能力?

王仲远:至少需要好几年,科研是非常严肃的,很可能卡在某个地方三五年都没有突破,也有可能忽然就突破了。未来3到5年会是世界模型持续演进和迭代的阶段,我们也期待数字世界大模型的成熟,能够加速面向物理世界的人工智能基座模型、世界模型的迭代。

“智能体爆发是量变到质变的必然结果”

北京商报:最近北京发布了《北京市关于加快智能体引领发展的若干措施》,站在技术的角度,能否解读一下智能体与基座模型的关系?

王仲远:智能体的爆发,本质上是基础模型能力与工程架构长期积累后从量变到质变的必然结果。智能体的能力上限,取决于基础模型在复杂推理、任务规划、在线学习和自主进化等方面的水平,而其运行效率与稳定性,则有赖于上下文工程、记忆机制、任务持久化、多智能体协作和工具调用等系统工程能力。当前,我国基础模型总体仍处于跟跑与并跑交织阶段,大语言模型与国际前沿水平差距持续缩小,具身模型、世界模型等下一代模型正处于并跑阶段。

北京商报:2026年,Token经济是热门话题,它会怎样拉动宏观经济?

王仲远:智能体多轮自主调用将带来Token消耗的指数级增长,未来Token消耗量堪比工业时代的耗电量,将有望成为智能时代的核心经济指标。当前,我国基础模型推理调用的Token综合成本仅为国际先进水平的1/10—1/6,成本优势极为显著,叠加海量应用场景,有望推动北京乃至中国在全球范围内率先引爆Token经济。

北京商报:国际社会对AI安全都有共识,

现在关注点有没有更聚焦?

王仲远:前几年大家讨论AI安全,比如超级人工智能出现,会不会对人类产生毁灭性的打击和影响,这两年的AI安全问题是实实在在的,比如效率提升以后带来的产业链变革和重组。可以看到,每次技术进步一定都会对落后产能、低效率的方式有影响,我们乐观地相信会创造出新的增量价值和机会,人类社会一直都是和新技术共存,并伴随着演进和演化。

因为AI具备越来越强大的能力,智能体会绕过很多原来人类设下的安全机制,最先进的模型对于生物安全、管理安全的影响也迫在眉睫。这不是一家科研机构能够单独解决的,需要升级产业链、联合跨领域专家、制定政策法规,需要全社会共同努力。

北京商报:站在当下,您对AI的未来有哪些预判?

王仲远:大模型能力会继续上探,也必须能够解决实际问题, AI Coding (AI编程)、智能体和端侧AI是大模型落地的核心方向。当Coding能力不断提升,意味着整个数字世界可能被AI重构,智能体能帮助人类完成具体的任务。智能已经达到相对可用的水平,且不断往小模型压缩、降价,会出现AI普及的可能性。

北京商报记者 魏蔚



智源世界模型的技术演进 ▶▶▶

2023年智源大会上

杨立昆(Yann LeCun)
阐述了新一代世界模型的概念

2024年智源大会上

智源研究院提出的人工智能大模型
技术路线预判
明确指出世界模型是下一代大模型技术

2024年 ● 发布**悟界·Emu3**

2025年 ● 发布**悟界·Emu3.5**

更是全球首个原生多模态世界模型

2026年 ● 智源推出了
悟界·Physis-v0.1

●● **悟界·Physis** ●●

通用世界基座模型

●● **悟界·RoboBrain Orca** ●●

以下一个物理状态预测为核心的
具身大脑