

走近科学家

编者按/科技兴则民族兴,科技强则国家强。在经济发展范式转移与产业迭代的进程中,科学技术的价值有目共睹。科学家作为新一轮科技革命和产业变革的重要推动力量,总能前瞻性地发现某一领域运行中的新规律、新价值。本报特推出《走近科学家》系列报道,与科学家对话,走进其思想深处,在了解技术发展前沿方向的同时,有助于我们在纷繁喧嚣的行业发现中发现先机、把握规律、洞察本质、坚定信心。

对话肖仰华:机器智能的天花板在哪里?

本报记者 郑瑜 北京报道

人工智能(AI)逐渐成为各行各业转型升级的重要驱动力。

但硬币总有两面。随着AI大模型的加速渗透,数据质量、模型“幻觉”(指大模型生成的内容表面上合理,但实际上是错误的)以

大模型要“吃饱”也要“瘦身”

大模型瘦身的过程本质上是一种“遗忘”的过程,这与人类知识的发展过程非常相似。这个遗忘的过程,实际上是在提炼和沉淀真正重要的知识。

《中国经营报》:现有的大模型在实际应用中,数据的使用效率似乎较为“低下”,你能否详细解释一下这一现象的原因?

肖仰华:这一问题从根本上讲源于其训练过程的不透明性,整个训练过程仍处于“黑盒”阶段,我们对大模型训练的内在过程和机制缺乏清晰的理解。

尽管我们认识到大模型的工程本质是数据工程,但对于具体哪些数据能够发挥什么样的作用,对哪些能力起着决定性的影响,认知还较为初级。同时,对数据之间的关联性、数据分布的形态、数据类型与大模型能力之间的关系也缺乏全面的认识。以金融大模型为例,选择什么样的金融语料才能训练出一个胜任金融工作的大模型,我们所知甚少。

如何通过特定的数据和训练过程来激发和形成大模型的特定能力。这些问题的复杂性和不确定性,反映了我们在理解和优化大模型训练过程方面仍然面临的巨大挑战。我们需要更深入地理解大模型的训练过程,探索数据、训练方法与模型能力之间的关系,以期在未来能够更有针对

“协同”才能解决“幻觉”

协同是一种解决幻觉非常有效的思路,甚至可能是根本思路。

《中国经营报》:面对大模型对数据的“贪婪”需求,如何在数据的存储、流通与消费上找到平衡?

肖仰华:大模型在数据方面面临的根本挑战在于我们尚未建立起大模型的数据科学。

具体来说,我们还不清楚在大模型的训练和优化过程中到底需要什么样的数据,需要怎样的数据配方,需要怎样的学习课程。数据和模型能力之间的因果关系尚未建立,这是根本性的问题。

大模型的训练同样应该遵循类似的原则。如果我们希望训练出的大模型具备相对高级的认知能力,就必须对其“喂养”的数据进行精心选择。

我们需要建立起大模型的数据科学。这包括几个关键方面:数据的分类机制;我们需要了解不同类型的数据对模型能力的影响;数据的筛选机制;建立标准来选择高质

技术治理应该未雨绸缪

AI的发展一旦踏上加速的油门,是有可能让我们的刹车失灵的。对于这种可能性,我们应该保持高度警惕。

《中国经营报》:你强调AI技术的发展仍要“以人为本”,能否详细阐述对“机器取代人类”这一观点的看法?

肖仰华:发展AI要加强的首要原则就是以人为本,科技的尽头是人文,科技的终极目的是人文关怀。人不仅仅是万物的尺度,也是AI等先进技术的尺度。

我们要有所为有所不为,但凡伤害人之为人之本性的AI应用,我们要谨慎,要加以限制。事实上,技术的每一次进步都有可能带来人的某种能力的倒退,比如说汽车普及了,我们的四肢功能就要靠健身才能保有;键盘普及了,我们很多人就提笔忘字了。

当通用人工智能与脑机接口日益成熟,机器人和工具将大量代替人类从事脑力劳动,脑力实践活动的减少,会不可避免地带来人类智力

及如何有效利用私域数据等问题逐步浮现,正在成为决定市场发展进程和天花板的关键因素。

大模型是越“大”越好吗?如何在数据的存储、流通与消费上找到平衡?怎样进一步开发庞杂而敏感的私域数据?最终,机器智能的天花板又在哪里?

性地提升大模型的性能和应用效果。

《中国经营报》:你提到过大模型的发展过程中将经历“水分”挤掉的过程,能否进一步解释这个过程?

肖仰华:大模型的发展可能会经历几个关键阶段。首先是一个“吃饱”甚至“撑得太饱”的阶段。在这个阶段,我们用海量的数据和语料来训练模型,可能会产生拥有千亿甚至万亿参数的大模型。然而,这样训练出来的模型往往显得臃肿,容易出现各种问题。

接下来,我们就需要进行“瘦身”,让大模型变得更加精炼。从学术角度来看,这个瘦身的过程本质上是一种“遗忘”的过程。这与人类知识的发展过程非常相似。这个遗忘的过程,实际上是在提炼和沉淀真正重要的知识。

对于大模型来说,那些相对琐碎的知识完全可以放到外部的数据库或知识库中,让模型能够按需检索。真正智慧的系统不应该用自己的“认知载体”去存储琐碎的信息,而是用来沉淀高度精华的智慧。

更深层的做法,是逐步将大模型的多项能力进行解耦与剥离。

量、有价值的数据;数据的配比机制;研究不同类型数据的最佳组合比例;评测评价标准;建立科学的评估体系,衡量数据质量和其对模型能力的影响。

只有建立起这样的数据科学,我们才能避免大模型训练过程中的盲目性,也才能避免数据使用上的过度或不足。这将使得大模型的训练更加精准和高效。

《中国经营报》:生成式大模型的幻觉问题一直备受关注,你能解释一下这个问题的本质和可能的解决方案吗?

肖仰华:生成式大模型的幻觉问题从表面上看,源于其本质是一个概率生成模型。模型在生成过程中,本质上是在预测下一个词的生成概率,在整个词表空间计算概率分布,选择概率最大的词输出。这种基于概率计算的本质决定了它有一定概率会出错。

从深层次原因来看,幻觉的本

质与人类智力退化有关。大模型是通过我们能够收集的语料、文本或各种数据来训练的统计模型。然而,我们能够收集到的所有数据,包括我们能够表达的所有可能的思维过程等数据,从概率分布上来讲总是有偏差的,很难精确表达或还原真实世界。

大模型训练出来的模型多多少少被认为是有偏差的,会受到样本分布偏差的影响,带来结果预测的不准确,在特定情况下产生所谓的“幻觉”问题。

解决幻觉问题的一个思路是给模型构造足够长的上下文,这也是RAG(检索增强生成)技术要解决的主要问题。因为给模型提供的上下文提示越丰富、越合理,它预测的概率就越准确,从而缓解幻觉问题。比如我们可以把最新的金融行情信息作为检索来源,增强大模型对实时金融行情的认知,从而缓解大模型幻

想无法承受的。我们要从根源弄清楚某一项AI技术的大规模应用,对人类社会所能造成的长远影响。AI的发展一旦踏上加速的油门,是有可能让我们的刹车失灵的。对于这种可能性,我们应该保持高度警惕。

《中国经营报》:展望未来,你对下一代万亿级参数模型以及AGI的实现有哪些预期和想法?

肖仰华:这个问题的回答本质上需要首先澄清:机器智能的天花板在哪里?目前,我们已经看到人工智能实现了类似人类大脑系统一的直觉思维,并初步具备了系统二的理性思维或“慢思考”能力。一些乐观的观点认为,这已经预示着人工智能正在全面超越人类智能水平,也就是所谓的实现AGI。我倾向于认为,人类独具的高

度,也是AI先进技术的尺度。事实上,技术的每一次进步都有可能带来人的某种能力的倒退。在肖仰华看来,随着技术的不断进步,人类应该更加关注对技术的治理,AI是否能够真正实现通用人工智能(AGI)的终极目标,仍然是一个值得关注的前沿课题。

例如,将认知能力和事实知识进行分离,或者将逻辑推理和直觉推理进行分离。目前的大模型参数和神经网络结构实际上是将知识与能力糅杂在一起。未来,我们可能会通过一些技术手段,将大模型的能力相互剥离和拆解,从而能够根据需求组装出具有特定能力组合的大模型。这有点类似于人脑的不同区域完成不同功能的方式。

如果某个应用场景只需要部分能力,我们就可以只组装相应的大模型功能模块,而不需要启用参数规模巨大的完整模型。

《中国经营报》:私域数据在各行业的深度应用,被认为是未来AI模型突破的关键,你认为目前行业在利用私域数据时最大的挑战是什么?

肖仰华:在利用私域数据时,行业面临着诸多挑战,包括技术层面的和生态层面的挑战。生态层面的挑战在于当前数据要素市场尚不健全,使得私域数据的汇聚和交易流通也面临诸多困难。如何完善数据要素市场建设,让数据供得出、流得动,仍然缺乏制度保障。

技术挑战从根本上来讲仍是因为大模型自身能力不足。首先,大

模型对专业性、复杂性数据的理解不足。私域数据有着较强的私有性和专业性,对大模型理解数据进而利用数据造成了困难。比如很多金融信息系统有着各自行业数据编码标准、行业分类标准,大模型难以理解这些私有化的表达;如果没有背景知识的支撑,大模型也难以理解专业性极强的数据(比如会计审计中各项数据)。其次,大模型难以理解复杂数据结构。数据库的复杂表结构、数据库的复杂数据模型、非结构数据库的复杂数据建模,均给大模型理解背后的数据带来了巨大障碍。

肖仰华
复旦大学计算机科学技术学院教授、博导、上海市数据科学重点实验室主任



肖仰华
复旦大学计算机科学技术学院教授、博导、上海市数据科学重点实验室主任

模型对专业性、复杂性数据的理解不足。私域数据有着较强的私有性和专业性,对大模型理解数据进而利用数据造成了困难。比如很多金融信息系统有着各自行业数据编码标准、行业分类标准,大模型难以理解这些私有化的表达;如果没有背景知识的支撑,大模型也难以理解专业性极强的数据(比如会计审计中各项数据)。其次,大模型难以理解复杂数据结构。数据库的复杂表结构、数据库的复杂数据模型、非结构数据库的复杂数据建模,均给大模型理解背后的数据带来了巨大障碍。

现象。

另一个思路是认识到大模型自身能力的局限性,让它与其他人工智能组件协同工作,包括知识图谱和小模型。从成本和经济效益考虑,这种协同方案可能更具经济性。

大模型参数量大,推理和训练成本高,而且不可控、不可编辑、不可理解、不可解释,存在诸多缺陷。而这些缺陷恰恰是小模型和知识图谱的优势所在。小模型参数数量低,训练代价低,可控可理解可解释;知识图谱的知识可编辑,表达更精炼。因此,协同是一种解决幻觉非常有效的思路,甚至可能是根本思路。

金融行业是信息化与数字化基础较好的行业,已经建立大量的金融小模型与金融知识图谱,将其与大模型有效协同,而非简单丢弃不用,是未来金融智能化的主要实现途径之一。

级认知能力是当前甚至是未来机器智能难以达到和超越的。人类高级认知能力中自省、反思、ToM(Theory Of Mind,也就是对他人持有信念的理解能力,比如“我知道别人知道什么”“我知道别人知道我知道什么”),是当前AI难以实现的。

如果不断剥离上述高级认知能力的表层,我们就会触及人的根本认知能力:自我意识。人类的自我意识具有递归性。比如我可以认识当下的我在思考,进一步我也能对我在当下的思考状态进行思考,如此循环往复。这种递归性的反思能力是人类能够不断超越当前提升自我的根本所在。自我超越的能力似乎只有人类能够具有,目前AI在自我完善、提升、超越仍在研究阶段,即便AI具备形式上此类能力,它仍然缺乏人类的内在动机。

元宇宙与AI融合加速 产业亟待一体化监管

本报记者 蒋牧云 何莎莎
上海 北京报道

随着AI技术的不断成熟,元宇宙呈现出与AI深度融合的发展趋势。

《中国经营报》近期从多位产业人士处了解到,基于AI数据需求的不断增加,相关数据资产的需求井喷,有元宇宙科技公司正以高仿真的3D数字资产库为核心发力方向。

深度融合已成发展趋势

作为上海科创中心的重要发展方向之一,元宇宙产业正在获得更多资源倾斜。

记者近期了解到,为落实《上海市培育“元宇宙”新赛道行动方案(2022—2025年)》的具体任务,上海数字文创元宇宙产业联盟正通过构建产业研究体系、产业联动生态和项目落地机制,促进各市场主体交流合作,市区联动推动产业集聚和项目落地。

据悉,该联盟通过开展常态化的元宇宙项目路演活动,吸引了众多投资人、企业家及行业专家参与,包括政府产业基金、上市公司战投、文化集团公司战投以及风险投资机构,已累计为多家企业进行股权融资服务。

近期,上海市科学技术委员会根据《上海市“元宇宙”关键技术攻关行动方案(2023—2025年)》,发布了上海市2024年度“科技创新行动计划”元宇宙项目指南,为研究与开发相关关键技术或产品的企业提供非定额资助。

在技术路径上,元宇宙则在加速与AI融合。

汇付天下元宇宙研究院院长迟轩告诉记者,AI技术是构建元宇宙底层技术的关键,包括图形图像、感知交互、数字孪生和数据可信流通等领域。这些技术使得元宇宙能够提供更加真实的体验,并实现虚拟与现实的无缝连接。AI技术可以自动化地生成元宇宙中的内容,提高内容创作的效率。

汇付天下元宇宙研究院院长迟轩告诉记者,AI技术是构建元宇宙底层技术的关键,包括图形图像、感知交互、数字孪生和数据可信流通等领域。这些技术使得元宇宙能够提供更加真实的体验,并实现虚拟与现实的无缝连接。AI技术可以自动化地生成元宇宙中的内容,提高内容创作的效率。

技术、监管框架等仍待优化

苗青同时也指出,在全国范围内,上海的元宇宙综合指标仍有提升空间,需要不断加强商业模式创新、提升用户体验、丰富场景和内容。

该联盟相关负责人进一步向记者表示,一方面,元宇宙产业技术成熟度还不足,尽管VR、AR等技术已经取得了一定的进展,但目前的设备仍面临显示清晰度、算力不足、重量较重等技术问题,在提供沉浸式体验方面仍有不足。另一方面,在商业模式上,大多数企业仍处于探索阶段,这一领域的资金投入也具有一定的不确定性,而这些原因都限制了元宇宙优质内容的生产。

迟轩则告诉记者,除了技术成熟度不足将影响应用的稳定性和用户体验之外,监管合规以及用户接受度都需要不断完善与加强。监管方面,元宇宙的跨界特性使得监管变得更加复杂,需要新的法律框架和监管机制来适应这一新兴领域。接受度方面,目前用户对新技术持观望态度,对元宇宙中的金融交易安全性和便捷性仍然缺乏信任。

从监管的复杂性角度,万商天勤律师事务所合伙人、上海区块链法学会智库专家和科技评价专家张烽具体向记者阐述到,由于元宇宙技术应用连接和产品服务非常广泛,涉及传统政策及监管部门非常多,因此,需要建立相应的协同监管框架,以明确相应职责、加强跨部门协作,从而形成合力。

张烽进一步指出,要形成元宇宙各部门监管一体化,需要采

多位业内人士告诉记者,在AI技术的加持下,元宇宙能够提供更加真实的体验,并实现虚拟与现实的无缝连接。

不过,受访业内人士也指出,不论是元宇宙商业模式,还是元宇宙技术以及监管框架等都有待进一步完善与优化。由于元宇宙技术应用连接和产品服务非常广泛,涉及传统政策及监管部门较多,业内也在期待一体化的监管框架。

上海数字文创元宇宙产业联盟相关负责人也表示,AI能够为元宇宙提供更加智能化的服务和功能,如智能推荐、自动化管理、内容生成等,从而提升用户体验和元宇宙的整体效能。同时,元宇宙也为AI提供了丰富的数据和场景,有助于AI技术的进一步发展和优化。例如,AI驱动的数字人、虚拟角色的交互能力得到大大提高,除了具有不同外貌、性格和行为模式之外,AI能够理解用户的意图并给出回答和反馈,让用户体验更加自然。

两者的融合,也催生出更多应用场景。不久前,上海银行正式发布了市场首家元宇宙银行。该银行借助元宇宙、区块链及AIGC技术,以全新的虚拟数字空间为载体,使得金融消费者获得更好的个性体验、情感认同、社交满足和隐私保护。

迟轩也告诉记者,汇付天下也在相关场景中探索元宇宙技术的应用,例如,为了确保商户的知识产权和数字金融安全,公司打造了“区块链+链上可编程支付”解决方案,支持数字文创、版权分配、数字权益证明及发放、会员及积分体系、游戏、元宇宙等Web3.0多元消费场景。

上海数字文创元宇宙产业联盟副秘书长苗青认为:从虚拟环境的精细打造到智能角色的创造,再到沉浸式交互的实现,人工智能作为其核心驱动力,正在重构元宇宙体验。

取相应措施构建产业监管一体化。他建议,基于现有监管机构组织职能、流程和运行,在元宇宙产业发展战略指导下,结合其产业体系构建要求,根据各相关机构定位划分为产业发展决策层、产业发展执行层、产业发展平台层和产业发展保障层,形成一个浑然一体的产业发展组织体系,因而从根本上形成监管合力。

从企业的角度,迟轩则告诉记者,为应对前述的发展挑战,汇付天下在加大对元宇宙相关技术研发投入的同时,也在积极与监管机构合作,探索适应元宇宙特点的法律和监管框架,确保合规运营。与此同时,也将通过相关教育引导用户正确使用相关服务。

多位受访的业内人士均认为,元宇宙将与数字经济互相促进。

迟轩表示,元宇宙为数字经济提供了新的应用场景、商业模式和技术发展方向,比如支付就是元宇宙最重要的基础设施。

前述联盟相关负责人也表示,元宇宙作为一个由人工智能、虚拟现实、区块链等技术构建的虚拟世界,为数字经济提供了新的平台和机遇,而数字经济则为元宇宙的发展提供了坚实的基础和动力。

他认为,近年来,元宇宙相关项目在上海出现了现象级的火爆,参与的上下游单位不断增加,产生了新的消费模式和经济增长点。同时,元宇宙在多个行业的应用,可以为传统产业提供数字化转型的新路径和新手段。