

声音

中国AI大模型的差异化发展之路

访北京瑞泊控股(集团)有限公司创始人、董事长乙壤月

近两年来, AI大模型持续占据全球科技界的“顶流”位置, 至今热度居高不下。因具有实质性突破的ChatGPT、Sora均出自美国人工智能公司OpenAI, 中国如何发展大模型、实现差异化突围成为行业内外关注的焦点问题。对此, 业界有一个形象的比喻: 中国发展大模型不能跟美国“打篮球赛”, 一直

跟随着“满场乱飞”, 而应当使用“下围棋”的方法, 在技术相对落后的情况下, 通过大模型赋能各个产业领域, 以实现人工智能在应用上的突破和提升。

北京瑞泊控股(集团)有限公司(以下简称“瑞泊”)是中国最早进入预训练大模型技术领域的团队之一, 自2018年起与OpenAI这些国际主流AI实验

室同步开始预训练大模型及相关应用的研发, 瑞泊核心团队在自然语言处理技术上已有超过十年的技术积累。瑞泊先后获得国家级科研机构2021年科技成果转移转化一等奖、被誉为人工智能界“奥林匹克金牌”的2022年“中国AI金雁奖”技术创新大奖、2023年中国十大大模型案例及国家特色产业集群

赋能典型案例大奖。目前, 瑞泊VIDYA智慧认知大模型已应用于金融、工业、IT运维等众多行业领域。

为了深入了解中国大模型技术的发展状况, 尤其是在各个行业领域的赋能价值与应用潜力, 《中国经营报》记者近日专访了瑞泊创始人、董事长乙壤月博士。

在行业落地方面中美起步差不多

大模型是“大力出奇效”的工程成果, 以大参数、大算力、强算法为显著特征, “规模定律”(Scaling Law)也被大模型厂商奉为圭臬。目前我国已有超过100个大模型, 但由于算力被“卡脖子”, 我国大模型的算力参数与美国相比差距在逐渐拉大。基于这种客观现实, 中国大模型的从业者需要运用“下围棋”的智慧, 从虚与实、攻与防、利弊权衡之间找到一条新的突围路径。

《中国经营报》: 你怎么看中国和美国在大模型领域的差距?

乙壤月: 我们都知道, 从ChatGPT到Sora, OpenAI的“暴力美学”屡试不爽。这种方法论, 实际就是在当前技术和硬件条件下, 通过大规模地投入来推动人工智能发展的可行性和有效性。在通用人工智能大模型的基础研发方面, 中国与美国是存在客观差距的, 且这个差距短期还有可能进一步拉大, 我们从人才教育培养, 以及算力、芯片、深度学习框架等核心技术攻关层面, 都有较大的追赶空间。

在正视差距的同时, 我们也要对中国人工智能的发展保持信心。需要强调的是, 这里的信心并不是单纯的民族情绪, 而是有真正底层支撑的信心, 因为中国在发展人工智能方面拥有独特的优势——这些优势可能是包括美国、欧洲等其他国家和地区市场所不具备的, 总结起来就是三点: 需求规模大、产业配套全、应用场景多。从中央到地方

的政策层面也能看到, 新质生产力、人工智能+行动、智能制造、面向金融及工业领域等关键垂直的大模型等等, 都在试图将优势扩大深化。

我一直认为, 应当把通用大模型与垂类大模型分开来看, 在后者大模型的各行业应用, 我们和美国的起步是差不多的, 我们并不落后, 甚至是领先的。

《中国经营报》: 垂类大模型与通用大模型相比有什么区别?

乙壤月: 大模型的底层技术原理实际是一样的, 无论是哪一类模型, 共同目标都是利用大数据和复杂算法来模拟、理解以及生成人类的语言和知识, 进而在各自的应用场景中发挥作用。

通用大模型是以广泛的、跨领域的的数据为基础构建的, 其训练资料来自互联网、书籍、报纸等多种多样的来源, 能够掌握和生成包含广泛知识的内容。行业大模型相当于专门业务领域的专家, 通过深度学习特定行业或领域的数据, 在专业领域内提供更加精确、深入的分析和解决方案。

简而言之, 二者的专业深度和应用范围有较大的区别。可以类比通识教育, 到了大学本科的三四年级可能就要往专业方向上走, 变成一个专才, 才能在人才市场更具优势。通用大模型擅长处理广泛的问题, 拥有宽广的知识面, 但在特定领域的深度和精准度上远不及专门训练的行业大模型, 行业大模型则专注于深入挖掘特定领域的



知识, 提供更专业、更精确的服务和解决方案。

除了专业性, 另一个方面是涉及企

业的信息和数据安全, 很多企业的数据库按照法律法规的要求是“不出门”的, 那就必须本地化部署自己的专属大模型。

企业应避免盲目追求新技术

AI大模型浪潮势不可挡, 在众多企业投身AI大模型赛道的同时, 各行各业也都在争相拥抱AI大模型, 希望大模型能够推动行业的变革与创新发展。作为行业大模型提供商的瑞泊团队真切地感受到需求侧对于运用AI大模型的迫切与热情, 唯恐错过而被市场所淘汰。

《中国经营报》: 从瑞泊的业务感知来看, 需求侧现在对于大模型的认识与理解跟去年相比有变化吗?

乙壤月: 变化非常明显, 政、产、研的客户或合作伙伴, 都对大模型表现出很高的热情。去年, 我们在与客户的沟通中经常需要进行技术科普, 解释大模型的基本概念和原理。而今年, 很多客户更关注的是如何将大模型应用到具体的业务场景中, 以实现成本降低和效率提升。这一转变反映出社会对新技术的快速适应和渴望通过技术创新促

进企业成长的共识。

《中国经营报》: 在这一波的AI大模型浪潮中, 瑞泊的定位是什么? 你们着重解决的是什么问题?

乙壤月: 瑞泊VIDYA大模型专注垂直行业领域, 瑞泊的定位是关键行业的“超脑构建者”。行业大模型是“超脑”的组成部分之一, 超脑还涵盖行业专业模型、大模型原生技术组件、前端AI应用集群等。因为单靠大模型是无法完成那么多智能化的行业应用需求的, 瑞泊行业超脑已成功在金融、工业、航运、IT运维等关键行业落地生根, 我们可能是国内最早商业化的, 以大模型为核心的超脑研发企业。瑞泊提供的是嵌入式的服务, 通过联合实验室, 甚至股权投资的方式完全跟客户绑定, 为企业集团、政府机构、城市建设“量身定制”注重降本增效及价值创造的数字化

建设及转型解决方案, 跟瑞泊合作的客户基本是基于长期主义的, 奔着未来十年、二十年一直合作下去的。

《中国经营报》: 对于企业应用大模型的热情, 你觉得有提醒注意的地方吗?

乙壤月: 虽然大模型技术带来了许多重要价值, 但企业在投入使用前应进行充分的成本效益分析。确保技术投资与预期回报相匹配, 避免因过度投资新技术而影响企业的财务健康。同时, 合理设定技术应用的期望和目标, 管理内外部预期, 避免过高期望带来的失望和挫败感。

首先, 企业在选择时应重视产品的质量、安全性和可控性, 避免因技术选型不当而带来潜在的法律风险和安全隐患。尤其对于涉及国家核心产业和关键领域的企业, 使用不可控的开源大

模型和API接口存在较大风险。在当前的国际环境下, 确保技术的安全性和数据的主权尤为重要, 故建议优先考虑国产、可信的技术解决方案。

其次, 避免盲目追求新技术, 企业应深入分析自身的业务需求和痛点, 以确保技术的应用能够带来实际商业价值和社会价值, 避免因追求技术而忽略了技术应用的实际效益和合理性。

再次, 在使用大模型技术时, 企业必须重视数据隐私保护和安全性问题。不论是在中国, 还是在全球范围内, 安全合规的重要性日益凸显。

最后, 技术的适配性与集成问题需关注, 技术集成是实现大模型技术落地的关键步骤, 企业需要考虑大模型技术与现有系统的兼容性, 确保新技术的引入不会干扰原有业务流程的正常运行。

未来20年是AI大模型普及的阶段

大模型是通往AGI(通用人工智能)的终极密码吗? 目前业界仍众说纷纭, 莫衷一是。但是当被问及如何看待以大模型为代表的新一轮AI热潮时, 乙壤月在采访中给出了一个颇为肯定的回答, 他认为“大模型技术的突破, 至少是文明级的技术突破”。

《中国经营报》: 你为什么认为“大模型至少是文明级的技术突破”?

乙壤月: 人类的发展是一个漫长的历史过程, 科学技术的发展对人类文明的形成与演进, 起着至关重要的作用, 科技的发展是叠加性进步且不断加速的。上一次人类文明级的技术突破是蒸汽机, 蒸汽机技术把热能转换成动能, 使人类第一次完成了在生产劳动中以“热力”代替人的“体力”的革命性飞跃; 而这一次大模型技术的突破, 是将“电能”转换成了“智能”, 增加了脑力劳动的不断输出, 是机器智能对人类智能赋能及替代的开端。

说大模型是文明级的技术突破, 也是因为大模型的“涌现能力”, 机器开始能够理解人类的语言并且开始“会说话”了。人与其他动物最本质的区别, 在于人类具有高度发达的语言能力, 语言被认为是人类智能的标志性能, 当

人类没有语言的时候, 我们其实无法去广泛认知这个世界, 语言的边界其实就是世界的边界。正如《人类简史》中所描述——“智人之所以能够成为这个星球唯一霸主, 一个重要的原因是他们发展出了虚拟的语言。”我们常说的大模型全称是“预训练大语言模型”, 随着大模型技术的突破, 人类独有的语言能力被打破, 人机关系呈现出新的特征, 因为机器带来的语言边界的扩大将导致我们认知的世界边界的扩大。机器掌握了语言, 强人工智能时代就来到了。当然, 实现AGI这一过程一定是渐进的、不断迭代的。

《中国经营报》: 瑞泊创立之初, 国内外的大模型领域是一个什么样的状态?

乙壤月: 大模型技术属于人工智能科学中自然语言处理的子领域, 很长时间以来都是比较冷门的, 因为难度大、门槛高, 譬如对于算力、技术、数据的要求都很高, 更关键的是它还走了与其他“门派”大相径庭的技术路线——历史上大部分人工智能研究都非常依赖人类的直觉和经验, 试图通过精心设计的规则和算法来模拟人类的智能, 但它的问题在于人类知识的有限

性以及特定领域的偏见。相反, 大模型是利用海量的算力和数据, 通过让模型自监督学习的方法来取得让人惊讶的成果。强化学习之父、加拿大计算机科学家理查德·萨顿在《苦涩的教训》一书中说: “过去七十年, AI研究走过的最大弯路就是过于重视人类既有经验和知识。”所以大模型研究团队一直是“孤勇者”。

一般机构下不了决心做这种投入很大、很少人选择且前景很不明确的领域, 因为算不过来账。瑞泊开始进行投入预训练大模型技术的时点, 跟国际主流的AI实验室基本是同步的, 当时国内也就大概七八个机构是明确有团队在跟踪这项技术的, 美国的情况也差不多, 但是美国的投入确实要比我们大很多, 差着好几个数量级。我们的感受是这个过程很孤独, 像在茫茫大海中寻找新大陆, 没有信仰是不可能坚持下来的。

《中国经营报》: 你怎么看中国人工智能发展的未来? 中国的企业家应当扮演什么样的角色?

乙壤月: 未来20年将是AI大模型普及的年代, 生活和工作的方方面面都会充斥着大模型。大模型涉及国家安全

与国际竞争, 我国会避免芯片、操作系统这些核心领域“落后”的历史重演, 尤其是避免生态被外部所控, 核心领域不会使用国外或者不可控的开源系统。

人工智能技术的发展并不是孤立发生的, 它需要与企业的核心业务深度融合, 才能发挥出真正的商业价值。这就要求企业家不仅要关注技术本身, 更要深刻理解自己的业务和市场, 从而引领企业在正确的方向上进行技术创新和应用。

我很欣赏美团创始人王兴对于企业经营的一个总结: “我们吃了很多亏, 交了很多学费, 才明白这个世界没有神话, 只有一些很朴素的道理。便宜的打败贵的、质量好的打败质量差的、认真的打败敷衍的、耐心的打败浮躁的、勤奋的打败懒惰的、有信誉的打败没信誉的。”

对于中国的企业家和创业者们来说, 建设人工智能和大模型, 这些道理其实就是: 数据的质量胜于数量、耐心和细致的研究胜过浮躁和急躁、持续的学习和适应能力胜过一成不变的模式。最后, 我想把诺贝尔物理学奖获得者理查德·费曼的一句话与大家共勉: “凡是我还不能创造的, 就是我还不能理解。”

老板秘籍



如何让大模型更好地在各行业落地?

单靠大模型是无法完成那么多智能化的行业应用需求的, 瑞泊提供的是嵌入式的服务, 通过联合实验室, 甚至股权投资的方式完全跟客户绑定, 为企业集团、政府机构、城市建设“量身定制”注重降本增效及价值创造的数字化建设及转型解决方案, 跟瑞泊合作的客户基本是基于长期主义的, 奔着未来十年、二十年一直合作下去的。

对于企业应用大模型技术的建议?

虽然大模型技术带来了许多重要价值, 但企业在投入使用前应进行充分的成本效益分析。确保技术投资与预期回报相匹配, 避免因过度投资新技术而影响企业的财务健康。

首先, 企业在选择时应重视产品的质量、安全性和可控性, 避免因技术选型不当而带来潜在的法律风险和安全隐患。

其次, 避免盲目追求新技术, 企业应深入分析自身的业务需求和痛点, 以确保技术的应用能够带来实际商业价值和社会价值。同时, 企业要合理设定技术应用的目标和期望。

再次, 在使用大模型技术时, 企业必须重视数据隐私保护和安全性问题。

最后, 技术的适配性与集成问题需关注, 企业需要考虑大模型技术与现有系统的兼容性, 确保新技术的引入不会干扰原有业务流程的正常运行。

深度

迈向AGI是一个渐进的过程

人工智能科学从20世纪中叶诞生至今, 经历了几波冷热交替的发展浪潮, 其间机器学习、数据挖掘、知识图谱、专家系统、图像识别、语音识别、自然语言处理等细分领域技术轮番登场。或许是因为长期以来受到科幻小说、影视作品的影响, 人们对科技的发展与进步总是寄予“一步到位”“瞬间变身”的想象, 而事实上, 科技的演进——从大模型这些技术迈向通用人工智能(AGI)是一个渐进式的、进步力量不断叠加的过程, 很多人可能身在其中却感受不到。

乙壤月和瑞泊团队从大模型技术最初的冷门阶段走来, 又在最近一年多来真切感受到行业的火热。他试图站在人类文明发展的历程中来观察, 并基于自己的认知方法论去阐释其大模型是文明级的技术突破论点, 这在众说纷纭的大模型观点交锋中提供了一种新的思考视角。

与此同时, 中国人工智能发展的实际水平究竟如何? 在日益白热化的全球科技比拼中, 我国能否占领高地? 这些最受关注的问题, 是对包括瑞泊团队在内的所有中国AI从业者的灵魂拷问。在采访过程中, 乙壤月多次强调中国发展人工智能的优势与信心, 并且给出了自己的解释与论证。他认为, 几乎所有产业都在或者即将因为人工智能系统及应用而发生深刻的变化, 同时将会创造出全新的服务产业形态和模式。

2024年《政府工作报告》中提出: 大力推进现代产业体系, 加快发展新质生产力。强调要深化大数据、人工智能等研发应用, 开展“人工智能+”行动。这是“人工智能+”首次被写入《政府工作报告》。显然, 人工智能正在成为新质生产力的核心引擎, 成为整个经济社会发展变革的基础性力量。作为中国的AI从业者, 乙壤月提醒要回归到最本质的初心, 重视数据质量, 投入耐心与细致的研究, 朝着社会文明变革的方向脚踏实地前进。

本版文章均由本报记者曲忠芳采访