复旦 管理学家圆桌谈

复旦大學管理学院 SCHOOL OF MANAGEMENT FUDAN UNIVERSITY

三问大模型

2023年，堪称“大模型元年”。

在国内，大模型更新迭代速度一日千里。3月，多家研究机构和科技巨头先后宣布推出大模型，至今已逾百个，可谓“百模大战”；8月31日，首批国产大模型产品获批开放服务，包括商汤科技“商量”、百度“文心一言”等，推动行业再度加速跑。

大模型时代，热度中也应有冷思考。大模型之争，应该拼什么？国产大模型的优势与挑战究竟在哪里？面对新的科技浪潮，我们又该如何治理？本期文汇-复旦管理学家圆桌谈，我们邀请复旦大学管理学院3位专家，从技术、应用及治理层面“三问大模型”。

■本报记者 徐晶卉 张天弛

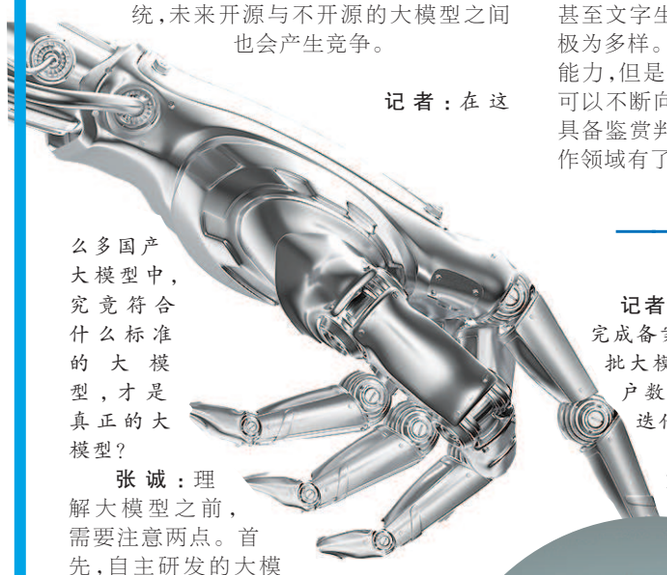
技术之问

记者：大模型的出现意味着什么？

胥正川：ChatGPT的出现引发了人工智能（AI）革命，人工智能从1.0时代跨越到2.0时代，这是一个原则性的变化。在1.0时代，人工智能是单一数据、单一模型、单一任务，处理每一项任务都要设立一个模式进行训练，效率低下且泛化能力严重不足。

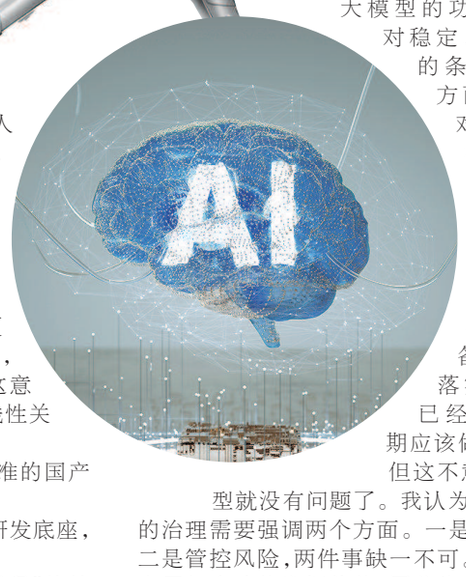
ChatGPT出现后的2.0时代，大模型几乎“什么都懂”，可以把所有通用数据拿来运用，且可以应用于多个任务。在商业领域，大模型应用最大的意义在于降低了人工智能的工程门槛，使人工智能的部署变得简单、成本降低。

从大模型自身能力来讲，目前应用比较成熟的有ChatGPT、“二号羊驼”、文心一言、腾讯混元等。比如，“二号羊驼”是一个开源大模型，可以与其他开发者共享其代码，有点类似于手机中的安卓系统与IOS系统，未来开源与不开源的大模型之间也会产生竞争。



记者：在这么多国产大模型中，究竟符合什么标准的大模型，才是真正的大模型？

张诚：理解大模型之前，需要注意两点。首先，自主研发的大模型，第一个条件应该是在底层技术上有一定创新，如果只是在别人的大模型上“套壳”做微调和应用，那叫“大模型的应用”，不叫自研大模型。其次，大模型并不是仅看参数有多大，虽然从ChatGPT3到ChatGPT4，参数从1750亿增加到了1万亿，但不代表仅凭参数增加就带来同等程度的效果提升。更直观的理解就是：参数提升10倍后，效果提升10倍了吗？显然不是，这意味着参数数量和模型效果并不是线性关系。



在这个基础上，我认为符合标准的国产大模型，至少具备4个特征。

基础层面，企业能够进行自主研发底座，这是最基本要求。

效果层面，它应该达到“智能涌现”的特征，这是生成式AI（AIGC）最显著的特征之一。如果看不到这种特征，我们很难认同参数复杂度就代表大模型，而且，现在认为的1000亿参数的“大”也许再过10年就是“小”，“大模型”本来就是动态的。

技术层面，它应该通过当代的图灵测试。图灵测试是一种测试机器是否具有人类智能的盲盒对比方法，但目前国内很多大模型仅有部分功能可通过图灵测试。当然，图灵测试自身也仅是大半个世纪前考虑的简单思路，自身还需要进一步完善。

学术层面，它应该要能开源，并且能在高等级学术会议或期刊上发表论文，获得学术同行的认可。从这些条件来看，真正的大模型屈指可数。

记者：随着ChatGPT的一路走红，AI在哪些领域已经超过了人类？

张成洪：汉斯·莫拉维克绘制过一份“人类能力地图”，他在地图上标示出在一些领域，例如死记硬背、算术、下棋、问答比赛等方面，机器已经超越了人类。随着时间的推移，像机器视觉、语音识别、语言翻译等领域，机器的发展已经慢慢接近并超越了人类。人类的能力也有其独特之处。例如，在科学、艺术以及AI算法设计等领域，人类可能还拥有一些独特的高峰。

随着ChatGPT的发布，我们也注意到一些变化，新技术具备了通用人类意图理解和思维链的能力，尤其是在与深度学习和生成式AI相结合的情况下，再加上基于人类反馈的强化学习，计算机已经能够完成编程、写作、绘画等任务，逐步接近了人类能力地图上的几座高峰。

> AI治理需要一个体系，只有政策法规层面还没法落地，还需要算法审计层面和合规算法开发层面。特别是AI算法审计需要政府或行业成立相应的监管机构，才有助于实现对AI大模型的真正监管。

张成洪
复旦大学管理学院信息管理与商业智能系教授

以绘画为例，我们普遍认为，艺术创造需要脑洞大开的想象力以及优秀的审美品位。而如今，AIGC可以实现文生图、计算机绘画甚至文字生成视频等，而且在风格和形式上极为多样。虽然大模型一开始不一定有鉴赏能力，但是基于人工反馈的强化学习，AIGC可以不断向人学习判断好坏的标准，越来越具备鉴赏判别能力，这使得机器在绘画和创作领域有了更多的发展空间。

应用之问

记者：8月31日，首批国产大模型产品完成备案并开放服务。过去一个月来，多批大模型陆续获批，这意味着什么？用户数据的反馈，对于AI大模型的更新迭代来说，是一件好事吗？

张成洪：国内的大模型从实验室走向市场和应用，一方面说明大模型的功能变得相对稳定，有了开放的条件。另一方面，一批相对成熟的大模型完成备案，也体现了政府部门的认可，说明在大模型的算法备案、责任落实等方面，已经完成了前期应该做的事。

但这不意味着大模型就没有问题了。我认为，AI大模型的治理需要强调两个方面。一是发挥价值，二是管控风险，两件事缺一不可。大模型只有用起来才能发挥价值，因此绝不能因为大模型不够完善或有潜在风险，就一刀切“关”起来，而应该在使用的过程中不断发现问题，进一步监管。

张诚：大模型产品的合规性通过了备案，意味着企业已经拿到了“软件著作权”，但这并不代表大模型的真正价值和社会影响。它的意义在于，国产大模型的比拼开始从上一阶段的“出生潮”向新一阶段的“应用潮”迈出重要一步。

大模型面向普通用户开放是一个重要的里程碑，它意味着更庞大规模的中文语料，从某种程度来说，也解决了规模化下数据收集的成本问题。但要指出的是，来自用户的数据并不一定是有质量的信息反馈，这就与语料的定义有关。一般来说，只有高质量的信息反馈对于算法才有价值，而并不依赖大规模数据本身。

从已经开放的AI大模型来看，既有企业的大模型，比如百度“文心一言”、商汤科技“商量”、阿里云“通义千问”等，也有不少高校以及科研机构的大模型，比如中国科学院旗下的“紫东太初”、上海人工智能实验室的书生通用大模型、清华系的“智谱清言”等。从技术上说，学院派在一定范围和时间内有领先性，从应用落地上来说则是企业的强项，未来这两类大模型如何发展还有待观察。

记者：大模型“批量上市”，但似乎同质化严重，暂时走在前面的这些企业，要保持领先优势，应该怎么做？

胥正川：现在一些企业开发的大模型都与自身的强项和资源优势相关，如阿里云推出的通义千问就结合了其电商运营的优势。百度推出文心一言，借助了其作为搜索引擎所掌握的大量丰富的通用知识数据。腾讯开发的混元大模型则是基于深厚的媒体舆论数据积累。

今天的人工智能最核心的一句话是“数据驱动”，拥有哪类数据，就掌握了做这种类型人工智能的先机，而数据的数量和质量则直接关系到大模型和人工智能产品的能力。大模型之间的较量，数据也是重要的决胜因素。

记者：除了这项核心竞争力因素，要想取得商业上的成功还需要具备哪些条件？

胥正川：我认为，国内最终能跑出来的大模型并不会太多。首先，从同类型产品中脱颖而出的大模型，要比其他同类数据更全、训练更充分、调优能力更强。而它一旦推出，就会吸引绝大多数的用户选择它，而此时，用户越多的大模型，数据就越多，得到的训练和反馈越多，就会产生“马太效应”。

未来大模型在行业应用方面将有更多的机会。我预测未来将有一批AI公司如雨后春笋般出现，把大模型转化应用于各行各业。打个比方，大模型多是通用模型，就相当于一名“本科生”；未来要为各行各业服务，就需要掌握行业的数据，进行这一行业的定向训练，就进修成一名“研究生”了。

另一层面，在落地应用中，需要大数据、大计算、大内存的大模型部署使用的成本也非常巨大，一些AI公司可以对其进行轻量化处理，以提高部署效率、降低大模型应用的成本。

记者：大模型会生出杀手级应用吗？

张诚：这是全球所有公司都在追问和寻找的答案，从技术上看，大模型的中间层还未成熟，应用层也尚需时日。尽管现在很难预测杀手级应用还有多久到来，但从最近大模型公测引发的全民体验潮来看，爆发点或许正越来越近。

我预计一年之内会有杀手级应用问世，同时，根据AIGC大模型的应用特征，与语言应用相关的所有产品都有可能。不过，杀手级应用会在哪个领域爆发、什么时候爆发，都有随机性。从过去技术潮所诞生的现象级爆品来看，最终应用很难预测，但杀手级应用的诞生，一定是AIGC的技术用户内心需求释放之间的耦合，比如密切符合社会情感需求的应用，在国外Character AI、Inworld等模式也在验证中。

值得一提的是，在2B行业，第一批被颠覆的行业已经在内容领域出现了。比如，随着AI绘画工具Midjourney、Stable Diffusion等应用的走红，游戏原画行业已经迎来变革，游戏公司的成本大幅下降。轻度内容创新的行业，比如广告文案设计等，也受到很大影响。

记者：那么多大模型会是一种资源浪费吗？这个行业会不会最终走上兼并收购之路？

张诚：这个问题要看以怎样的目的为标准来判断。从短期看，“百模大战”“千模大战”的出现肯定是资源浪费，但从长期进步来说，我们必须允许产品的试错——最后一个现象级产品的问世，一定是多种尝试后的结果，这在基础研发领域更为明显。长期来说，大模型的白热化竞争，终将带来技术创新，带来新的变化、新的融合。

移动互联网时代，美团点评、滴滴快的等企业的兼并收购案例耳熟能详，企业“跑马圈地”快速成长，通过资本的力量“大鱼吃小鱼”，最终形成赢者通吃的格局。但是在AI大模型领域，靠兼并收购的可行性比较小。原因很简单，移动互联网时代的应用最终是市场渠道的竞争，而渠道是线下的物理空间，收购兼并是资源整合的必然出路。而反观大模型，它属于典型的数字技术，具备赢者通吃的特性，边际成本接近零的特征，而且，企业的大模型底座不同，这意味着，相同类型、赛道的企业必然只有“你死我活”，很难“兼容”。但从收购方式看，科技巨头投资或者收购一些初创型AI大模型企业的通道仍然畅通，比如在国外，Databricks花13亿美元收购大语言模型初创公司MosaicML，以增强研发实力。

记者：未来大模型生态会是什么样的？

> 从短期看，“百模大战”“千模大战”的出现肯定是资源浪费，但从长期进步来说，我们必须允许产品的试错——最后一个现象级产品的问世，一定是多种尝试后的结果，这在基础研发领域更为明显。

张诚
复旦大学管理学院信息管理与商业智能系教授

胥正川：我认为会出现头部企业，其他组织围绕头部企业共同构成一个大模型生态。在“马太效应”中，脱颖而出的一到两款大模型会不断强化，它所带来的巨大流量便会吸引其他组织，后者会根据前者的需求进行一些具体功能的补充。此时，大模型的头部企业就成为了一个“指挥者”，调动周围的小模型，共同满足用户的需求响应，由此形成一个以大模型为主导的生态体系。

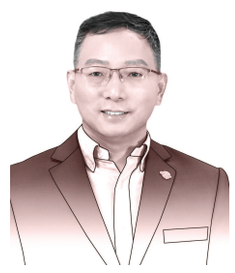
> 当下，人工智能最核心的一句话是“数据驱动”，拥有哪类数据，就掌握了做这种类型人工智能的先机，而数据的数量和质量则直接关系到大模型和人工智能产品的能力。大模型之间的较量，数据也是重要的决胜因素。

胥正川
复旦大学管理学院信息管理与商业智能系副教授

治理之问

记者：AI大模型有远虑和近忧吗？

张成洪：今年上半年，包括马斯克在内的千名全球科技人士联名发布了一封公开信，呼吁暂停推出更强大的人工智能系统。世界上最懂技术的人如此“恐惧”，有一定的远虑。在我看来，一方面，AI已经能够担任普通程序员的工作，当它某一天可以自己编写AI程序，可能就来到“智能爆炸”的奇点。另一方面，人类社会越来越依赖AI，因此必然希望AI是和人类的价值观对齐的，是有科技伦理的，未来AI可能越来越复杂和庞大，如果它的价值观出现了偏离，很难修正和管理。



至于近忧，它已经出现了。比如算法歧视，包括搜索排名与个性化推荐被操纵；比如信息茧房，网站会根据用户的喜好，只呈现“这类信息”和“这类观点”。而当AIGC出现之后，已经出现了因为“AI换脸”产生的欺诈案例，大模型还涉及到知识产权、个人信息泄露等问题，这些近忧都需要得到重视。

记者：在开发和调训大模型的过程中，怎么让大模型具有“道德感”呢？

胥正川：我认为，大模型本就是有“道德感”的，它只是没有自主意识。而我们所讲的“道德感”其实更偏向于价值取向。在开发和调训大模型的过程中，开发者可以为其灌输价值观，也必须对大模型的价值取向进行干预。

值得注意的是，大模型的价值观念并不是开发者强加上去的，而是由数据驱动的。更类似于人类教育下一代，在日常的言行间潜移默化地引导，让小朋友观察、模仿、从而得到自己的知识。大模型也是如此，在它基于数据的自主学习中，它便自然而然地形成了一套价值观。不过“数据驱动”也无法100%保证大模型不会出现道德偏差。

记者：AI算法治理包含了哪些层次和方面的内容？

张成洪：AI算法治理至少包含3个层次。第一个层次是算法治理的政策法规，由政府主管部门制定政策，给予指导。第二个层次是算法合规性的审计，依据法律法规或者行业要求对企业展开检查和监管。首先，要把政策法规的原则和要求，落实成每个场景的不同指标；其次，要解决指标如何度量、如何计算以及合规标准问题；再次，形成结果后，还需要有工具来展示和说明问题。这个层次中，不仅需要企业在内部做，也可能需要行业或国家成立专门的监管机构去做。第三个层次是开发出合规合法的AI算法，这事实上就对AI开发人员、算法工程师提出了要求。

另一方面，要实现AI算法治理，我们至少要在隐私保护、可解释性、公平性、稳定性、安全性等方面审查与保障，同时穿透上述3个层次，才能较好实现AI算法的“治”与“理”。

记者：在AI治理方面，有一种说法是AI大模型对抗AI大模型，您觉得这种方式可能吗？

张成洪：AI治理需要防范各种风险，而完全靠人工去发现AI的问题不太可行，需要称手的工具与程序，“人”“机”协同完成对AI的风险监测与防范。比较典型的情况是，利用基于强化学习的AI审计程序，自动检测出被审计AI算法的风险点或问题样本，可以说是用AI对抗AI；类似的，如果我们要检测大模型的幻觉问题、知识产权问题、信息泄露问题等风险点，也需要运用AI算法，甚至可以利用AI大模型去实现对另一个AI大模型的审查，即AI大模型对抗AI大模型。

记者：AI大模型出现之后，治理的迫切性超过以往任何一次科技革命，我国已先后推出多个监管举措。在您看来，监管是需要“及时”还是应该“让子弹再飞一会儿”？

张成洪：就像汽车不能没有刹车就上路一样，监管肯定需要“及时”，才能及早避免错误与风险。但治理始终要平衡“用”和“管”两件事，大模型不用肯定没问题，用了就一定会出现一堆问题，不能因为监管措施不够完善，就不让用；出问题就需要“及时”监管与纠正。因此，这不应该是一道选择题，而是监管既需要“及时”，更需要在使用过程中动态监管。

从2021年起，我国就相继颁布了一系列关于算法治理的法律法规，ChatGPT问世后，我国针对AIGC工具迅速推出两部专门性立法，分别是《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》，前者是算法备案，后者是开放备案，目前已有多批次大模型分别通过备案。总体来说，我国在监管政策上相当重视，做到了与时俱进，这是亮点。

不过，如前面讲的，AI治理需要一个体系，只有政策法规层面还没法落地，还需要算法审计层面和合规算法开发层面。特别是AI算法审计，需要政府或行业成立相应的监管机构，就像已有的财务审计、软件测评、安全测评等机构一样，才有助于实现对AI大模型的真正监管。

本版图片 视觉中国
人像素描 张旭菁