

# AI 大模型时代 开局就是“百模大战”

家家都是纸上谈兵，孰优孰劣难说清

自从 OpenAI 于 2022 年 11 月推出 ChatGPT 后，一场波及全球科技界的“AI 海啸”就此爆发。

在国内，今年以来市场隔三差五有一款大模型产品问世，在百度文心一言最先公开邀测后，阿里、京东、华为等互联网大厂，商汤、昆仑万维、出门问问、科大讯飞等 AI 公司先后发布大模型；毫末智行、知乎等垂直领域的头部企业也发布相应的垂直大模型。

据不完全统计，截至目前中国已有超过 40 家公司、机构发布了大模型产品或者项目。再加上各类名头响亮的创业团队纷纷公布的开发计划，一场 AI 大模型“百模大战”正席卷全国。

在国际市场，大模型同样热度不减。谷歌在短短几个月时间内连续迭代，日前发布了新的大语言模型 PaLM 2，并将其融合进了 Gmail、搜索、Pixel 等一系列产品中，誓与 ChatGPT 决一雌雄。一度高喊“暂停”大模型研发的马斯克也有了实际行动，成立大模型公司 X.AI，并将新计划命名为 TruthGPT……

突然涌现如此之多的各式各样大中小模型，旁观者难免有“乱花渐欲迷人眼”之感，很自然会想，这些大模型能不能比一比、测一测？到底孰强孰弱？哪些是能实质性地帮助到人类的真 AI？哪些又只是花架子、“纸上发布”？或者只是拉提股价的噱头话？中国的这些大模型，跟 ChatGPT 等当今最尖端技术相比，差距有多大？

■新快报记者 郑志辉

## 行业现状

### 缺乏标准，关公混战秦琼

那些急切想看到一个明确结论的读者要失望了，原因是目前全球对于大模型仍处于早期的探索阶段，尚未形成权威、通用的行业标准。这也导致了当前出现的一种情况是，一些厂商对于该如何评判大模型各有说法，怎样对自家有利怎么来。

科大讯飞董事长刘庆峰近日在发布星火大模型时谈到：“今天要看一个大模型系统到底好不好，首先要看它是不是能解决刚需、是不是真的有用，而不是一个简单的单点测试。”他认为，应该从 7 个维度衡量 AI 大模型产品是否达成通用人工智能 (AGI)：文本生成、语言理解、知识问答、逻辑推理、数学能力、代码能力、多模态能力。

按照这样的标准，刘庆峰声称，讯飞星火在长文本生成、数学能力和泛领域开放式知识问答三大能力已超越 ChatGPT。

但其后发生的两件事，让外界对讯飞和星火的能力有所保留。

5 月 9 日晚，一段关于讯飞星火的群聊记录和截图在各个群广为流传：得到内测资格的提问者问到星火提供的 API 接口是否 OpenAI 的？星火大模型回答道，“是的，我是由

OpenAI 开发的”。紧跟截图后面，是群内的议论，指向星火大模型疑似套壳 OpenAI？

此外，讯飞发布会后，一份自称“中文通用大模型综合性评测基准 SuperCLUE”发布，其评测排名榜单中，刚发布的讯飞星火大模型在总榜单和子榜单中均排名第一位，仅次于 GPT-4、GPT-3.5，而百度文心一言却排在榜单最后。

很快，有网友揭秘称，发布榜单的只是一新近冒起的国内民间组织，其测评成员顾问中排第一位的人士来自哈工大讯飞联合实验室，这样的测评难以令人信服。

另一边，在 5 月 9 日的技术交流会上，百度也给出了大模型服务的三个评判维度：大模型本身的能力、大模型企业服务的能力、全栈技术积累程度。

思必驰联合创始人兼首席科学家俞凯表示，目前去评判各模型能力如何以及好坏，其实不是合适的时间点。在他看来，业界现在更关注大模型是否有足够的“泛化性”，即广

## 纸上谈兵

### 大厂才具备追赶 chatGPT 实力

囿于各种商业原因，对于各公司大模型实际的数据、测试反馈指标、投入的资源情况，乃至用户数据等，外界很难全然知晓，自然也很难对其实力情况做出科学的判断。

虽然如此，根据各家所公布的自家大模型的一些最新参数、资料等，做一些简单的纸面对比，还是能得出一些有意义的结果的。最新的一份来自于中国移动研究院旗下的中移智库。

中移智库指出，截至 4 月 20 日，国内涉足人工智能大模型训练的机构，主要分为大型科技公司、科研院校和初创科技团队三类。

从大模型的布局体系来看，科技大厂在算力层、平台层、模型层、应用层进行了四位一体的全面布局。百度、阿里、华为三家均从芯片到应用进行自主研发的全面布局，如百度的“昆仑芯+飞桨平台+文心大模型+行业应用”、阿里的“含光 800 芯片+M6-OFA 底座+通义大模型+行业应用”、华为的“昇腾芯

片+MindSpore 框架+盘古大模型+行业应用”。垂直行业科技企业和科研院校，主要以研发大模型算法及细分领域应用为主，自有算力相对薄弱，很少涉及芯片领域自主研发。

从大模型参数量看，科技大厂的参数量远大于科研院所：阿里通义千问大模型参数在 10 万亿级以上、腾讯混元大模型和盘古大模型参数量均在万亿级以上、百度文心一言大模型参数量在 2 千亿级以上、京东言犀大模型的参数量为千亿级；垂直行业科技企业已经上线的参数量普遍在千亿级以上；而科研院校大模型的参数量在千亿级及以下。

从大模型应用方向看，大部分企业前期以内部应用为主，后续主要向 B 端企业拓展服务，预计少数企业将在 C 端市场形成规模。目前，百度文心大模型、华为盘古大模型、中国科学院紫东太初大模型均在 B 端垂类市场积累了标杆应用案例，腾讯混元大模型、阿里通义大

### 大模型发展不应一味追求参数数量

需要指出的是，大家在看上面的对比表时，部分数字还需理性看待。

一般认为，算法、算力、数据是 AI 大模型研发中必不可少的基础要素，也是制约大模型成果水平的关键因素。在这一轮“百模大战”中，不少企业都宣称其推出的 AI 大模型数据量超大、参数达千亿条，有的比肩甚至超过 ChatGPT。

对此，ChatGPT 之父萨姆·奥特曼已经明确表示，大语言模型的规模已接近极限，并非

越大越好；大模型发展不应一味追求参数数量，未来大模型或以多个小模型相互协作的方式运行。

商汤科技首席科学家王晓刚用了一个比喻来解释这一情况，“大模型你可以把它想象成为一个非常有天赋的运动员，而人就是教练。教练在教这样的运动员的时候，并不需要一招一式地去示范动作，而是通过一些方法论的输出，就能让这个运动员解锁完成很多新动作。”

泛使用，但从产业角度而言，国内大模型的用户量级也还未达到泛在化。在未达到广泛的通用性之前，以通用性的标准去做评判，还需慎重。

尽管如此，复旦大学教授、上海市数据科学重点实验室主任肖仰华认为，从诊断与评测两个视角，建立与健全大模型的诊断与评价体系，建立大模型的评测基准，是大模型产业发展所亟需的，是形成差异化发展路线的关键，具有战略意义。

模型则更多聚焦公司自身业务。

而在 C 端市场应用方面，百度文心一言、阿里通义千问、腾讯混元助手三类大模型最有可能向此方向拓展，但目前只有百度文心一言大模型正在进行友好客户测试，阿里通义千问大模型则计划在 9 月份进行公测，腾讯混元助手大模型则处于计划开发状态。

中移智库还表示，从大模型业界评估看，国内大模型与 GPT-4 有较大差距，但科技大厂具备追赶实力。百度、阿里巴巴、腾讯、华为四家在大模型研发投入、技术能力和人才团队等方面综合实力较强；商用推进方面，四家企业均依托现有业务领域更容易形成大模型应用规模效应。

另外，商汤的 AI 大装置“SenseCore”是亚洲最大的算力平台之一，可以同时支持 20 个千亿级参数的大模型训练，最高可支持万亿参数超大模型的训练。

“而不管是中模型或者小模型，它是一个专属的模型，可以把它理解成是一个比较刻苦但天赋不是很高的运动员，要教他就有两种方式：一种是教练人工去标注大量的数据，一招一式地展示给他看；第二种方式，就是让天赋很高的运动员一遍遍地吧新动作展示给他看，情形就是一个很强的大模型自动产生了很多数据，再去喂给小模型让它去学，那就形成了专属领域里面的模型。”