

近期，字节、百度、阿里云等多家企业宣布生成式人工智能服务集中调价——

大模型为何纷纷降价

本报记者 王俊岭

“比行业同类产品便宜90%”“两款产品免费”“全线降价”——近日，字节、百度、阿里、智谱AI等国内大模型领军企业纷纷宣布调降生成式人工智能大模型（以下简称“大模型”）产品和服务使用价格，不少大模型产品价格降幅达50%以上，有的产品价格降幅甚至超过90%。

消费者如何看待大模型降价？“降价潮”对行业发展有何影响？带着这些问题，记者进行了采访。

门槛更低，服务触手可及

家住河南开封的小伙子刘多经常使用文心一言等生成式人工智能大模型起草发言提纲、制作展示材料。最近，他发现不少大模型的服务价格明显降低。

5月21日，阿里云宣布通义千问GPT-4级主力模型Qwen-Turbo、Qwen-Plus等9款主力AI大模型全面降价。其中，Qwen-Long API输入价格从0.02元/千tokens降至0.0005元/千tokens，最高直降97%。据估算，降价后1元钱大约可以买到200万字符，相当于5本《新华字典》的文字量。

字节跳动旗下火山引擎推出的豆包大模型以“更强模型，更低价格，更易落地”为服务特色，降价后豆包大模型最强版本的主力模型推理输入定价只有0.0008元/千tokens。

对于大模型服务价格的降低，用户特别是年轻用户普遍持欢迎态度。

“我在开发编程的过程中经常会遇到一些问题，这时大模型可以给出很多解决灵感。作为经常使用大模型的人，我对于价格比较敏感。大模型降价可以让人有更多的选择机会和更低的使用门槛，同时能显著提高工作和学习效率。”在北京一所高校读计算机专业的冯业彭说。

在湖北武汉从事网络运营工作的罗怡然说，近几年国产大模型的使用体验不断提升，能够满足自己寻找策划灵感、收集背景资料等现实需求。平时，她愿意“阶段性付费”购买更多优质高效的大模型服务。在她看来，在应用场景不断扩展的时候，适时下调价格是好事也是趋势。

不过，也有消费者面对“降价潮”态度较为保守，认为最关键的是服务品质。

那么，大模型自己又是如何看待“降价潮”的呢？记者以“如何看待人工智能大模型使用价格大幅降低”为题向部分大模型进行了“咨询”。



第七届中国建设峰会上，现场体验区准备了多项互动体验项目，涵盖AI大模型应用实践、各领域数字成果等内容，大批市民前来沉浸式体验5G、虚拟现实、增强现实等技术成果。图为市民在该峰会体验区与仿生机器人进行互动。

谢贵明摄
(人民视觉)

2024年中国品牌日活动现场，观众在了解中国电信的星辰通用大模型。

王初摄(人民视觉)



文心一言给出的回复是：随着算法优化、硬件性能提升以及数据资源的丰富，大模型的训练成本逐渐降低，使得厂商有能力提供更低价格的服务。各大厂商为了抢占更多市场份额、吸引更多用户，将降价作为一种有效的竞争策略。为了保持竞争优势，厂商将不断对大模型进行优化和升级，提高模型的性能和准确性。同时，厂商还将加强与其他技术的融合创新，推动人工智能技术的不断发展。总之，人工智能大模型使用价格的大幅降低是市场竞争、技术进步和客户需求共同作用的结果。

360智脑认为，使用成本的降低，会进一步加速人工智能在各行各业的普及和应用。不过，这可能会让企业和开发者过度依赖云服务提供商，从而在一定程度上失去自主性。尽管总体而言，大模型降价是积极趋势，但人们还是应该关注随之而来的各种挑战，并采取相应的措施来应对。“未来大模型价格可能会呈现逐渐下降的趋势，但这种下降有望在一定水平上趋于稳定。”360智脑预计。

抢抓市场，加快创新步伐

一顿秒创、抖音剪映、讯飞公文写作助手、腾讯混元……如今，任意打开一个网址导航，都可以找到许多各具特色的大模型产品，内容涵盖写作、办公、视频、音频、绘画、设计、编程等方面。

参与“降价潮”，大模型服务提供商们有着自己的考虑。

有的希望以此探索更多应用场景。5月21日起，百度智能云千帆大模型平台ERNIE Speed、ERNIE Lite两款主力模型全面免费。这两个模型在文案创作、聊天对话、知识问答、内容续写等场景中有非常高的客户调用量。百度集团执行副总裁、百度智能云事业群总裁沈抖对本报记者说，人工智能发展需要新的开发范式、新的技术工具做支撑。推动部分产品降价以及一些主力大模型免费，是为了最大限度降低企业创新

试错成本。

“价格只是消费者选择大模型的一个考量因素。比起价格，更重要的是模型效果、应用场景、已落地案例。这几年，百度一直在用技术手段降低企业使用大模型的成本，尤其是业界最高效的AI异构算力管理平台和‘多云多芯’。我们希望通过降价让大家创新的胆子更大一些、步子更快一些，快速探索并复制更多大模型可能的应用场景。”沈抖说。

有的着眼释放市场潜力。

阿里云自研的通义大模型家族已形成基础大模型及系列行业模型的矩阵。不到一年时间，通义先后开源数十款不同规格的模型，包括大语言模型、多模态模型、混合专家模型、代码大模型。阿里云资深专家朱迅说，降价是希望让更多的普通用户和中小企业用上大模型，加速AI应用市场潜力的释放和爆发。

朱迅说，对阿里云来说，用户每次使用大模型所需的成本主要是模型推理过程中消

耗的算力资源成本，在公共云上这一成本远低于用户在本地进行私有化部署。“我们基于自研的异构芯片互联、高性能网络HPN7.0、高性能存储CPFS、人工智能平台PAI等关键技术和产品，构建了极具弹性的AI算力调度系统，大幅降低了模型推理成本。公共云带来的技术红利、规模效应、成本优势，是我们降价抢占市场的底气。”朱迅说。

有的聚焦全新价值创造。

火山引擎总裁谭待向本报记者介绍，目前火山引擎提供全栈大模型服务，包括AI算力底座、火山方舟大模型服务平台、火山引擎智能数据洞察DataWind AI助手等，可以为互联网、金融、汽车、大消费、智能终端等各行各业提供服务。降低大模型的价格可以让更多企业和开发者轻装上阵，在应用中创造更多意想不到的全新价值。

“降低成本是推动大模型快速进入‘价值创造阶段’的一个关键因素。火山引擎对大模型的超低价，来源于我们有信心在技术上优化成本。”谭待说，通过对模型结构优化调整、将单机推理演进到分布式推理、把不同负载的推理混合调度等，可以让各种底层算力用得更好，助力企业用户以更低成本探索数字化智能化转型，从而在各自领域进一步开展创新、提高生产力。

布局长远，赋能生产生活

“降价潮”背后，是很多大模型提供商乃至广大互联网企业对“人工智能赋能生产生活”的努力与期盼。

智谱AI首席执行官张鹏说，AI技术发展进入了一个全新阶段。“技术曲线的提升”和“让更多人享受到AI强大生产力”之间，还有非常大的差距。如何把强大的大模型能力真正用来帮助人们工作、学习和生活，还有很多工作要做。

伴随着功能更强大、价格更亲民，大模型走进生活的步伐正在加快。

蚂蚁集团首席技术官何征宇介绍，蚂蚁大模型坚持面向产业、全栈布局，推出3个“智能管家”创新应用产品。其中，生活管家“支付宝智能助理”，连接了400万家机构小程序和8000项数字生活服务，不仅“有脑有嘴能对话”，还“有手有脚能办事”，食住行游购娱等生活服务均可问可办；金融管家“支小宝”，能通过文字和语音对答提供行情解读、投资分析、保险配置、政策科普等理财服务，让投资者拥有一个“私人理财专家”，目前已有4300万用户体验；即将推出的医疗健康管家运用了与亚运会“数字人”同款的技术，让患者在就诊全程都有贴心的数字陪诊员。“只有当一个相机变成‘傻瓜相机’之后，才会全面进入普通人的生活。大模型技术也应该像其他先进技术一样，为每个人服务。”何征宇对本报记者说。

大模型的这一轮“降价潮”将给行业带来哪些影响？

北京市社会科学院管理研究所副研究员王鹏在接受本报记者采访时说，随着大模型技术的不断发展和优化，模型训练所需的边际时间和资源也逐渐减少。特别是算法改进、计算效率提升、硬件成本下降都为大模型降价提供了可能。从这个角度看，国内外大模型的这一轮降价潮内在原因基本相似，各大厂商都旨在通过降价来推动人工智能技术的普及和应用。

一方面，降价使得大模型服务更加亲民，可以有效扩大市场规模和用户基数；另一方面，短期内大幅降价也会加剧市场竞争，促使各大模型提供商不断提升服务质量来留住用户。随着时间的推移，一些效率低下或资金不足的商家可能被淘汰出局，一些有特色但又难以做大的商家可能被吸收整合，实现“聚小成大”甚至“强强联合”。这些都是行业放眼长远、向前发展必须经历的过程。”王鹏说。

专家普遍认为，大模型行业未来还需要探索多元化应用场景、注重数据质量与隐私保护、增加衍生收入来源、完善用户反馈机制，最终实现更加健康、更可持续的发展。

北方地区出现大范围高温天气

本报记者 李红梅

6月8日起，京津冀、河南、山东及苏皖北部等地高温迅速发展。气象监测显示：9日，北京、天津南部、河北中南部、山西西南部、山东、河南东部和北部、苏皖北部及新疆东北部等地区出现35—39摄氏度高温天气。10日，河北中南部、山西西南部、河南中北部、山东大部、苏皖北部以及新疆吐鲁番等地部分地区出现35—39摄氏度高温天气。

9日18时，中央气象台升级发布高温橙色预警，当天，中国气象局启动高温四级应急响应。10日18时，中央气象台继续发布高温橙色预警。中央气象台预计至20日，北方地区高温将继续发展，华北中南部、黄淮大部、山西南部、陕西关中等地35摄氏度以上高温日数有6—8天，河南等地部分

地区可达9—10天。预计至14日，华北、黄淮、苏皖北部、湖北西部和北部、陕西、内蒙古西部和东南部、新疆盆地地区等地日最高气温在35摄氏度以上，河北中南部、河南中北部、山东西北部和中部等地部分地区可达40—42摄氏度；其间，河北、河南、山西、山东等地部分地区的日最高气温可达到或突破历史同期极值。预计15日至16日，上述地区高温范围和强度有所减小，但河南等地高温仍将持续；17日至6月下旬前期，华北、黄淮等地高温天气将再度发展，部分地区气温可达40摄氏度以上。

中央气象台首席预报员符娇兰介绍，近期华北、黄淮等地受高压脊天气系统控制，天空云量较少，在晴空辐射和下沉增温的共同作用下，促进了

高温发展。本轮高温天气是北方地区今年以来影响范围最广、强度最强的高温过程，河北、河南、山东等地高温持续时间长、日最高气温有极端性，部分地区发生中暑气象风险高，建议公众减少户外活动，做好防暑降温措施。

“全球变暖的大背景和大气环流的异常，是造成高温天气的直接影响因素，对我国来说也是如此。”国家气候中心首席预报员郑志海介绍，今年夏季，我国的高温天气出现相对较早，高温天气过程比较多。

数据显示，今年春季我国平均气温为12.3摄氏度，较常年同期偏高1.4摄氏度，为1961年以来历史同期最高。国家气候中心预计，6月我国大部地区气温接近常年同期或偏高。华北、黄淮等地气温偏高，高温日数较常年

同期偏多，阶段性高温热浪风险较高。

据监测，厄尔尼诺现象将结束，拉尼娜现象将出现，是否会对我国夏季气温有影响？郑志海介绍，拉尼娜对全球和我国的气温影响有非常大的区域差异和季节差异，从目前的预测来看，进入拉尼娜状态要到夏季后期，对今年夏季气温的影响相对不大。

9日，中国气象局公共气象服务中心发布今年第一期高温中暑气象预报。中国气象局公共气象服务中心高级工程师梁莉提示，高温中暑容易导致人体体温调节功能失衡，增加中暑的风险。高温天气期间，公众需注意尽量避免在午后气温最高的时段进行户外活动，外出时应注意及时补充水分，防止中暑和热射病的发生。



6月9日，位于河北省秦皇岛市的中交一航局五公司船机中心船舶修造基地内，4000匹级多功能拖轮“一航津远”顺利进入渤海湾中。“一航津远”应用AI技术，具备全回转自动舵功能、多元感知航辅功能、可视环船影像功能等，极大提升拖轮的操作性能。因为多功能拖轮“一航津远”。
曹建雄摄(人民视觉)