

OpenAI拒绝中国开发者，“百模大战”全面进入下半场

近日，不少使用 OpenAI 的开发者都收到了官方邮件，邮件中称检测到开发者的账号存在 OpenAI 目前不支持的地区的 API 流量。在委婉的解释了一下官方的政策后，OpenAI 开门见山表示：

目前，在 OpenAI 支持的 188 个国家和地区中，并不包括中国，这也意味着以上区域的开发者想要继续接入 OpenAI 的 API，需要承担更多额外的风险。根据 OpenAI 的使用条例来看，违反条例的账户会遭到诸多限制，比如暂停提供服务和账号封禁等。

OpenAI 对违规账号的封禁并非近期才开始，实际上早在 ChatGPT 4.0 发布后不久，OpenAI 就已经在间歇性封禁部分来自受限区域的账号，只不过对于通过服务器跳转等方式访问的 API 流量，OpenAI 则是采取了睁一只眼闭一只眼的态度。

本来大家心照不宣的状态，如今突然结束，确实是让许多开发者都感到措手不及。有些开发者更是直言，OpenAI 的警告信已经让自己的项目面临暂停的风险，后续会根据具体的情况，考虑实际的应对措施。

OpenAI 的这轮举动，会对中国的 AI 开发造成怎样的影响？

比开发者更慌的，是“套壳大模型”

虽然，国内的一众 AI 模型都表示自己达到或接近 ChatGPT 4.0 的水平，但是也不得不承认，目前 ChatGPT 4.0 仍然是全球最佳的 AI 大模型，在通用领域暂时还没有真正的对手。

而在中文领域和部分特定领域中，国内的文心一言、通义千问等 AI 模型确实拥有了与 ChatGPT 4.0 抗衡的实力。对于多数国内的 AI 应用开发者来说，以上 AI 大模型都可以满足日常的开发需求。

不过，对 OpenAI 需求最大的企业，其实恰恰就是 AI 大模型的开发公司。目前国内除了百度等头部企业在推动 AI 大模型的研发外，还有大量的初创公司也在进入 AI 大模型赛道。为了能够高效训练自己的 AI 大模型，它们往往会选择用 ChatGPT 4.0 的回答作为参考答案，调整自己的 AI 大模型。

在这些企业的项目中，ChatGPT 4.0 其实更像是作为“导师”而存在着。而 OpenAI 的封禁措施，显然会对这些企业的 AI 大模型训练项目造成影响，导致项目进度被拖慢，被迫投入更多的服务器资源和人力。

此外，影响最严重的恐怕就是一些“套壳”AI 大模型。作为目前投资市场上最热门的领域，AI 大模型并非全部都是“真·自研”，其中并不乏一些滥竽充数的套壳企业。何为“套壳”？简单来说就是通过一些手段，将 ChatGPT 的回答包装成自己的 AI 大模型的回答，以此来骗取投资。

这些企业往往是对 OpenAI 依赖度最高的，因为如果改换其它国内的 AI 大模型，很有可能会面临法律层面的风险。而 Gemini、Claude 等海外 AI 大模型，性能并不比国产 AI 大模型好到哪里去，对投资者的吸引力远不如“ChatGPT”。如此看来，OpenAI 或许还能帮国内清掉一些滥竽充数的 AI 企业。

OpenAI 的封禁，对于国内的 AI 开发者来说确实会造成一些麻烦。但是 OpenAI 并非不可替代的，在常用领域中，国产 AI 大模型的如今已经可以做到 80% 乃至 90% 的替代，随着后续国产 AI 大模型的迭代升级，相信效果也会变得更好。

虽然还是有不少人对于 OpenAI 的做法愤愤不平，但是事已至此，我们显然也没必要用热脸去贴冷屁股。而且过去的经验告

诉我们，自力更生才是最佳的选择，与其想办法绕过限制，不如与国内企业一起将国产 AI 做大做强。

所以，现在最大的问题其实是：国产 AI 大模型都有哪些可以替代 OpenAI？他们各自的优势、成本，以及目前的迁移方案都有哪些？

“接回”开发者，国产大模型拼了

OpenAI 拒绝中国开发者，是国产大模型供应商千载难逢的机会。

鉴于 OpenAI 是先行者，因为很多开发者笃定地认为 GPT 就是比国产大模型强大许多，国产大模型再努力都没证明自己的机会。此外，也有开发者对雷科技 (ID:leitech) 表示，虽然此前已经看到智谱 AI、百川以及 Kimi 们的进步，但由于迁移成本高，因此也一直没有动力更换底层大模型。

现在，中国开发者被迫要迁移回国产大模型。抓住这一机会，国内的多家主 AI 大模型都纷纷发出“招揽令”，并且给出各种优惠政策和迁移方案，让开发者能轻松转移自己的 AI 项目“回国”。国产大模型提供搬家服务，“接回”开发者，既是雪中送炭，也是顺势而为。那么，哪家大模型能抢到被 OpenAI 抛弃的开发者呢？

1、文心大模型：0 成本迁移，轻量级产品免费用

百度智能云发布名为“故乡的云”的国产大模型普惠计划，即日起 (6 月 25 日) 为所有新注册的企业用户提供以下优惠政策：首次调用文心旗舰模型不收费，并且赠送 ERNIE3.5 旗舰模型的 5000 万 Tokens 包，此外还提供免费模型精调训练、零成本 SDK 迁移工具和专家服务。

其中 ERNIE3.5 旗舰模型是目前文心大模型的一个重要版本，支持 5K tokens 输入和 2K tokens 输出。相较于上一个版本进行了全方面的优化与升级，能够满足复杂任务场景的使用需求，并且在语义、语境理解等方面的表现远超其他版本，更适合用于虚拟助手、在线客服、社交媒体等场景。

在综合能力测试中，ERNIE3.5 旗舰模型的评分超过了 ChatGPT 3.5，并在中文测试中超过了 ChatGPT 4.0。另外，得益于百度在搜索领域的权威地位，ERNIE3.5 旗舰模型在 AI 搜索等领域的表现优于其他大模型，十分适合有实时更新数据需求的 AI 应用和开发项目。

此外，文心大模型还提供了 ERNIE Speed、ERNIE Lite、ERNIE Tiny 三个完全免费的 AI 模型，这些 AI 模型可以被部署在轻量级的使用场景中，为企业节省成本，并且也能够获得超过传统智能服务的对话、问答体验。

如果你想更详细了解文心大模型的相关方案，可以前往百度智能云的微信公众号查看相关文章。

2、通义千问：价格只要 GPT-4 的 1/50

阿里云的 OpenAI 迁移活动，那叫一个言简意赅，全文不到 50 个字就将两个要点全部展示出来：1、AI 大模型实力并不比 GPT-4 差；2、价格仅为 GPT-4 的 1/50。

阿里云还为 OpenAI 用户提供百炼专属迁移服务，确保开发者可以高效、轻松地将 API 接口迁移至阿里云，此外还将提供 2200 万免费的 Tokens 额度。作为通义的超大规模语言模型增强版，Qwen-plus 的 API 接入支持高达 30K Tokens 的文本输入，可以更好

地处理超长篇幅的内容。

同时，阿里云百炼大模型平台还提供针对多种应用场景的不同 AI 模型，支持从文生图、AI 客服、AI 办公到 AI 视频生成的多种用途，使用专门的 AI 大模型不仅可以提高 AI 创作质量，而且也可以降低对应的成本。

此外，阿里云还提供 AI 大模型托管等多种服务，有兴趣的朋友可以去阿里云微信号或者官网查看。

3、讯飞星火：送 Lite 版、送 Tokens

讯飞星火也是我们的老朋友了，毫无疑问的国产 AI 大模型第一梯队尖子生，在 AI 语音、翻译等多个领域更是名列前茅。此次，正值讯飞星火 V4.0 版本正式发布，讯飞星火推出了全新的普惠政策：星火 Lite、Pro、Max 三款不同的模型，分别提供永久免费、1 个月免费和赠送 1 亿 Tokens 的活动，并且就连最新的星火 4.0 Ultra 版也提供买一送一的大促。

不得不说讯飞星火的诚意已经快要溢出屏幕了，除最新的星火 4.0 Ultra (对标 GPT-4 Turbo) 外，其余星火大模型都提供大量的免费额度，其中星火 Lite 更是宣布永久免费，让开发者可以零成本部署轻量化的 AI 应用。

其中星火 Max 在中文领域的综合能力并不弱于 GPT-4，并且调整后的价格仅为 0.21 元/万 Tokens，仅需 2.1 元就可以生成与余华《活着》相当的文本内容。

另外，在 6 月 27 日发布的讯飞星火 V4.0，将会为开发者和用户带来更进一步的体验与更强大的 AI 大模型，而价格也依然会远低于 OpenAI。如果是有着复杂任务需求的朋友，可以关注一下发布会及后续的价格方案。

4、智谱清言：OpenAI 能给的我也给

虽然在应用领域的名气不如文心、通义等 AI 大模型，但是在多项 AI 大模型测试中，智谱清言也是排行榜上的常客，此次针对 OpenAI 的开发者推出了专属的搬家计划。

智谱清言为每个开发者都提供 1.5 亿 Token (5000 万 GLM-4 + 1 亿 GLM-4-Air) 的免费使用额度，并且提供对应的迁移培训，确保开发者可以轻松将应用迁移到 GLM 大模型平台。

对于需求较多的用户，智谱清言还提供了额外的优惠，如与 OpenAI 使用规模对等的 Token 赠送计划 (不设上限) 和与 OpenAI 对等的并发规模，此外还有专属的技术顾问及专家支持。

作为国内领先的 AI 大模型，GLM-4 大模型支持高达 128K 的上下文窗口长度，能够轻松处理长文本的提问及回答需求，而且支持图像、声音等多模态数据处理，使其作为通用大模型使用时的表现十分不错。

有兴趣的朋友可以前往智谱微信公众号查看相关信息，并添加客服进行咨询。

5、百川大模型：API 兼容 OpenAI

前段时间，百川大模型发布了新一代的 Baichuan 4，实测性能在国内名列前茅，丝毫不逊色于海外主流 AI 大模型，并且提供了多样化的 AI 模型选择，除 Baichuan 4 外还开放了 Baichuan3-Turbo、Baichuan3-Turbo-128k、Assistant API 等多个 API 接口。

为了方便 OpenAI 的开发者进行迁移，百川大模型除了提供 1 千万 Token 免费额度外，还提供 Assistant API 的免费使用权益，并且上线对应的技术专家答疑服务，最快 5 分钟即可完成 API 迁移。

从查询到的信息来看，Baichuan 4 对比 Baichuan 3 在通用能力上提升了 10%，代码和代码能力也分别有着 14% 和 9% 的提升，并且拥有出色的多模态处理能力，可以针对多用途的 AI 应用和开发项目提供更好的服务。

此外，Baichuan 4 还为开发者准备了更易用的开发平台，后续将推出零代码 Agent 创建平台产品，让开发者可以通过自然语言的方式创建 Agent，即使是不懂代码的用户也可以轻松使用，进一步降低 AI 领域的入门门槛。

值得注意的是，百川大模型的优惠活动针对所有开发者，不管你是 OpenAI 的迁移用户还是新开发者，都可以前往百川大模型的微信号了解详情。

6、天工大模型：中文能力比 GPT-4 更

强

针对迁移的 OpenAI 开发者，天工大模型也提供了一系列的优惠政策：免费的 1.5 亿 Tokens 和提供专属的迁移服务技术支持和备案培训及辅助，为开发者提供更加全面的服务。

天工 AI 大模型是由昆仑万维自研的超大规模语言模型，在相关测试中，天工 3.0 取得了优异的评分，在中文任务中的实际表现甚至超过了 GPT-4。而且，天工 3.0 还是目前全球模型参数最大、性能最强的开源 MoE 模型之一。

对于有复杂任务需求的开发者来说，天工 3.0 提供的多模态处理能力和长文本处理能力足以应付不同的场景需求。而对于使用场景比较轻量化的开发者而言，也可以选择性价比更高的天工 2.0 大模型，其实际表现并不逊色于 ChatGPT 3.5，同样拥有业内领先的 AI 性能。

有意向的朋友可以前往天工 AI 智能助手微信号进行咨询。

7、腾讯混元大模型：新用户送 1 亿 Tokens

作为 BAT 三巨头之一，腾讯显然也不会缺席。根据最新的消息，腾讯云针对 OpenAI 及新迁移的企业用户，赠送腾讯混元大模型的 1 亿 Tokens，并且提供免费的专属迁移工具和服务。

在国内外的多项测试中，腾讯的混元大模型都取得了出色的成绩，加上腾讯在 AI 领域的多年技术积累，使其可以为开发者提供多种规模、不同领域的专属 AI 大模型，进一步降低开发者的使用成本和开发难度。

此外，腾讯未来预计还将上线一系列的 AI 游戏开发工具。对于游戏开发者及相关从业者来说，腾讯的混元大模型在娱乐领域的表现更值得期待，有兴趣的朋友可以前往腾讯云微信公众号了解。

大模型大战全面进入下半场

谁能想到，OpenAI 的封杀令，竟然会一石激起千层浪，让国内的一众 AI 大模型纷纷进入抢客模式。

不过细想之下倒也不足为奇，早在 5 月份，国内的 AI 大模型就已经开始内卷，除了加快新版本的上线速度外，关键的调用成本也开始暴跌，部分国内领先的 AI 大模型调用成本甚至降低到原有的 10%。

可以说，国内的 AI 大模型正愁着如何招揽开发者的时候，OpenAI 就送上了一份“大礼”，确实是要说声谢谢。其实，对于多数开发者来说，他们的 AI 应用和项目只需要国产 AI 模型就可以满足，并不一定需要用到 ChatGPT，除了少数高要求、高复杂度的项目外，相信很多开发者都会逐渐将重心转移到国内的 AI 大模型平台上。

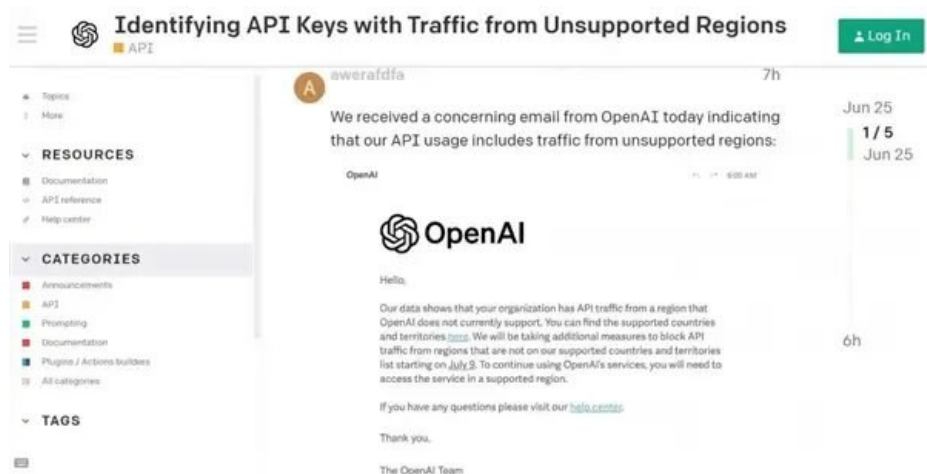
对于国内的 AI 生态而言，这无疑是一件好事，不过也会进一步加剧国内 AI 大模型的竞争，丰厚的迁移优惠固然是吸引开发者的“萝卜”，但最终能否让开发者留下来并长期使用，还是要看 AI 大模型的性能是否能够真的成为 ChatGPT 的平替。

2023 年，如火如荼的“百模大战”推动了国产大模型的技术进步。2024 年，“百模大战”走向结束，头部大模型初创六强浮出水面 (智谱 AI、MiniMax、月之暗面、百川智能、零一万物、面壁智能)，与百度文心、阿里通义、讯飞星火、字节豆包、腾讯混元、华为盘古、昆仑万维天工 AI 等巨头大模型构成第一阵营“十三太保”，基础大模型的竞争已告一段落，大模型进入“落地赛阶段”。

如今，不论是 AI 大厂还是初创公司的大模型都将重点放在了“落地”上。不同基础大模型、垂直大模型正在差异化发展，在各自擅长的领域形成特长。如百川智能强调语言认知能力，以及医疗等场景优势；kimi 长文本优势凸显；智谱 AI 除了技术强大外，对开发者门槛低强调普惠理念……

在这样的关键时刻，位于中间层的钉钉、字节扣子、阿里云等平台推出了“模型可选”模式，基于开放模式，让开发者可以选取最合适应用场景的大模型，并可无缝切换、快速迁移，类似于大模型的“分发入口”，这也意味着，基础大模型的竞争只会更加激烈。

在 OpenAI 拒绝中国开发者的“泼天富贵”面前，新一轮的厮杀已经开始，围绕技术、产品、服务、品牌、价格、BD 乃至生态的综合较量，只会更加激烈，更加精彩。



从 7 月 9 日起采取额外措施，阻止来自非支持国家和地区列表中的地区的 API 流量。若要继续使用 OpenAI 的服务，需要在支持的地区访问该服务。