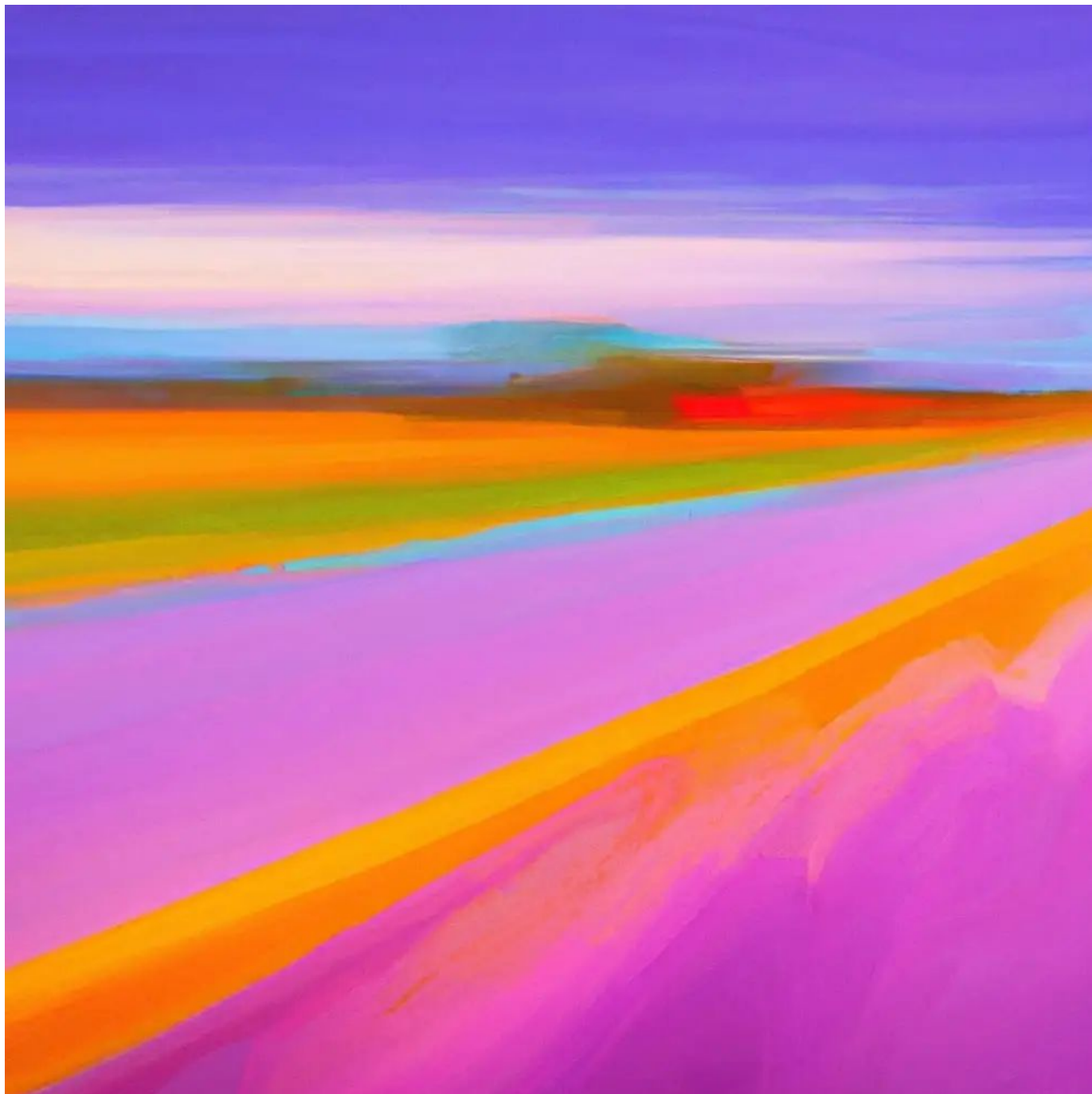


OpenAI 又发新博客了，有五大更新

原创 我是哥飞 哥飞 2024-01-26 09:18:56 广东

大家好，我是哥飞。



今天 OpenAI 又发新博客了，原文可以在下面两个链接查看。

原文：

<https://openai.com/blog/new-embedding-models-and-api-updates>

中文翻译：

<https://readweb.ai/zh/page/953128da1a6a598740eaf88a40ea65d0>

哥飞说下自己的理解和总结。

1、发布两款新的嵌入模型

`text-embedding-3-small` 体积更小、效率更高，价格是之前嵌入模型的五分之一。

`text-embedding-3-large` 性能更强，维度更高，高达3072维度，意味着更精准。

以上两个模型都支持传入 `dimensions` 参数来缩短向量长度。

如`text-embedding-3-large`可以缩短到256长度，都比之前嵌入模型1536长度更精准。

2、发布了新的 `GPT-3.5-Turbo-0125` 模型

价格大降，输入价格降低到之前的一半，输出价格降低到之前的四分之三。

同时性能更强了一些些，修复了一些bug。

3、发布了新的 `GPT-4-0125-preview` 模型

提升了代码生成任务的准确度和性能，修复了一些非英语的 `utf8 bug`。

4、发布了新的审核模型

新模型 `text-moderation-007` 号称迄今为止最强大的有害信息审核模型。

5、更新了密钥的权限控制精度

Usage

Cost Activity

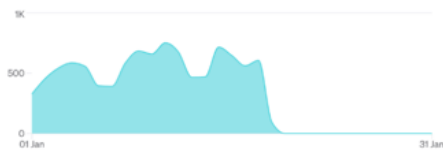
Models ▾

< January >

Export

GPT-3.5-turbo-0613

API requests 10,179



Tokens 3,756,913



Top Users

hello@openai.com

Top Organizations

MyOrg

Top API Keys

LaundryBuddy Key

Playground

Untracked

Start tracking new usage for untracked API keys [here](#)

GPT-4-0613

API requests 1,002

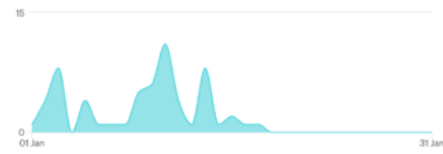


Tokens 588,460



GPT-4-1106-preview

API requests 60



Tokens 907,761



启用跟踪后，可以按照密钥维度查看个导出用量报告。

可以为某个密钥设置为只读模式，适用场景是内部跟踪用量数据。

可以限制某个密钥只能访问某些模型接口，更精准的控制密钥权限。

#ChatGPT