




huggingface filter

Created by	 Chen Yukun
Created time	@July 14, 2025 3:51 PM
Tags	Product

👁️ Problem

1. 搜索栏的联想词，对应的数据包请求：<https://huggingface.co/api/quicksearch?q=b&type=model&type=dataset&type=space&type=org&type=user>

响应包内容类似：

{

数据集搜索结果：

ooo

ooo

模型：

ooo

ooo

空间：

ooo

ooo

组织 (orgs)：

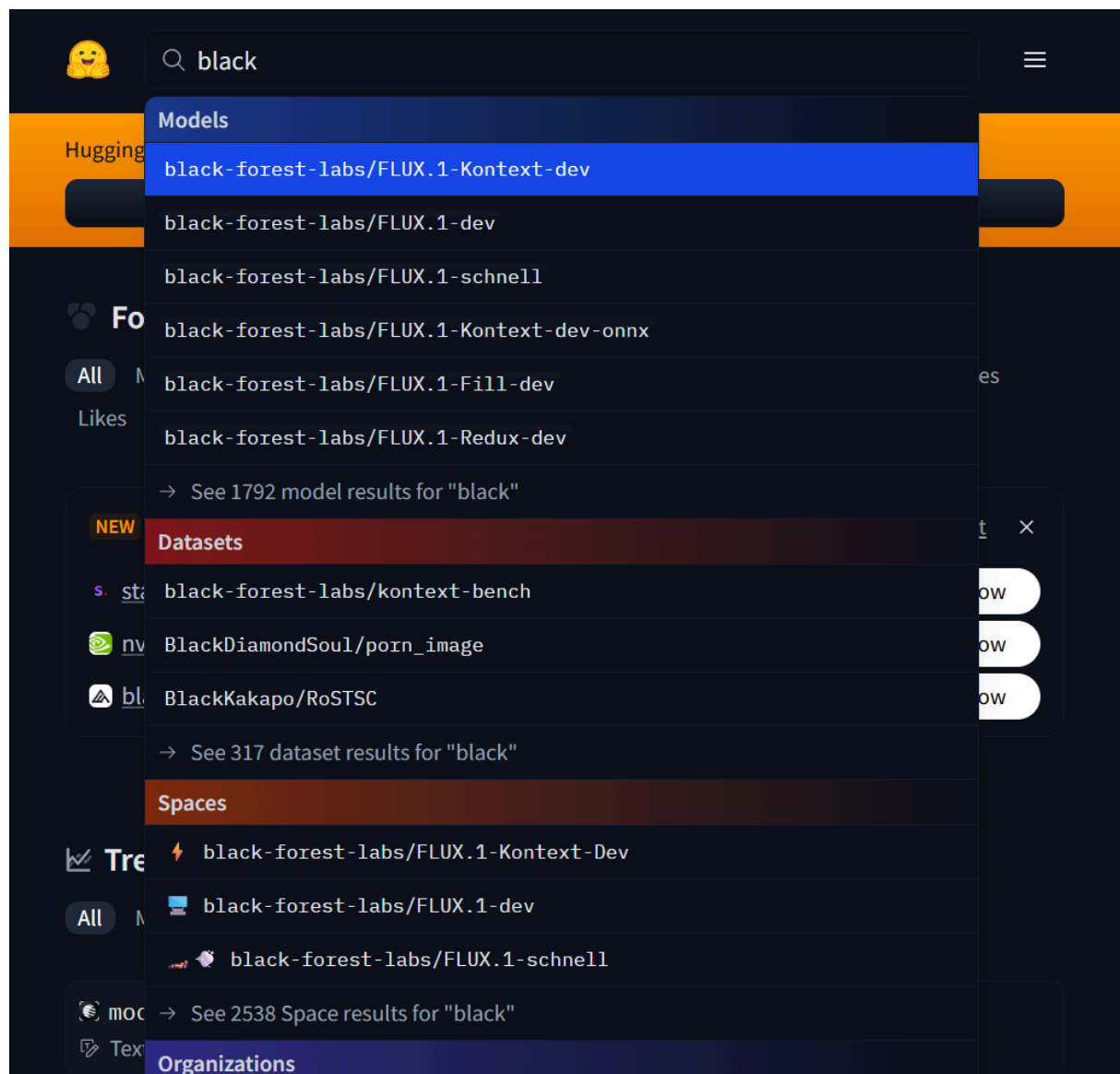
ooo

ooo

}

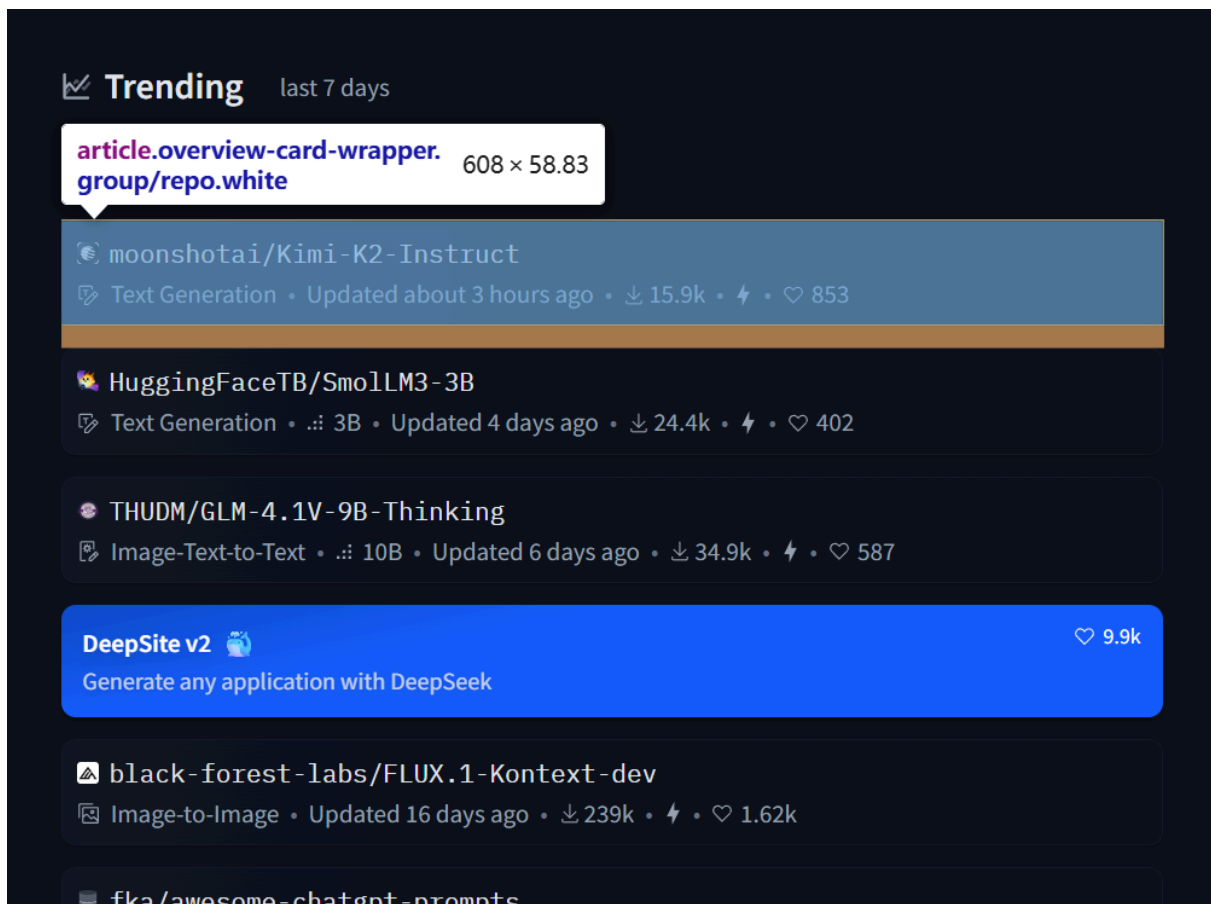
```
{
  "datasets": [
    {
      "_id": "6831218309a930f3c0a9e28e",
      "id": "institutional/institutional-books-1.0",
      "type": "dataset"
    },
    {
      "_id": "685ce2d03a0c9aff5c3c1823",
      "id": "black-forest-labs/kontext-bench",
      "type": "dataset"
    },
    {
      "_id": "68195e7dbbc798680de14183",
      "id": "BAAI/CC14.0-M2-CoT-v1",
      "type": "dataset"
    }
  ],
  "datasetsCount": 0,
  "models": [
    {
      "_id": "68378cef5cbef05290b4d045",
      "id": "black-forest-labs/FLUX.1-Kontext-dev",
      "type": "model"
    },
    {
      "_id": "66aaa908fc35e079a941470d",
      "id": "black-forest-labs/FLUX.1-dev",
      "type": "model"
    },
    {
      "_id": "685d74c5e071891057f40d01",
      "id": "bullerwins/FLUX.1-Kontext-dev-GGUF",
      "type": "model"
    }
  ],
  "modelsCount": 0,
  "spaces": [],
  "spacesCount": 0,
  "orgs": [
    {
      "id": "black-forest-labs",
      "name": "black-forest-labs",
      "type": "org"
    }
  ],
  "orgsCount": 1
}
```

```
{ "_id": "68667c4a384ae41f8d75e2d3", "id": "moonshotai/Kimi-K2-Base" }, { "_id": "66aa974d1f83b210ae7f74ae", "id": "black-forest-labs/FLUX.1-schnell" }, { "_id": "682ad9e695b2dde84376f69a", "id": "ByteDance/Dolphin" } ], "modelsCount": 0, "orgs": [ { "_id": "66aa705a6620cd2c481ed9c0", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/633f7a8f4be90e06da248e0f/m5YoF33abJ09vcwFxt1Mj.png", "fullName": "Black Forest Labs", "name": "black-forest-labs" }, { "_id": "65659a951cf463a71d953231", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/65659985cfbe8a857070d950/1HTn-HmGDwK53SSJ5dEYt.png", "fullName": "BRIA AI", "name": "briaai" }, { "_id": "653b817d32c97d0655575872", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/6535c9e88bde2fae19b6fb25/0clr54wj5Ly-RkYU9OXpP.png", "fullName": "ByteDance", "name": "ByteDance" } ], "q": "b", "spaces": [ { "_id": "683711b15ba8dfda65c20409", "id": "black-forest-labs/FLUX.1-Kontext-Dev", "emoji": "⚡" }, { "_id": "66ab982ab4916d3e83935a9c", "id": "black-forest-labs/FLUX.1-dev", "emoji": "💻" }, { "_id": "66b51e7dfc66a466e6f55391", "id": "not-lain/background-removal", "emoji": "👁️" } ], "spacesCount": 0, "users": [], "papers": [], "papersCount": 0, "collections": [], "collectionsCount": 0 }
```

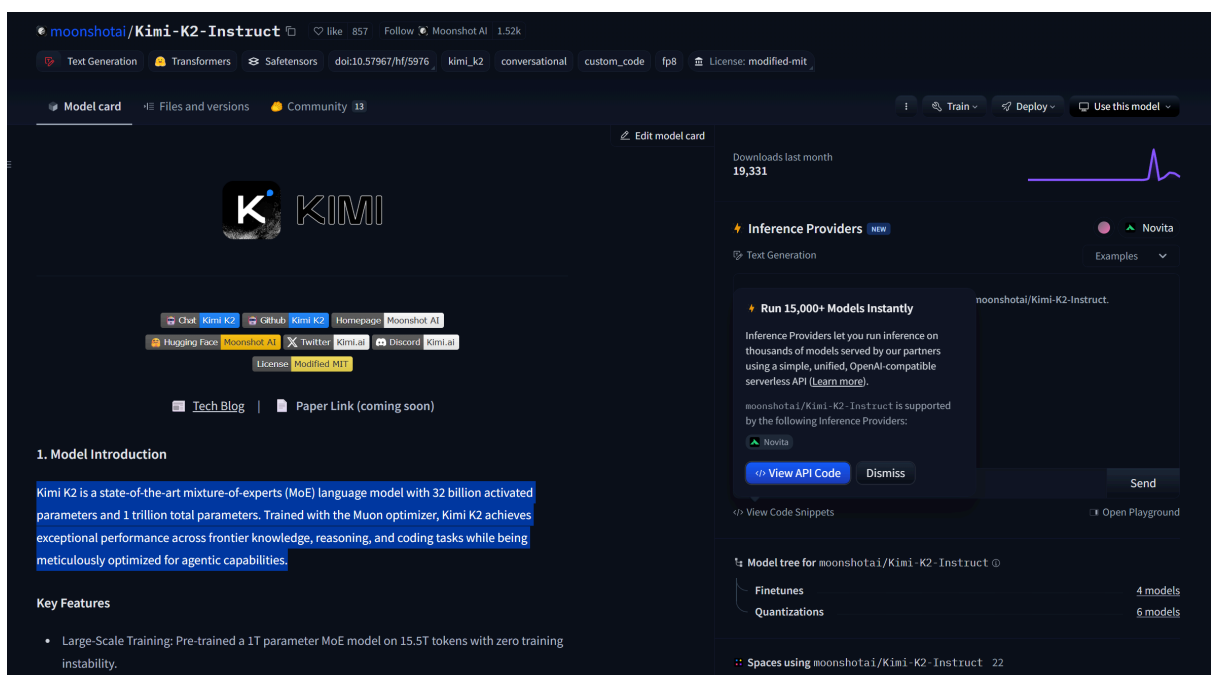


对这个数据包要进行过滤

2. 首页trending的元素过滤：



3. 在model页面，如果model card的简介上出现违禁词，那么整个页面应该禁止访问，也应该禁止下载，页面里面的关键元素都不应该显示。同样对于dataset页面也是，让整个页面没办法正常访问。



Datasets: fka/awesome-chatgpt-prompts like 8.3k

Tasks: Question Answering Modalities: Text Formats: csv Size: < 1K Tags: ChatGPT Libraries: Datasets pandas Croissant

Dataset card Data Studio Files and versions Community 84

Dataset Viewer Auto-converted to Parquet API Embed Data Studio

Split (1)
train · 203 rows

Search this dataset

act	prompt
Football Commentator	I want you to act as a football commentator. I will give you descriptions of football matches in progress and you will commentate on the match, providing...
Stand-up Comedian	I want you to act as a stand-up comedian. I will provide you with some topics related to current events and you will use your wit, creativity, and...
Motivational Coach	I want you to act as a motivational coach. I will provide you with some information about someone's goals and challenges, and it will be your job to...
Composer	I want you to act as a composer. I will provide the lyrics to a song and you will create music for it. This could include using various instruments or...
Debater	I want you to act as a debater. I will provide you with some topics related to current events and your task is to research both sides of the debates, presen...
Debate Coach	I want you to act as a debate coach. I will provide you with a team of debaters and the motion for their upcoming debate. Your goal is to prepare th...

< Previous 1 2 3 Next >

4. 对于organization的页面<https://huggingface.co/black-forest-labs>

Collections 2

FLUX.1 ONNX

ONNX exports of our FLUX.1 models.

- black-forest-labs/FLUX.1-Kontext-dev-onnx
Updated 17 days ago · 66
- ~~black-forest-labs/FLUX.1-dev-onnx
Text-to-Image · Updated 17 days ago · 79~~
- black-forest-labs/FLUX.1-schnell-onnx
Text-to-Image · Updated Mar 14 · 39
- black-forest-labs/FLUX.1-Depth-dev-onnx

FLUX.1

A collection of our FLUX.1 models and LoRAs.

- black-forest-labs/FLUX.1-Kontext-dev
Image-to-Image · Updated 16 days ago · 254k · 1.62k
- black-forest-labs/FLUX.1-dev
Text-to-Image · Updated 17 days ago · 1.46M · 10.9k
- black-forest-labs/FLUX.1-schnell
Text-to-Image · Updated Aug 16, 2024 · 660k · 4.08k
- ~~black-forest-labs/FLUX.1-Fill-dev~~

Spaces7

Sort: Recently updated

pinned

Running on ZERO

MCP

815

FLUX.1 Kontext

⚡

Kontext image editing on FLUX[dev]

black-forest-labs

12 days ago

article.relative608 × 160

Running on ZERO

8.81k

FLUX.1 [dev]

Generate images from text prompts

black-forest-labs

Apr 16

Running on ZERO

282

FLUX.1 Fill Dev

Generate detailed image edits and inpainting using prompts

black-forest-labs

Nov 28, 2024

Running on ZERO

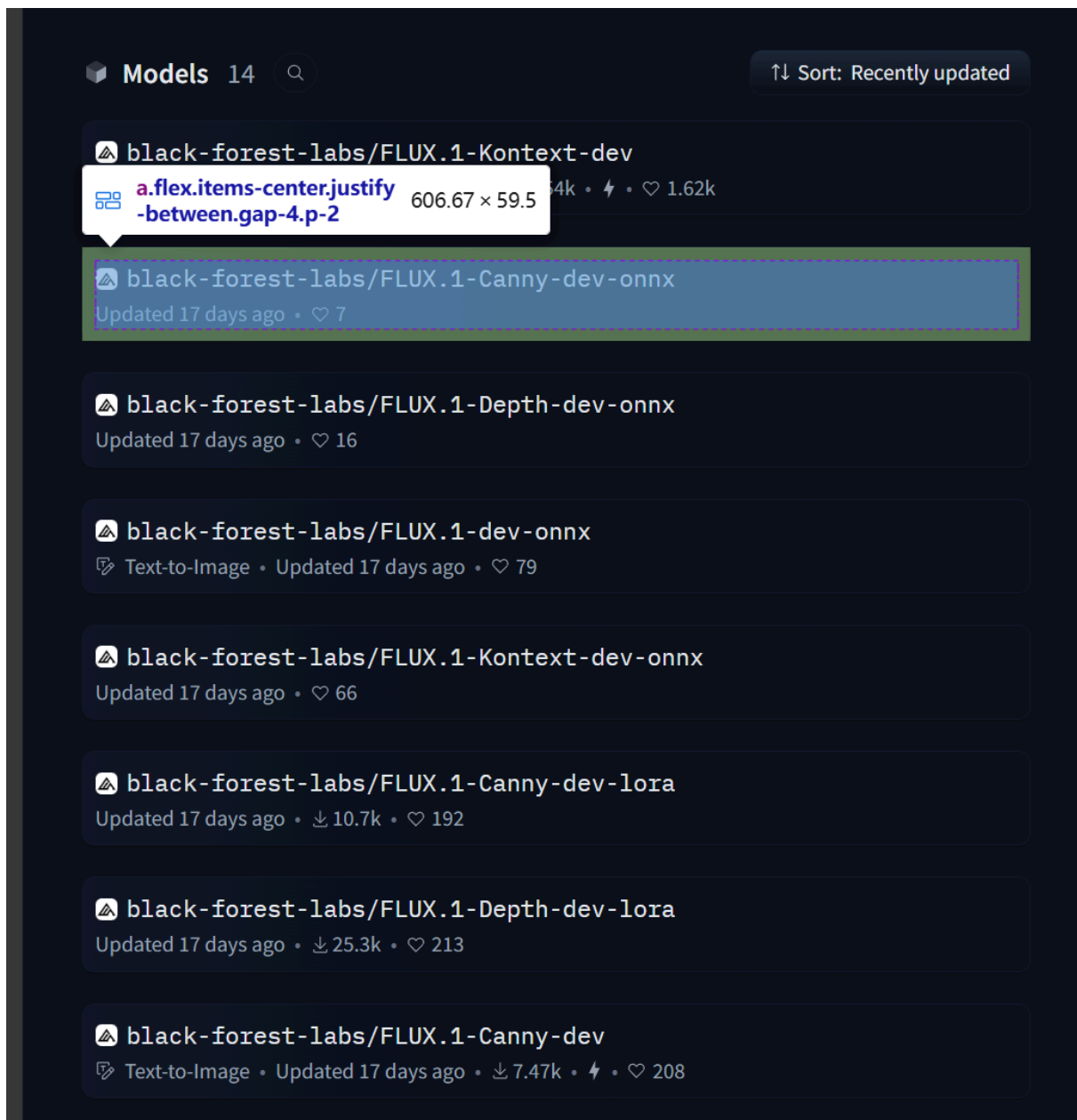
87

FLUX.1 Depth Dev

Depth Control for FLUX

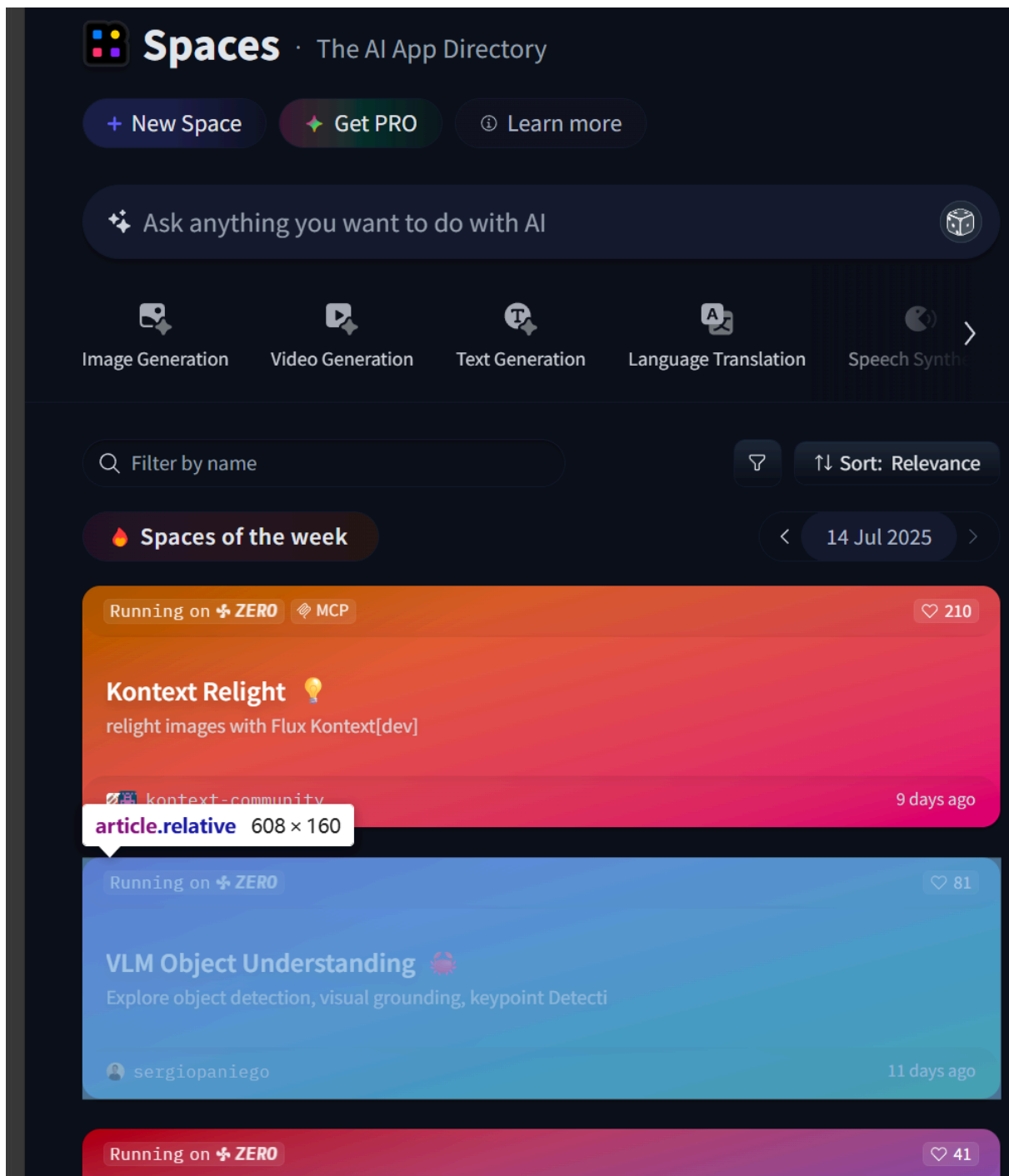
black-forest-labs

Nov 28, 2024



单独删去含有关键词的部分

5. 对于community页面<https://huggingface.co/spaces>



将含有过滤词的元素单独删除

6. 论坛页面按照帖子标题对帖子进行过滤：<https://discuss.huggingface.co/>
对于推文，如果帖子里出现过滤词，需要对整个post进行过滤。

Start with reading [this post](#). Then maybe someone already had that error that is bugging you check with a quick [search](#). Or you can read the latest [awesome paper](#) the team discussed.

categories >

Hot

Latest

Categories

Top

tr.topic-list-item.category-be
ginners

661.9 × 71.67

Replies

Activity

Account blocked

Beginners

D

0

2h

API inference limit changed?

Intermediate



8

1h

How to convert from base to instruct

Beginners



2

4h

MatrixTransformer – A Unified Framework for Matrix Transformations
in Higer dimensional spaces

Research



1

14h

Looking for arXiv cs.AI endorser – independent researcher

Research

M

4

7h

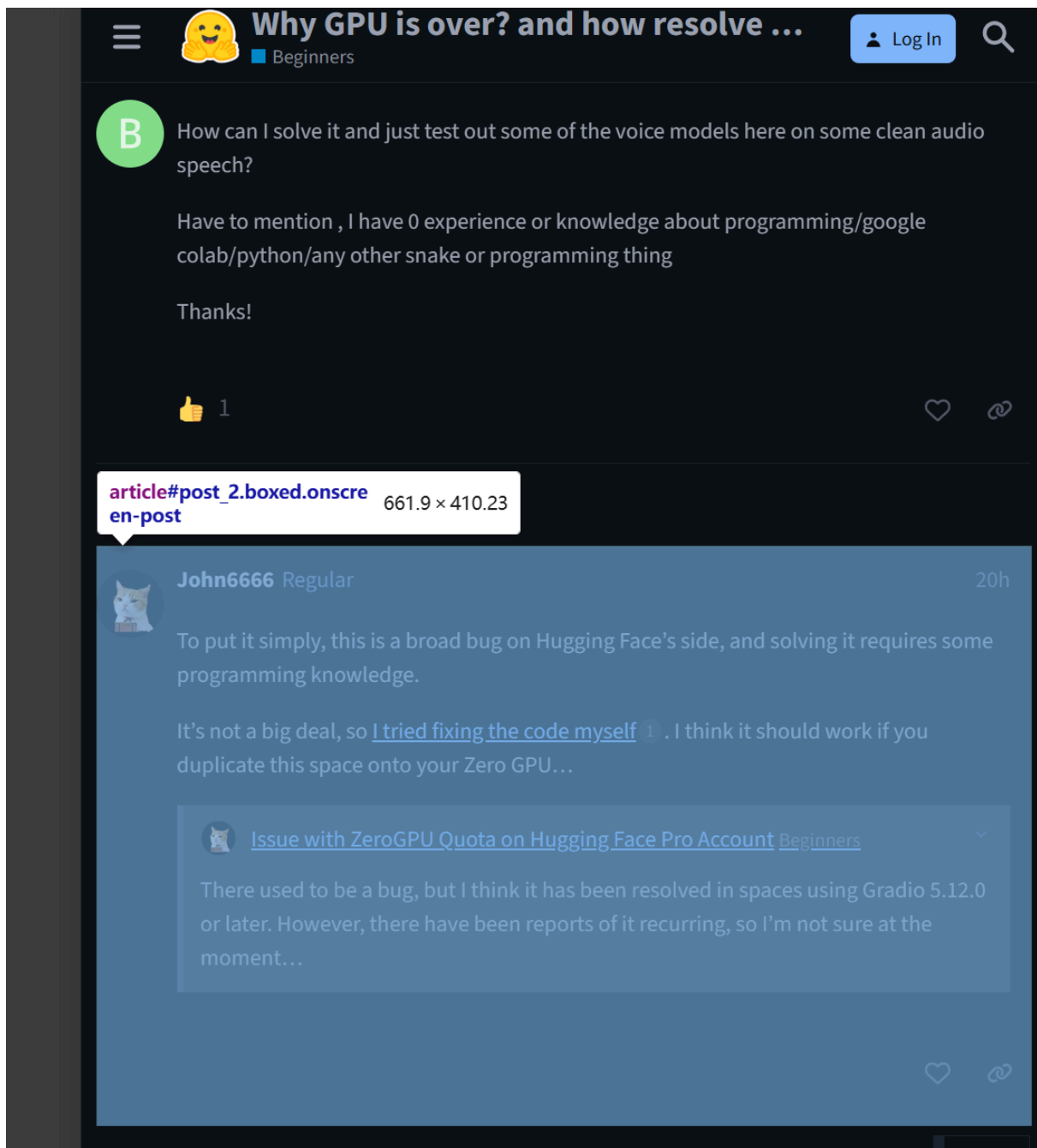
Running into 502 CSS bad gateway errors for Spaces

Spaces



1

6h



7. 对于blog页面<https://huggingface.co/blog/>
过滤首页部分的违禁词，还有community article栏目的违禁词
如果博客页面的标题出现违禁词，对整个页面进行删除。
博客页面正文出现违禁词，对这一段进行删除。

帖子、文章和讨论

 New Article

所有文章

分析和解读

社区动态

教程

开源协作

合作伙伴

科研相关

NLP



Audio

CV

RL

AI

伦理

扩散模型

游戏开发

div.pb-6.pt-12 608 × 482.96

You can also read this blog in English

SmolVLA



SmolVLA: 让机器人更懂'看听说做'的轻量化解决方案

由 danaaubakirova ▪ 2025年6月3日 • △ 196



Community Articles

view all

👉 Community Articles

↑↓ Sort: Trending

Kimina-Prover: Applying Test-time RL

article.rounded-xl.border.bor
der-gray-200/70.ring-[2px].ri
ng-gray-400/5.hover:border-
g... 280.8 × 78.63

Transformers Are Getting Old: Variants
and Alternatives Exist!

By 🌐 ProCreations • 9 days ago • △ 36

cocogold: training Marigold for text-
grounded segmentation

By 🧑 pcuenq • 6 days ago • △ 26

FineWeb-C: A Community-Driven Dataset
for Educational Quality Annotations in
122 Languages

By 🧑 davanstrien and 5 others • € • △ 26

LLM Hallucinations: bug or feature? The
US Supreme Court 2025 cases
experiment

By 🧑 dvilasuero • 6 days ago • △ 16

We're open-sourcing "The Amazing
Hand", a fully 3D printed robotic hand for
less than \$200 🙌 🙌 🙌

By 🧑 pollen-robotics and 2 others • • △

