stable-diffusion-webui-api

Overview

In this notebook, we will explore how to build generative fill application and host Stable Diffusion/ ControlNet / segment anything models on SageMaker asynchronous endpoint using DLC container.

You will find 2 Jupyter Notebooks: 1 for running with Amazon SageMaker Studio and 1 for running with Amazon SageMaker Notebook.

IAM role recommendations

1. Running with Amazon SageMaker Studio

- Permissions policies

```
AmazonSageMakerFullAccess
AmazonEC2ContainerRegistryFullAccess
AWSCodeBuildAdminAccess
IAMFullAccess
```

- Trusted entities

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": [
                    "sagemaker.amazonaws.com",
                    "codebuild.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

- Tested image, kernel, and instance:

```
image: Pytorch 2.0.1 Python 3.10 CPU Optimized
kernel: Python 3
instance: ml.m5.4xlarge
```

2. Running with Amazon SageMaker Notebook

- Permission Policies

```
AmazonSageMakerFullAccess
AmazonEC2ContainerRegistryFullAccess
```

- Trusted entities

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": [
                    "ecs.amazonaws.com",
                    "sagemaker.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

- Tested kernel:

```
kernel: conda_pytorch_p310
```

Note

1. You may need to adjust IAM roles definition to achieve fine grained access control.

2. Amazon Web Services has no control or authority over the third-party generative AI service referenced in this Workshop, and does not make any representations or warranties that the third-party generative AI service is secure, virus-free, operational, or compatible with your production environment and standards. You are responsible for making your own independent assessment of the content provided in this Workshop, and take measures to ensure that you comply with your own specific quality control practices and standards, and the local rules, laws, regulations, licenses and terms of use that apply to you, your content, and the third-party generative AI service referenced in this Workshop. The content of this Workshop: (a) is for informational purposes only, (b) represents current Amazon Web Services product offerings and practices, which are subject to change without notice, and (c) does not create any commitments or assurances from Beijing Sinnet Technology Co., Ltd. (“Sinnet”), Ningxia Western Cloud Data Technology Co., Ltd. (“NWCD”), Amazon Connect

Technology Services (Beijing) Co., Ltd. (“Amazon”), or their respective affiliates, suppliers or licensors. Amazon Web Services’ content, products or services are provided “as is” without warranties, representations, or conditions of any kind, whether express or implied. The responsibilities and liabilities of Sinnet, NWCD or Amazon to their respective customers are controlled by the applicable customer agreements.