



Stable Diffusion  
welcome to AI creative world

Changes in the world due to AI

## Stable Diffusion

StableDiffusionを学ぶことで、自由自在に画像を生成し、自分だけの作品を作れるようになります。  
本講義では、StableDiffusionの基礎から実践的な画像と動画生成のスキルを身につけていきましょう。

Lesson①

**Stable Diffusionの基本と仕組み**

- ① 他の画像生成AIの違い
- ② StableDiffusionの基礎

Lesson②

**プロンプト設計のコツ**

- ① プロンプトの基本
- ② プロンプトの種類

Lesson③

**ControlNet**

- ① StableDiffusionの機能理解
- ② ControlNet

■ 最終ゴール「思い通りのAIクリエイティブを作れるようになる」

### まずStableDiffusionとは？

Stable Diffusionは、テキストプロンプトから画像を生成するAI技術であり、以下の特徴を持ちます。

高精度な画像生成

小さなディテールまで表現可能。



多様なスタイルに対応

リアルな写真から抽象的なアートまで生成可能。



オープンソース

開発者が自由にモデルをカスタマイズ可能。



### 他の画像生成AIとの違い

Stable Diffusionや画像生成AIは、AIがノイズから画像を生成する技術です。

(1)StableDiffusion

テキストプロンプト + "自作または配布"の学習モデル > 生成画像

(2)その他生成AI

テキストプロンプト + 固定またはテンプレートの学習モデル > 生成画像

つまり StableDiffusionはオープンソース だからこそ

ユーザー間での研究・研鑽が活発に行われていて、  
"表現の幅や再現性"という意味で強みのある画像生成AIである。

## Stable Diffusionの基本と仕組み

このセッションで学ぶ内容は「Stable Diffusionの基礎」



## プロンプト設計の基本

StableDiffusionにはプロンプトで生成を制御するアプローチが2つあります。

生成する要素  
Positive Prompt

生成しない要素  
Negative Prompt

この2つを組み合わせることで画像を生成していきます。  
早速、サンプルを見て理解を深めましょう。



使用モデル: NTR MIX | illustrious-XL | Noob-XL

### プロンプト例

#### Positive Prompt 生成する要素

1boy black\_pants blonde\_hair blue\_jacket,  
closed\_mouth denim\_jacket grey\_shirt jacket long\_hair  
male focus pants ponytail. profile sanpaku shirt sitting solo,  
masterpiece, best quality, absurdres, sketch, flat colors

#### Negative prompt 生成しない要素

chibi, loli, futanari, lowres, (worst quality, bad quality), bad anatomy,  
jpeg artifacts, signature, watermark, simple background, 3d



使用モデル: NTR MIX | illustrious-XL | Noob-XL

### プロンプト例

point① 英語で記述: AIモデルは英語をベースに学習されています。

point② 単語の羅列または文章形式

単語例: sunset, ocean, realistic

文章例: A sunset over the ocean in a realistic style

順番が重要: 前の単語ほど画像生成に強く影響します。

※これはネガティブプロンプトにも該当する。

※使用するモデルSDXLやSD1.5などによっても優位性が変わる。

## プロンプトの種類

品質に関するプロンプト	・ masterpiece 傑作 ・ best quality 最高品質
画質に関するプロンプト	・ high resolution 高解像度 ・ 4K 4K解像度
キャラクターに関するプロンプト	・ l boy/girl ・ foodie ・ jeans ・ sneaker
背景に関するプロンプト	・ garden 庭園 ・ simplebackground シンプルな背景
基本的にはMidjourneyと同じ	・ masterpiece 傑作 ・ best quality 最高品質

## よくあるプロンプトで聞く単語ばかりですがStableDiffusion特有の構文がある

### — 強調構文 —

(prompt)で要素を強調。  
[prompt]で要素を減少。

### — 強調構文(係数) —

prompt: 1.1~2で要素を強調。  
prompt: 1.1は強調。  
prompt: 0.9は減少。

### Point!

括弧の数を増やすことで効果を乗算することができます。

(prompt) は1.1倍

((prompt)) は1.21倍

((((prompt)))) は1.33倍

(((((prompt)))))) は1.46倍

複数の[]を使用した場合の効果は次のとおりです。

[キーワード] は0.9倍

[[キーワード]] は0.81倍

[[[キーワード]]] は0.73倍

[[[[キーワード]]]] は0.66倍

そして以下のような構文を用いることで  
2つのキーワードを混ぜ合わせることができます。

[prompt1:prompt2:係数]

この構文内の係数はキーワード1がキーワード2に切り替わるタイミングを制御するもので、0から1の間の数値になります。

例えば、次のようにシルバーの髪とピンクの髪をブレンドする場合を考えてみましょう。

[Silver hair:Pink hair:0.7]

[[Silver hair:1.3):(Pink hair:1.3):0.3]



[[Silver hair:1.3):(Pink hair:1.3):0.4]



[[Silver hair:1.3):(Pink hair:1.3):0.5]



[[Silver hair:1.3):(Pink hair:1.3):0.6]



[[Silver hair:1.3):(Pink hair:1.3):0.7]



もう意味わからなくなってってるかもですけど、触りながら覚えましょう!!。

## 『Break』でグループを区切る

Stable Diffusionのプロンプトは、【 75トークンで1グループ 】となります。

グループ内に序列があり、【 1番目のトークンが一番強く、75番目のトークンが一番弱い指示 】になります。

また1グループ目75番目のトークンより、【 2グループ目1番目のトークン(全体76番目のトークン) 】の方が強くなります。



## BREAKE構文 例



## Break構文を使わなかった場合

## 修正前

l girl,standing pose,cute,cute eyes,detailed hair, anime,highlight hair,front sun light,beautiful hair,illustration,face up, smile,white teeth, school uniform, blue ribbon, gentle smiling expression of a woman,black hair, Blue eyes, acing front,

青い瞳 (blue eyes) と指示しましたが、片方の目が若干紫になっています。Break構文を使って、紫の瞳を青くしてみましょう。

## BREAKE構文 例



## Break構文を使った場合

## 修正後

l girl,standing pose,cute,cute eyes,detailed hair, anime,highlight hair,front sun light,beautiful hair,illustration,face up, smile,white teeth, school uniform, blue ribbon, gentle smiling expression of a woman,black hair, **break**, Blue eyes, acing front,

【 blue eyes 】の前にBreak構文を入れた結果、きちんと青い瞳になりました。生成した画像を修正したいなら、修正した箇所に関するpromptの前にBreak構文を入れてみましょう。

## 【 SteableDiffuisonの機能理解 】

## StableDiffusionのUIの種類

## AUTOMATIC1111 WebUI ★☆☆

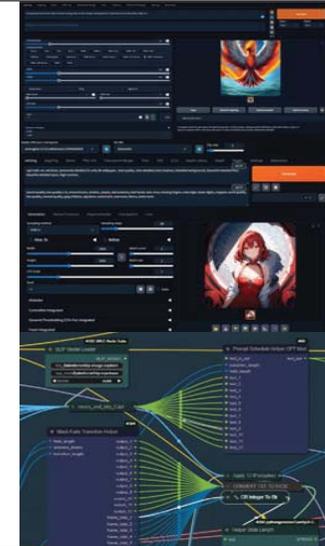
最もポピュラーなWebUIであり、多くのカスタム機能や拡張が利用可能です。特に初心者にとっては使いやすく、コミュニティも活発でサポートが充実している。

## WebUI Forge ★☆☆

パフォーマンスの最適化、メモリ処理に優れたWebUIで、生成速度が早く、大量の生成や連続処理に強いのが特徴。

## ComfyUI ★★★

ビジュアルプログラミングに近い操作性を持つWebUIで、画像生成の各ステップを細かく制御できる。ノードベースで、複雑なプロンプトや生成プロセスを視覚的に管理できる。



難易度

### 基本生成における必須要素

- ①Sampling Steps...ノイズを除去する回数(レイヤーを重ねるイメージ)
- ②Sampling method...ノイズ除去する方法(ゲームで言うサーバーみたいなイメージ)
- ③Batch count...現在のプロンプトを使って生成する画像の枚数を指定
- ④Batch Size...Batch countで指定した画像生成を繰り返す回数を指定
- ⑤Seed...画像生成した際にランダムに割り振られる値(住所)みたいなもの
- ⑥CFG Scale...プロンプトの反映強度(ベースは7)

### アップスケール(生成した画像の高画質化)

HiresFix...画像生成時に最後の処理で解像度を数倍に  
Extras...画像生成後にアップスケールモデルを使い解像度を上げる

### 追加Extensionによるアップスケール

TiledDiffusion...画像生成時に最終出力した画像を分割して各部分にプロンプトを乗算させてよりディティールの細かい画像へ。  
MultiDiffusion...img2imgを用いて生成した画像をより精細にアップスケールできる拡張機能。最大で8Kの高解像度化が可能。

### システム上設定必須の要素

- ①Checkpoint...SDのモデルの事(SD1.5/SDXL/Pony/Flux/etc..)
- ②VAE/Text Encoder...画像生成の最終段階で画質や色のバランスを調整する補助的なモデル。
- ③clipskip...Stable Diffusionが画像生成時に使用する「テキストエンコーダーの層数」を調整する設定。

- 層数をスキップ → よりアニメ的・シンプルなスタイルになる。
- 層数をスキップしない → リアル画像向けの複雑なスタイルになる。
- 設定値の例
- 1(スキップなし):  
— CLIPのすべての層を使用する。
- リアルな画像生成や細部にこだわる場合に有効。
- 2:1層スキップする(より抽象的なテキスト理解を適用)。
- アニメやキャラクター生成に適している。

もうワケ分からん!!とりあえず生成してみましょう!!  
今回、皆さんに使っていただくのは【 Mimic PC 】です。



クラウド上で簡単に  
StabilityAIのアプリケーションを  
利用できるプラットフォーム。  
  
一緒に立ち上げまで行いましょう。

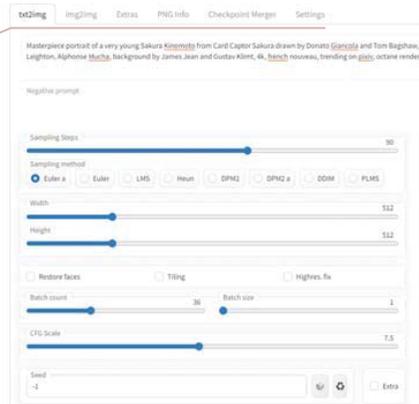
## 【 AUTOMATIC1111 WebUIの基本操作 】

Step1: ホーム画面の確認  
txt2img、img2img、Settings等が並ぶ。

Step2: txt2imgの操作手順  
promptから画像を生成する。

Step3: img2imgの操作手順  
参考画像を元に画像を出力する。

Step4: 設定の確認と変更  
SettingタブやVAEの切り替えを行う。



## 【 CivitAIの活用 】

### CivitAIの活用

AIモデルやプロンプトを共有・ダウンロードできるオンラインプラットフォームです。

URL: <https://civitai.com>

モデルのダウンロード

プロンプトの参照

参考生成結果の検索

### 検索方法

- ①サイトにアクセスし、検索バーに「モデル名」や「キーワード」を入力。
- ②フィルター機能を使ってモデルのカテゴリ(アニメ、リアル系など)や評価で絞り込む。
- ③各モデルのページでサンプル画像や推奨プロンプトを確認。

### ダウンロード手順

- ①欲しいモデルを選択し、対応するCheckpointファイル(.ckptまたは.safetensors形式)をダウンロード。
- ②AUTOMATIC1111 WebUIの「models」フォルダに配置。

### 参考プロンプトの活用方法

- ①モデルページの推奨プロンプトをコピーし、txt2imgで試して調整。
- ②良い生成結果を得るためにNegative Promptも確認。

### おすすめモデル例

- ①AnythingV4: アニメ・キャラクターデザインに最適。
- ②DreamShaper: 高品質なリアル画像生成向け。
- ③Counterfeit-V3: アートスタイルの幅広い表現が可能。

### プロンプト実践演習 プロンプトを試してみよう!

題1

簡単な風景画像生成

題2

キャラクター生成

【 良いプロンプトの例 】  
明確で具体的な描写を含む(例: vivid colors, cinematic lighting, detailed textures)

【 悪いプロンプトの例 】  
曖昧で不明瞭な記述(例: beautiful, good quality)

【 ControlNet 】

### ControlNetの基本概念

Stable Diffusionの拡張機能で、ポーズや構図を細かく指定できる。  
プロンプトで指示しにくい要素(ポーズ、線画など)の制御が可能。



## 【 ControlNetの使い方 】

## ■ 導入手順

- ①AUTOMATIC1111 WebUIのExtensionsタブでsd-webui-controlnetを検索し、インストール。
- ②WebUIを再起動し、ControlNetが有効化されていることを確認。

## ■ 主なモデルの用途

- OpenPose: 人物のポーズを指定。  
 Canny: 線画や輪郭を抽出し、画像生成に反映。  
 Scribble: 手描きのラフスケッチを基に画像を生成。  
 Inpaint: 部分的な画像修正や補完を実現。

## ■ 応用例

- ・写真からイラスト風画像への変換。
- ・キャラクターデザインでのポーズ指定。
- ・下書きからの清書生成。

## 【 ControlNetの使い方 】

## ■ ControlNetを有効化する手順

- ①AUTOMATIC1111 WebUIのtxt2imgタブまたはimg2imgタブを開く。
- ②左下にあるControlNetセクションを見つけ、「Enable」にチェックを入れる。

## ■ プリプロセッサとモデルの設定

- ①Canny: 線画や輪郭を基に画像生成。
- ②OpenPose: 棒人間のポーズを基に生成。
- ③Scribble: 手描きのラフスケッチを元に画像生成。
- ④IPAdapter: 画像の情報をプロンプトにし画像生成。

※プリプロセッサと対応するモデル

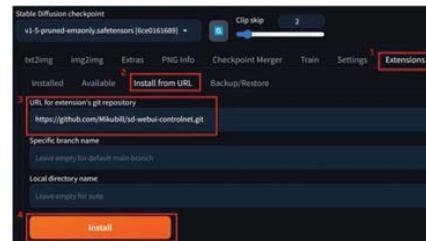
(例: Cannyモデルを選択する際はcontrol\_v1lp\_sd15\_cannyを指定)。

## 【 ControlNetの使い方 】

## ■ 操作手順の具体例

- ①入力画像をドラッグ&ドロップでアップロード。
- ②プリプロセッサ(例: Canny)を選択。
- ③モデル(例: control\_v1lp\_sd15\_canny)を選ぶ。
- ④パラメータ(例: Control Weightを0.8-1.0)を調整。
- ⑤「Generate」をクリックして画像を生成。

## 【 ①具体例 】



## 【 ①ラフスケッチを変換 】



## 【 ②写真のスタイル変換 】



## ■ 応用例

- ①ラフスケッチをリアルな画像に変換。
- ②写真を基にしたスタイル変換。
- ③OpenPoseを使用したキャラクターポーズ生成。

## 【 ControlNetの使い方 】

## ■ 演習課題

- ①OpenPoseを使用してポーズを指定し、画像を生成。
- ②Cannyを使って輪郭を抽出し、スタイル変換を実施。
- ③Scribbleを利用してラフスケッチをリアルなイラストに変換。

## ■ 手順の詳細

- ①AUTOMATIC1111 WebUIのControlNetセクションを開く。
- ②各プリプロセッサ(例: OpenPose)を選択し、対応するモデルを指定。
- ③サンプル画像をアップロードし、適切なパラメータを設定(例: Control Weightを0.8-1.0)。
- ④「Generate」をクリックして結果を確認。



Changes in the world due to AI

## AnimateDiffとSVD(StableVideoDiffusion)

Stable Diffusionは基本的には画像生成のためのツールですが、外部ツールや動画モデルを活用することで動画生成を行うことができます。

■ 最終ゴール「StableDiffusion基軸の動画生成ができるようになる」

- ① ComfyUIを使用したStable Video Diffusion (SVD)の基礎操作を理解。
- ② AnimateDiffでキャラクターアニメーションを制作。
- ③ 動画生成ワークフローを体験し、応用のヒントを得る。

■ SDを使った動画生成手法は大きく2つ。

- ① Diffusionモデルを時間軸に適用したStable Video Diffusion (SVD)。
- ② フレーム間の一貫性を強化するAnimateDiff。

### AnimateDiff

2Dアニメのような短い動きを作るのに最適!



フレーム間の動きを補完してアニメーションを生成(バラバラ漫画のイメージ)

### SVD

リアルな動画や映像作品向け!



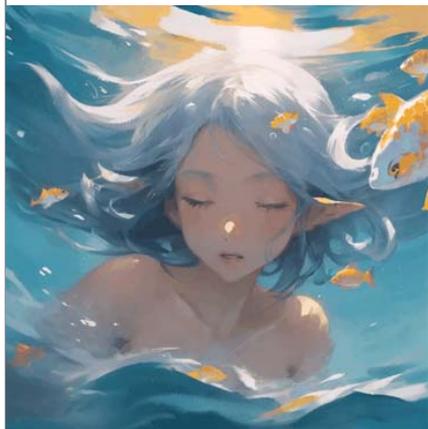
画像生成モデルを時間軸に拡張し、連続した動画を生成

## Stable Video Diffusion (SVD) の特徴



- ① 高解像度の動画生成  
フレーム間の連続性を保持しながら、テキストプロンプトから動画を生成可能。
- ② 時間的整合性  
動画の各フレームにおいて、一貫したスタイルと動きを持つ。

<https://youtu.be/No2RQ1dJsCA>



### AnimateDiffの特徴

- ①時間的整合性のあるフレーム生成
  - ・各フレーム間で一貫したスタイルや動きを保持。
  - ・動画の各フレームが独立していないため、連続的で自然な動きを実現。

- ②高度なプロンプト対応
  - ・プロンプトを基にしたカスタマイズが可能。

例: 「a cat jumping over a log, cinematic lighting」と入力すれば、ジャンプの動きが含まれる動画を生成。

<https://youtube.com/shorts/QATskqG69nY?feature=share>



### AnimateDiffの特徴

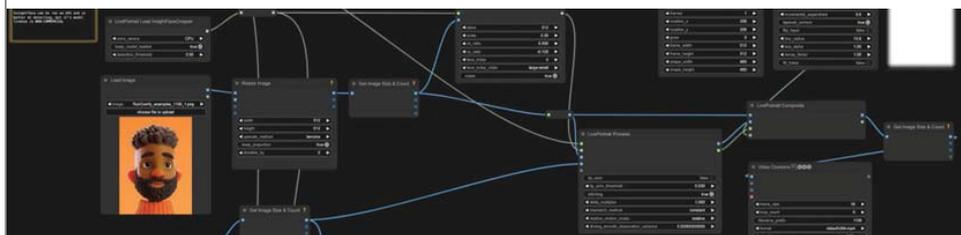
- ④フレームごとの高解像度・フレーム単位での画質が高い。
  - ・720pや1080pの解像度にも対応可能。
  - ・ディテールを犠牲にせず、スタイルの一貫性を保つ。

- ⑤効率的な生成時間・他の動画生成手法に比べ、比較的短時間で結果を出力。
  - ・ノードベースのワークフローを使用することで、設定や修正が簡単。

<https://youtube.com/shorts/PPphtGmxKZY?feature=share>

### 今回皆さんに使っていただくのは【 Comfy UI 】

ComfyUIはStable Diffusionを扱うノードベースのインターフェースで、ユーザーが自由にワークフローを組み立てられる。特に動画生成において柔軟で視覚的に操作しやすい。



### 【 SVDとAnimateDiffのワークフロー演習 】

### ■ mimicPCを使ってComfyUIを起動してみましょう

皆様にお渡ししたワークフロー(json)を活用して動画生成を行います。

予備 <https://89.gigafile.nu/0601-a2600d9f07577f303ef40a3b573d9972>

### ■ Diffusionモデルを時間軸に適用したStable Video Diffusion (SVD)について

現在、Stabel Video Diffusionでは3つのモデルが公開されています。

#### 【 SVD 】

svdモデルでは14フレームの動画を生成することができます。フレームとは1枚1枚の描画される絵のことを指します。描画された絵をつなぎ合わせることで1つの動画が生成されます。

#### 【 SVD-XT 】...(ちょっと長くなった)

svd\_xdでは25フレームの動画を生成することができます。

#### 【 SVD-XT-I-I 】...(性能が向上した)

25フレームで動画を生成するモデル。出力が安定するように生成される条件が固定された上でsvd-xtiに対して微調整が行われています。

### ■ フレーム間の一貫性を強化するAnimateDiffについて

AnimateDiffは、Stable DiffusionにMotion Moduleと呼ばれる機構を付加することで、Text2ImageをText2Videoに拡張する手法です。以下のようなモジュールがあります。

#### 【 mm\_sd\_v14.ckpt 】

417Mパラメータ・1.6GB. ダイナミックな動きが特徴

#### 【 mm\_sd\_v15\_v2.ckpt 】

453Mパラメータ・1.7GB. 安定した動きが特徴

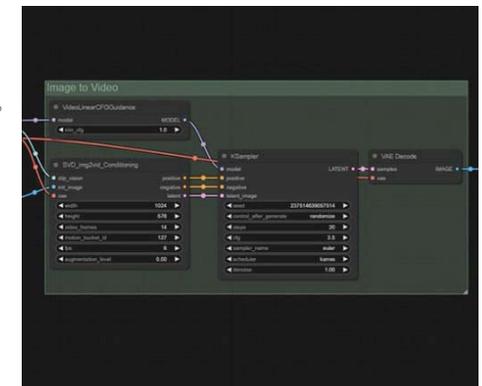
### ■ StablevideoDiffusion(SVD)パラメータの調整と生成までのステップ

- ①フレーム数: 20~30フレーム。
- ②解像度: 720p推奨(必要に応じて1080p)。
- ③ステップ数: 50~60。

動画生成の実行

「Generate」をクリック

ノードの接続を確認した後、生成を開始。



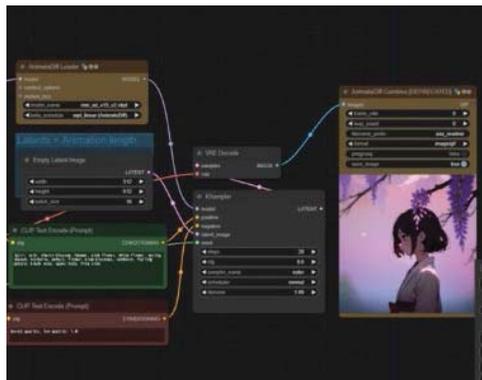
## AnimateDiffパラメータの調整と生成までのステップ

### 推奨設定

- ①フレーム数: 16~24フレーム。
- ②解像度: 720p。
- ③ステップ数: 40~50。

### 動画生成の実行

「Generate」をクリック



## 注意点とトラブルシューティング

### ①VRAMの不足

- ・高解像度・高フレーム数の場合、VRAM不足が発生する可能性あり。
- ・解像度やフレーム数を調整することで対処。

### ②成結果の品質向上

- ・プロンプトを具体的に記述。
- ・Negative Promptを適切に使用。

### ③エラー発生時の対応

- ・ノード接続を確認。
- ・モデルファイルやComfyUIのバージョンを最新に更新。



## LoRA学習について

### LoRA作成と応用技術

LoRA (Low-Rank Adaptation)の基礎から 応用技術を学び、  
Stable Diffusionでの高度な画像生成を習得する。

最終ゴール「LoRAモデルの使用から作成!」



## LoRA学習

LoRAとは、Stable Diffusionや他のAIモデルに新しいスタイルや  
機能を軽量かつ効率的に追加する技術です。

### ①軽量性

元モデルに小さなパラメータを  
追加することで、モデル全体を  
再トレーニングする必要がない

### ②高効率

VRAM使用量が少なく、  
コンシューマ向けGPU (8GB以上)  
でも扱いやすい。

### ③応用性

キャラクターデザインや  
アートスタイルの変更に  
適用可能。

要は、一番カンタンな機械学習。

## LoRA学習

### LoRAの主な用途

たとえば目とか...

色合いとか...

- ①キャラクターデザイン
- ②スタイル変換

特定のアニメキャラクターやスタイルを再現。  
古典絵画、漫画風、リアルな絵画などを  
特定のプロンプトで再現。



## LoRAの活用

### LoRA・ファインチューニングのビジネスへの活用

#### 「AI写真集」

LoRAを使用して、基盤モデルとなるStable Diffusionをファインチューニングすることで、「アジア人女性」などの特定の画像を生成できる。Amazon.co.jpの電子書籍読み放題サービス「Kindle Unlimited」をAI生成画像の「写真集」が席卷している。



#### 「AIモデル」

自社ブランドのファッションアイテムを学習させ、AIモデルに活用させることができる。AI model株式会社は、2022年6月14日にサービス「AI model モデル撮影サービス」を提供開始した。



## LoRAの活用

LoRAは、Stable Diffusionモデルと同様にファイル形式で配布され、以下の2つの主要プラットフォームで入手可能です。

### Hugging Face

①AI関連のさまざまなモデルを提供する総合的なプラットフォーム。

②Stable Diffusion関連のLoRAも配布されている。

### Civitai

①Stable Diffusionで生成された作品やモデルを共有するプラットフォーム。

②グラフィックを確認しながら、目的に応じたLoRAを簡単に探せる。

## LoRAの活用

### CivitaiでのLoRAダウンロード方法

#### STEP 1: CivitaiのLoRAページへアクセス

- ①Civitaiの公式ページ(<https://civitai.com>)にアクセス。
- ②「LoRA」カテゴリをクリックして、目的のLoRAを検索。

#### STEP 2: LoRAのダウンロード

- ①気になるLoRAをクリックして詳細ページを開く。
- ②「.safetensors」形式のデータをダウンロード。

#### STEP 3: Stable DiffusionにLoRAを配置

- ①Stable Diffusion Web UIを開く。
- ②「models」>「Lora」フォルダに移動。
- ③ダウンロードした「○○.safetensors」ファイルを配置。



【 LoRA学習 】

## LoRA学習の前準備

実はそんなに難しくなくて

明確なコンセプトが重要。  
「最低20枚」の高品質な画像が必要。

-大切なこと-

- ①必要情報以外を取り除く(背景を白または透過にする)。
- ②ツール例: PhotoShopやRembg(背景削除ツール)。
- ③高品質な画像を使用(低解像度やピンボケ画像は避ける)。
  - ④解像度は1024×1024以下に調整。
  - ⑤様々な表情を含める(任意)。

## LoRA学習の前準備

ただ、フォルダの名前の付け方とかが面倒くさい。  
学習データの整理が重要。

例: [繰返し回数]\_[トリガーワード]

例: 10\_okara。

『ファイル名のルール』

- ①画像を連番で命名(例: 001.png, 002.jpg)。
- ②ツール例: serial\_rename.py(連番付け用スクリプト)

では、早速mimicPCからKohya's GUIを立ち上げて作ってみよう。

## 【 LoRA学習の実行手順 】

### 『手順』

- ①Kohya's GUIを起動。
- ②データセット(教師画像フォルダ)を指定。
- ③トリガーワードを設定(例: okara)。
- ④学習パラメータ(学習率やエポック数)を設定。
  - ・推奨値: 学習率  $1e-4$ , エポック数 10。
- ⑤トレーニングを実行。

『注意点』 過学習を避けるため、適切な繰返し回数を設定。