

stable-diffusion で遊ぶ

なんか乗り遅れた感があるけど、遊んでおこうかな、と。

遊ぶだけなら [Stable Diffusion Demo](#) もいいけど、休んでいる GPU マシンでも動かしてみた。

セットアップは <https://github.com/CompVis/stable-diffusion> 通りに、Anaconda 使ってセットアップ

```
conda env create -f environment.yaml
conda activate ldm
mkdir -p models/ldm/stable-diffusion-v1/
```

で <https://huggingface.co/CompVis/stable-diffusion-v-1-4-original> から sd-v1-4.ckpt ダウンロードして

```
models/ldm/stable-diffusion-v1/model.ckpt
```

としてコピー。

想定されているのは VRAM が 10GB の GPU だけど、手元の GPU 環境は VRAM が 8GB な GTX1080 なので、

テキストから画像を生成する AI「Stable Diffusion」を Windows にインストールしてみたを参考に、

```
python scripts/txt2img.py ¥
  --prompt "a photograph of an astronaut riding a horse" ¥
  --plms --W 384 --H 256
```

と、サイズ小さ目に。

... 全然、オンラインデモの結果と全然違う。

折角なので、他にもいくつか試してみた。それぞれ、それっぽい(?)のを抜粋。

- a cute girl eating a piece of cake
- a boy eating a piece of cake

なんかいろいろ試している先人を真似してみると、それっぽいのが出力されてびっくり。うまいかないものも多いので --n_iter の値を変えて候補を増やした中から抽出。

これなんかは、ちょっと上手に怖い。

特に工夫せずに、

- japanese anime of a beautiful girl

だけだと

とか、そんな感じ。文字列で絵を描かせてる、って感じだなあ。

ちなみに、なんか、謎の歌手がでてくるのなんでだろうと思ったら、Twitter で教えてもらった。

<https://twitter.com/hotpepsi/status/1579022560419024897>

Rick Astley さんっていうのか。

... 別にエッチな画像だしたかったわけじゃないのに、ちょっとずかしい。

あらためて、実行結果みたら、

Potential NSFW content was detected in one or more images.
A black image will be returned instead. Try again with a different prompt and/or seed.

って、ちゃんと書いてあった。'A black image' じゃなくて洒落が利いてたけど。

こういう(?) 学術的なコード公開でも nsfw とか気にするあたり、
ディープラーニング境界は倫理関係に注意はらってるんだなあ、と、感心する。

単に描かせるだけじゃなくて、追加学習させたモデル使ってみるとか、
自分で追加学習させようとかも、いろいろあって楽しい。