

# Assignment 4: Diffusion

For this assignment, you will be using colab.

Please complete and submit this assignment by April 9, 11:59 PM. Download and submit the .ipynb file and share the notebook with the TA ([swetha.sirnam@ucf.edu](mailto:swetha.sirnam@ucf.edu) and [swethacrcv@gmail.com](mailto:swethacrcv@gmail.com))

## Useful Resources:

### **PyTorch Colab Documentation:**

<https://pytorch.org/tutorials/beginner/colab.html>

### **HuggingFace Sample Notebooks:**

<https://huggingface.co/docs/transformers/en/notebooks>

### **Stable Diffusion v1.5 (SDv1.5)**

Model Doc/Weights: <https://huggingface.co/stable-diffusion-v1-5/stable-diffusion-v1-5>

### **Kandinsky**

Model Doc: <https://huggingface.co/docs/diffusers/en/using-diffusers/inpaint>

Weights: <https://huggingface.co/kandinsky-community/kandinsky-2-2-decoder>

### **ControlNet**

Model Doc/Weights: <https://huggingface.co/lllyasviel/sd-controlnet-canny>

## Tasks:

1. Setup a Google Colab and load the above the models (SDv1.5, Kandinsky, ControlNet) onto GPU from huggingface and show memory usage for each model. **[10 points]**

## **2. Text-to-Image Generation [20 points]**

- a. Generate 10 sample images of your choice – Text-to-Image generation using stable diffusion and Kandinsky models and compare the generated images.

## **3. Generate Images for Intersectional Biases [35 points]**

- a. Generate images for at least 10 intersectional categories of your choice, with the focus to assess model bias
  - i. SocialCounterfactuals paper[I] generates synthetic images for different races, gender and occupations. Explore other categories of your choice like Religion, Socio-economic status, location etc. Generate 10 images spanning various intersectional biases.
  - ii. List down the categories first before generating images and then show images for each of them.

## **4. Human Redaction [ 35 points]**

- a. Take 10 celebrity pictures that has face clearly visible in the image. Perform human redaction on these images by replacing the human with cartoon caricature. Use ControlNet to maintain the original silhouette.

### **References:**

- I. Howard, Phillip, et al. "Socialcounterfactuals: Probing and mitigating intersectional social biases in vision-language models with counterfactual examples." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024