

Interaction with Large Scale Models

Practical Sheet 6

Multimodal Prompting

- 1- Set up Python + Replicate (on Google Colab)

```
!pip install replicate
import replicate
import os
```

- 2- Setup tokens.

```
os.environ["REPLICATE_API_TOKEN"] = "your_token_here"
```

- 3- Create a Basic Text-to-Image Prompt

```
output = replicate.run(
    "stability-ai/stable-
diffusion:db21e45c742c2926e4f9e0c5b88c0e6a5981d43f38326c6c0
c3cd69a1469b8a6",
    input={
        "prompt": "a cyberpunk city at night, neon lights, rain, ultra
detailed",
        "num_inference_steps": 30,
        "guidance_scale": 7.5
    }
)
print(output) # This will return a URL to the image
```

Image Generation

Next, observe how different prompt keywords change the generated image.

- Write 3 different prompts for the same subject:
 - "A medieval castle"
 - Add a **style**: "A low-poly 3D render of a medieval castle at dawn"



hugomcp@di.ubi.pt, 2024/25

- Add **mood + lighting**: “*A medieval castle on a stormy night, lightning in the background, dramatic lighting, digital art*”

In a Colab cell, images can be shown using `IPython.display.Image`.

Ambiguity & Specificity

Goal: Experiment with how ambiguous prompts result in default assumptions.

Task:

1. Prompt: “*a person using a computer*”
2. Modify to:
 - “*a female scientist in a lab using a computer, sketch style*”
 - “*a male teenager gaming on a laptop, anime style*”

Each time:

- Compare the generated images.
- Note what changed in **pose, context, style, demographic assumptions**.

For the exercises below, you should use one of the following platforms:

Platform	URL	Free	Notes
Mage.space	https://www.mage.space/	✓	Uses Stable Diffusion 1.5/2.1/SDXL
Playground AI	https://playgroundai.com/	✓	Account required, offers free generation
Hugging Face Spaces	https://huggingface.co/spaces	✓	Hosts many web-based models

Prompt Structure Exploration

In this exercise, the goal is to understand how prompt composition affects image generation.

1. Open <https://www.mage.space/>
2. Generate an image with this **basic prompt**:

“*A cat on a chair*”

3. Next, modify the prompt step-by-step:



hugomcp@di.ubi.pt, 2024/25

- Add **style**: "*A digital painting of a cat on a chair*"
- Add **setting and mood**: "*A digital painting of a cat on a chair in a cozy room at sunset*"
- Add **camera and lighting**: "A 3D render of a cat on a chair in a cozy room, warm lighting, cinematic, 35mm lens"

Take screenshots of all 4 images and compare.
Analyze the effect of style, mood, and camera terms.

2: Style Transfer by Prompt

Next, the goal is to explore how style descriptors control model outputs.

1. Choose a **subject**: "*A futuristic city*"
2. Create four prompts, each with a different **style**:
 - "*A photorealistic futuristic city*"
 - "*A cyberpunk futuristic city at night*"
 - "*A watercolor painting of a futuristic city*"
 - "*A pixel art futuristic city*"
3. Use Mage.space or Playground AI to generate the four variations.

Compare the results. Which modifiers changed colors? Geometry? Detail level?

Create your own custom style term (e.g., "in the style of Studio Ghibli").

3: Ambiguity and Specificity

As we've seen for the text case, vague prompts lead to variable outputs, particularly in multimodal prompting.

1. Try this **ambiguous** prompt:

"*A person working on a computer*"
2. Now **refine** it step-by-step:
 - Version 1: "*A young woman working on a laptop at a cafe, watercolor painting*"
 - Version 2: "*A male robot typing code on a holographic interface in a spaceship, digital art*"



hugomcp@di.ubi.pt, 2024/25

- Version 3: "*An elderly man working on a computer at night, lit only by screen glow, photorealistic*"

Generate and compare the outputs.

What did the model assume by default? How did specificity change the image?

4: Bias and Representation

Again, as it happens for the text-only prompts, model bias can seriously be determined prompt formulation. Also, note that this can be particularly hazardous in multimodal outputs.

1. Prompt:

"A doctor standing in a hospital hallway"

2. Observe the following features of the generated data.
 - Gender?
 - Ethnicity?
 - Attire?
3. Try variants to **correct or steer** the bias:
 - "*A female Black doctor in a modern hospital, wearing a stethoscope*"
 - "*An Indian male doctor smiling in a pediatric ward, digital painting*"

Take screenshots of the outputs.

Write down how the model reacted to demographic cues.

Discuss how prompt design influences fairness and representation.
