Generative AI Challenge: Character Consistency

Text-to-image models, such as Stable Diffusion, DALL-E, and Midjourney are not designed to maintain consistent character appearance across multiple images, which is essential for storyboarding and visual narrative tools.

For example, consider the following descriptions of two consecutive storyboard scenes:

- **Scene 1**: *Sarah stands on the edge of a cliff, gazing out over a vast forest at sunset. The wind blows through her hair.*
- **Scene 2**: *Sarah stands on the porch of an old cottage, talking with an elderly woman.*

The images generated by a Stable Diffusion model show Sarah as a different person with completely different features and clothes:

Scene 1

Scene 2

Your task in this challenge is to propose, implement, and test a solution to improve character consistency across image sequences using existing text-to-image models.

Therefore, you must implement a program that takes as input a sequence of textual scene descriptions (such as the descriptions of Scene 1 and Scene 2 presented above), and produce a sequence of consistent images illustrating those scenes.

The basic requirements for this challenge are as follows:

- Automation: the proposed method must function with minimal human intervention, relying primarily on automated processes to generate consistent images.

- Generality: The solution must be applicable to any well-described sequence of scenes. You should assume that the scene descriptions are generated by another AI system and are not known in advance.

To assist with this challenge, basic Python notebooks are provided (see Section 5) to help you understand the problem and the structure of the solution. Additionally, a template notebook for your final implementation is included (see Section 6).

The next sections of this document present some notes and resources related to the challenge. Please read this entire document before starting the challenge.

1. Communication Related to this Challenge

A dedicated Teams channel ("GenAI Challenge - Character Consistency") has been set up for this challenge. If you have any questions, encounter issues, or wish to discuss any aspect of the challenge, please use this channel.

The challenge organizer, Edirlei Soares de Lima, will monitor all discussions in the Teams channel and provide assistance when possible.

While this is an individual challenge, you are encouraged to use the channel to share your results and exchange ideas. Discussing concepts and sharing materials that may benefit others is highly encouraged.

2. Understanding Image Generation and Diffusion Models

If you are not familiar with image generation and diffusion models, it is strongly recommended that you begin this challenge by reviewing the following resources. These materials will provide you with the foundational knowledge necessary for understanding the topic:

- Introduction to Image Generation: https://www.youtube.com/watch?v=JR9Gdo-_lx8
- The Illustrated Stable Diffusion: https://jalammar.github.io/illustrated-stable-diffusion/
- An Introduction to Diffusion Models and Stable Diffusion: https://blog.marvik.ai/2023/11/28/an-introduction-to-diffusion-models-and-stable-diffusion/
- How does Stable Diffusion work?: https://stable-diffusion-art.com/how-stable-diffusion-work/
- Stable Diffusion – Wikipedia: https://en.wikipedia.org/wiki/Stable_Diffusion
- Deep Learning Foundations to Stable Diffusion: https://www.youtube.com/watch?v=_7rMfsA24Ls

If you find any additional resources that may be useful for this challenge, please share them with your peers in the Teams channel of the challenge.

3. Software and Server Access

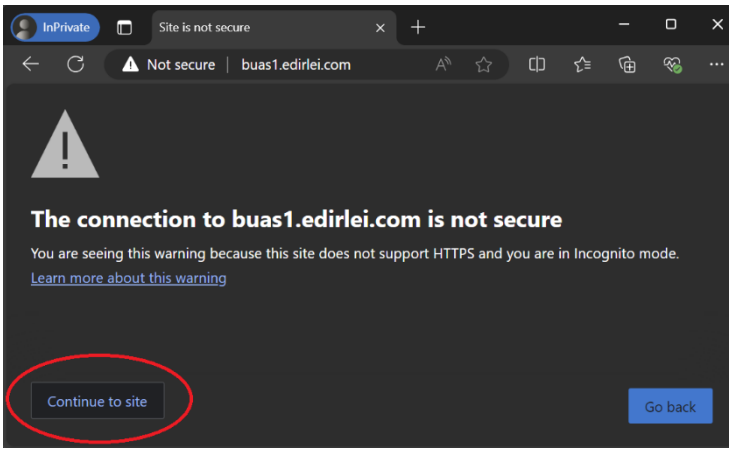
For this challenge, you will be using two popular frameworks for running and fine-tuning Stable Diffusion models: **ComfyUI** (https://github.com/comfyanonymous/ComfyUI) and **WebUI** (https://github.com/AUTOMATIC1111/stable-diffusion-webui). Although it is possible to run these systems and large models locally, doing so requires a high-performance GPU. Therefore, instances of ComfyUI and WebUI are available on the ADS&AI server for you to use during this challenge. You can access the instances using the following links:

- **ComfyUI**:
  - http://buas1.edirlei.com/
  - http://buas2.edirlei.com/
  - http://buas3.edirlei.com/
  - http://buas4.edirlei.com/
- **WebUI**:
  - http://buas5.edirlei.com/
  - http://buas6.edirlei.com/
  - http://buas7.edirlei.com/
  - http://buas8.edirlei.com/

The authentication details to access the systems are:

- Username: adsai
- Password: uuUP4whjX29cF3cxwrX3SY5Mm9TmtASkdgpaNCTHZ9

When accessing the links for the first time, your browser might show a security warning (like the one shown in the image below) because the systems are not running over HTTPS. In this case, you can simply click on the continue button to proceed to the system.



Important Notes:

1. **System Load**: As these systems are shared by all participants, they may occasionally become overloaded. If image generation takes longer than expected or you encounter access issues, consider switching to another instance or waiting a few minutes before trying again.
2. **Modifications**: You are allowed to modify settings and install components on these systems. However, please ensure you document all changes in case you need to replicate them on another instance. If an instance fails, it will be restored to its original state.

Ollama Server

Additionally, an **Ollama Server** (https://ollama.com/) is available on the ADS&AI server to support implementations that require querying large language models, such as Llama 3.1. The Ollama API is accessible at: http://buas9.edirlei.com/. An example notebook on how to execute queries using Ollama is provided in Section 5 of this document.

If you want to run other local large language models (check the supported models here: https://ollama.com/library), please post the details of the model in the Teams channel dedicated to the challenge. The organizer of the challenge will assist you in installing the models when possible.

4. Open Image Generation Models

The instances of ComfyUI and WebUI running on the ADS&AI server come preloaded with several Stable Diffusion models:

- Stable Diffusion XL Base: https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0
- Stable Diffusion XL Turbo: https://huggingface.co/stabilityai/sdxl-turbo
- Juggernaut XL: https://civitai.com/models/133005?modelVersionId=357609
- CyberRealistic: https://civitai.com/models/15003?modelVersionId=846876
- AbsoluteReality: https://civitai.com/models/81458?modelVersionId=132760
- Anything XL: https://civitai.com/models/9409/or-anything-xl
- Stable Diffusion 1.5: https://huggingface.co/stable-diffusion-v1-5/stable-diffusion-v1-5

How to add other models?

In ComfyUI, you can add additional models using the Manager:

If the model you need is not available in the Manager, or if you need to install a model in WebUI, please post the details of the model in the Teams channel dedicated to the challenge. The organizer of the challenge will assist you in installing the models when possible.

Where to find models?

- https://civitai.com/models
- https://huggingface.co/models?pipeline_tag=text-to-image&sort=trending

Custom Nodes and Extensions in ComfyUI?

ComfyUI also supports the installation of custom nodes and extensions via the Manager. If you cannot find the custom node or extension you require, please create a post in the Teams channel with the relevant details of what you need.

5. Example Notebooks

The following Python notebooks are provided to assist you in understanding the problem and the structure of the solution you are expected to implement:

- Notebook 1: Image generation and storyboard creation using ComfyUI (without any method to ensure character consistency). Link: https://bit.ly/3NLFQbe
- Notebook 2: Image generation and storyboard creation using WebUI (same as Notebook 1, but using the WebUI framework for image generation instead of ComfyUI). Link: https://bit.ly/4hrYeDY
- Notebook 3: A basic example of using Llama to enhance scene descriptions. Link: https://bit.ly/3UqpIQ8

Important: to run Notebook 1, the following file is required: https://bit.ly/4fnNIvn. Save the file workflow_juggernaut_xl.json in the same folder as your notebook.

6. What are the possible approaches to solve the consistency problem?

There are several potential approaches you can explore to address the problem of character consistency in image generation. Below are some possible directions to consider:

1. Prompt-based solutions: investigate automated methods to enhance the prompts provided to the image generation models. Improving the clarity or specificity of prompts can help guide the models toward more consistent character representations.
2. LoRA-based solutions: Low-Rank Adaptation (LoRA) models are fine-tuned versions of Stable Diffusion models that can be trained to represent specific characters. Certain LoRA models can be activated to ensure that specific characters appear consistently in the generated images. Some examples of existing LoRA models include:
   a. Star Wars Characters: https://civitai.com/models/206995?modelVersionId=237150
   b. Female Models: https://civitai.com/models/671782?modelVersionId=934538
3. ControlNet-based solutions: ControlNet allows the use of a neural network to control various aspects of image generation in Stable Diffusion, such as specifying poses, copying compositions, or generating images that closely resemble a reference image. This approach can be highly effective for maintaining character consistency across scenes.

7. Evidencing the Results of the Challenge

A template notebook for submitting your final solution is available at:

https://bit.ly/4e0T6DK

The template notebook is divided into six sections, which you are expected to use for developing and reporting your results. The sections are as follows:

1. Solution Overview: Provide a brief explanation of your approach to solving the character consistency problem.
2. Load Story: Describe the process of loading the story for image generation.
3. Preprocess Scene Descriptions: Outline any preprocessing steps applied to the scene descriptions.
4. Image Generation with Improved Character Consistency: Present your method for enhancing character consistency, detailing the solution step-by-step.
5. Results and Discussion: Explain the results obtained from your solution and discuss their significance.
6. Conclusions: Summarize your key conclusions, challenges faced, limitations of the proposed solution, and potential future improvements.

Important: it is strongly recommended to document everything that you do for the challenge in the template notebook. The challenge is about effort, so if you can show that you dedicated

enough time to it, you can obtain points for the challenge even if you do not achieve good results in the end.

When you complete the challenge, you must submit the template notebook with your solution on the Brightspace assignment of the challenge:

Brightspace → ADSAI General Information 2024-2025 → Course Elements → Assignments → Generative AI

Good Luck!