### **Assignment: Image Generation by maintaining different aesthetics from the Movie *Bajirao Mastani***

#### **Goals: In the developed text-to-image model, the prompts should lead to images with the one or all of the following characteristics:**

* **Character Consistency**: Ensure consistent portrayal of Ranveer Singh’s character throughout the movie, across various poses, expressions, clothing, and scenes.
* **Background Consistency**: Maintain the scene backgrounds reflecting the unique atmosphere and visual style seen in the movie.

**Preparing training dataset**: Annotate and categorize images provided in the dataset to first prepare supervised data to train text-to-image models and ensure consistent outputs in image generation tasks.

**Data Files Overview**

The following files will be provided to assist in the data annotation process:

1. **Clips.json**  
   This file contains information about the captions for specific video clips. It provides detailed descriptions of the video segments, which are essential for generating precise annotations for the video content.
2. **Frames.json**  
   This file includes information about individual frames, such as the frame number and its corresponding description. It is used for annotating each frame accurately and associating it with the appropriate visual context and details.
3. **Transcripts.json**  
   This file contains the full transcripts of the movie, including all captions throughout the film. The information provided in the transcript may be valuable for generating detailed data annotations, as it gives context to the dialogue and narrative, which can be leveraged for improving caption accuracy and scene-specific annotations.

### **Image Generation Tasks**

#### **1. Bajirao Pics: Face Consistency**

This task focuses on gathering a variety of facial expressions, poses, and clothing variations of Ranveer Singh’s character throughout the movie. These images represent different scenes in which Bajirao (Ranveer Singh) appears with varying expressions and attire.

* **Image IDs for Ranveer Singh’s Face Consistency**:
  + 283, 329, 332, 531, 853, 1004, 1228, 1257, 1414, 1419, 1766, 1792, 2063, 2147, 2491, 3569, 3580, 3692, 3741, 3754, 3789, 3847, 4025, 5335, 5349, 5359, 7446

These images should be analyzed to ensure that the character's face remains consistent across various emotions, poses, and clothing, contributing to the overall character continuity in generated images.

#### **2. Cloth Consistency**

Bajirao’s clothing remains consistent throughout the movie. The task here is to study his attire in different scenes, ensuring the clothing is accurately maintained across all generated images of Bajirao.

* **Image IDs for Cloth Consistency (Ranveer Singh)**:
  + 4309 - 4517 (Song 1 of Ranveer Singh)
  + 3027 - 3413 (Song 1 of Deepika Padukone)

These images should be referenced to understand Bajirao's costume design, ensuring consistency in clothing when generating new images of the character.

#### **3. Scene Aesthetic Consistency**

The key goal of this task is for the AI model to learn the aesthetic of the movie's scene settings. This includes the architectural style, lighting, mood, and overall ambiance of the scenes in *Bajirao Mastani*.

* **Scenes to Focus On**:
  + 3027 - 3413 (Song 1)
  + 7123 - 7353 (Song 2)

The task involves training the model to understand the distinctive aesthetic of these settings and using text prompts to generate images that reflect these settings accurately. The AI model should be able to replicate the feel of the scene, focusing on factors like color grading, architectural elements, and overall atmosphere.

**Data Annotation Tasks**

**1. Image Annotation Task**

In the frames.json file, each entry contains the image URL along with a basic caption. The task requires generating precise and detailed annotations, including the creation of accurate and sophisticated masks. These annotations should not only align with the basic captions provided but should also reflect a higher level of accuracy and detail, ensuring that all relevant features within the image are captured comprehensively.

**2. Video Annotation Task**

The clip.json file includes timestamps, specifying the start and end points for each video segment. The task is to extract the relevant video footage corresponding to these timestamps, followed by annotating the video using an AI model. A ground truth (GT) video description will be provided, and the objective is to enhance this description by generating a more sophisticated and detailed prompt. The goal is to create accurate and insightful annotations that extend beyond the basic descriptions provided, offering a deeper understanding of the video content.