Diffusion Model prompt Analysis – uncovering patterns in large diffusion prompt datasets

1. Introduction

Recent advancements of Generative Artificial Intelligence (AI) systems have propelled artificial intelligence into public consciousness. Diffusion-based systems such as Midjourney can be used to generate high quality images from human-language text prompts. Compared to previous image generation models these novel systems produce images on par with contemporary artists, even winning art contests.

Diffusion-models have been made accessible to the public by different services (e.g. Midjourney.com, Stable Diffusion Web), since then users created millions of images. They explored the services’ capabilities and shared their images together with the prompts online. Users quickly found tricks to influence the quality, fidelity and style of the generated images by modifying the text prompt.

There has already been some research into the user-generated text-to-image prompts [1]. Prompts can be analyzed to find popular image content, style descriptions and more. For example, the researchers of [1] analyzed the prompts’ primary elements and further related them to traditional photography concepts.

Our aim is to answer the following questions, by analyzing patterns within the prompts: Which artists occur conjoint with other artists? Which eras are they representing? Which styles are the artists known for?

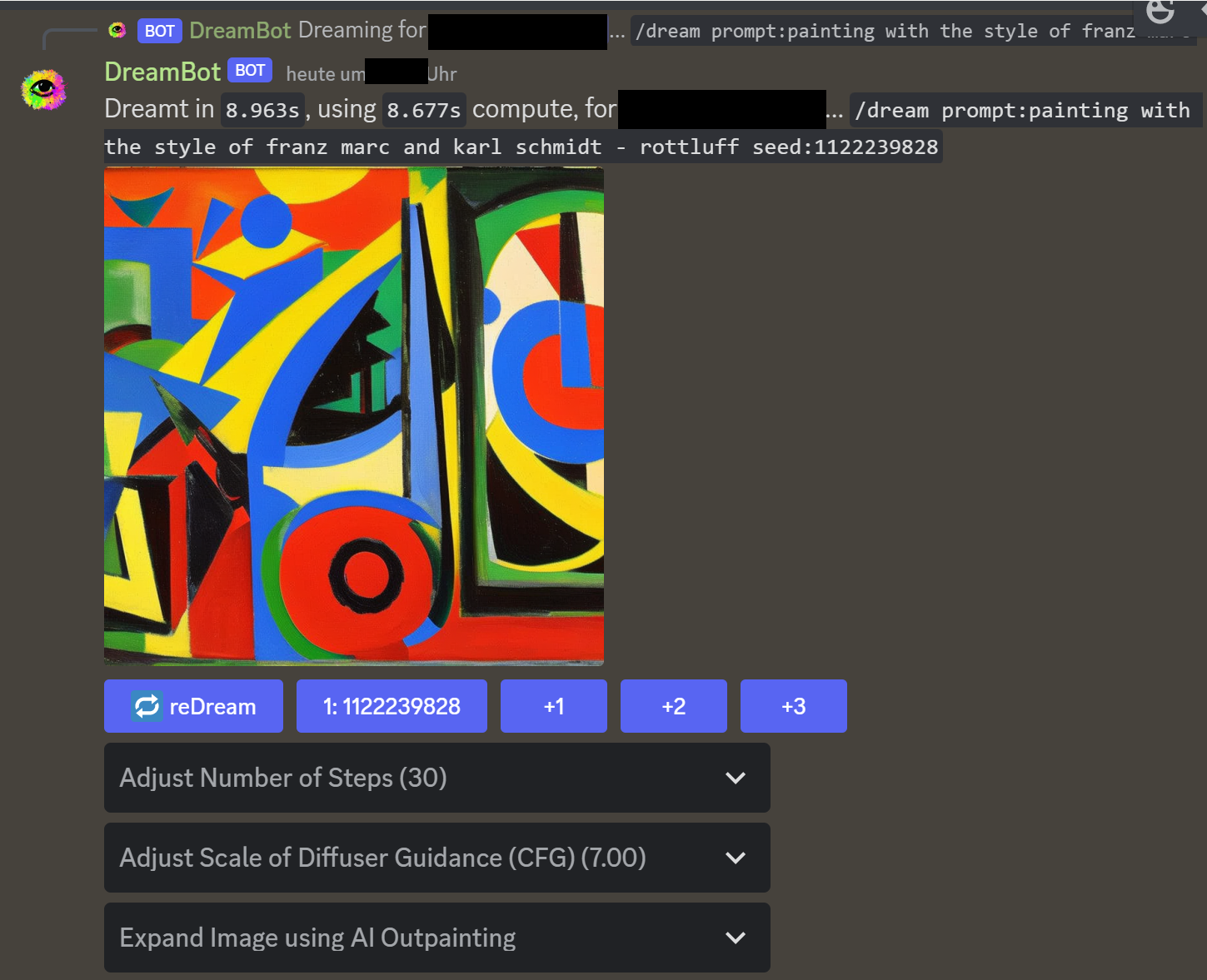
We want to investigate these questions to find the artists, styles and eras most appealing to users working with image generating ai.

2. Data

There are a lot of user-generated prompts available online (e.g. <https://lexica.art/> and <https://prompthero.com/>), however it is difficult to query some of these websites. Furthermore, the different websites and services format the prompts differently and may include other parameters such as aspect-ratios as well. Various datasets for experimentation and research have been compiled and published on huggingface.com, some of these datasets consist only of the prompts while others include prompts, images and metadata such as usernames.

We will use a publicly available dataset of prompts from huggingface, for our research we will only analyze the pure prompts without the produced images and metadata.

We will use the [diffusiondb](https://huggingface.co/datasets/poloclub/diffusiondb) dataset by poloclub from huggingface. It contains 1.8 million prompts of the Stable Difussion Discord Server and is publicly available.



Stable Difussion Discord Server generated image with the prompt: painting with the style of franz marc and karl schmidt - rottluff

We still need to decide which references to use to obtain descriptions of artists, styles and eras. One solution could be [this list](https://docs.google.com/spreadsheets/d/14xTqtuV3BuKDNhLotB_d1aFlBGnDJOY0BRXJ8-86GpA/edit#gid=0) of another study.

3. Methods

By using the co-occurrence method, we want to calculate the frequency with which each variable occurs in combination with other variables within a prompt. Therefore we would for instance like to find relationships such as „Franz Marc“ occurs conjoint with „Karl Schmidt-Rottluff“.

By using topic modelling we then want to go a step further and use machine learning algorithms to classify prompts into different topics based on the words that appear in them.   
The artists used in the prompt would be our proxys for an era or style of art (topic), which is labeled in our reference dataset, as seen below.

Ein Bild, das Tisch enthält.

Automatisch generierte Beschreibung

Ein Bild, das Text enthält.

Automatisch generierte Beschreibung

Consequently the aim of the prompt containing both artists could be to generate a vibrant and expressionistic image.

By using these two methods together, we can gain a deeper understanding of our data and potentially uncover insights that would otherwise go unnoticed.

4. Conclusion

We hope to get an interesting look into the users interests and therefore find patterns between artificially generated art and classic art.

Citations

# **[1] What is in a Text-to-Image Prompt: The Potential of Stable Diffusion in Visual Arts Education**

https://arxiv.org/abs/2301.01902