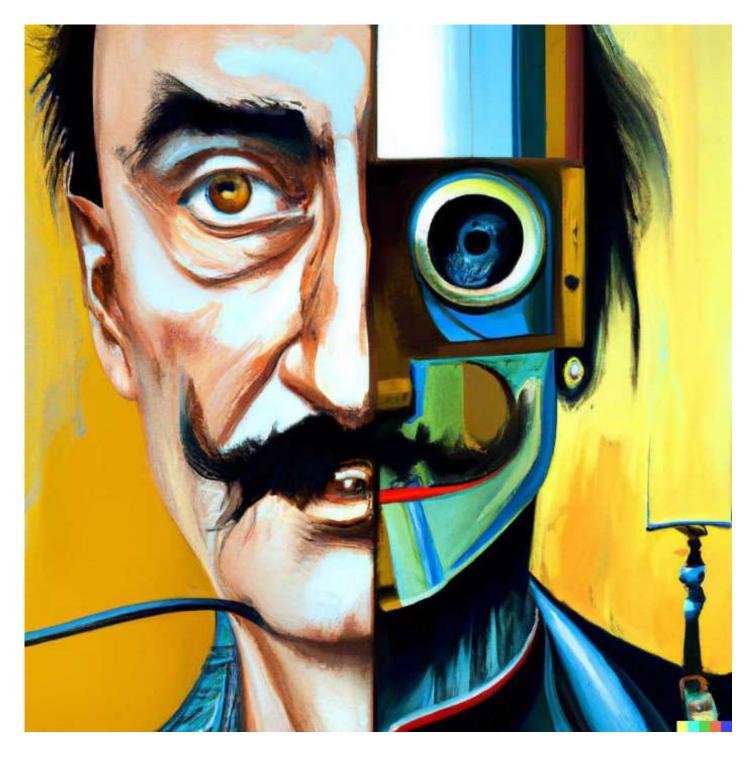# NLP for Text-to-Image Generators: Prompt Analysis [Part 1]

A full-code tutorial on how to optimize your text-to-image prompts using sentence similarity and KMeans Clustering

Text-to-image generators are currently a red hot topic in the field of AI art. With them, a user can provide text describing the artwork they'd like to output, and the machine generates different variations of this image (sometimes in less than a minute!). Deep learning powers this technology, which helps anyone without design skills create artwork with just their sheer imagination.

These AI systems have been trained using millions of images, along with their captions.

Some companies have trained their own AI "artists" and made them publicly available for people to use. Examples of such products include:

- DALL-E 2 by OpenAI is an AI system that pioneered this field. It is currently in public closed beta and you need to purchase credits to use it.
- Midjourney is an independent research lab that offers an AI program that creates images using textual descriptions. It is integrated with discord via a bot.
- Stable diffusion is a latent text-to-image diffusion model capable of generating photo-realistic images, given any text input. Created by Stability AI, they open-sourced a GitHub repository that has pre-trained models.

The main work involved in using these AI systems is coming up with textual descriptions of visuals you'd like to create. These descriptions, called "prompts," can be as vague or as detailed as you'd like. The more specific your prompt is, the higher the level of fidelity of the image generated. The use of specific keywords can help to boost the quality of your image. For your reference, check out the Dalle-2 Prompt Book to get started with creating quality prompts.

In this article, you are going to analyze a dataset of 200K+ prompts created by Midjourney users. This dataset is available in HuggingFace and you are going to use it to:

- Create and visualize word embeddings;
- Perform a semantic search that will help you find similar prompts;
- Explore topics using clustering algorithms and visualize the keywords in each cluster.

■