# Background - What are diffusion models

# Background - Text-To-Image

# Background - Stable Diffusion

In Stable Diffusion there are two set of layers known as the self-attention layers and the cross-attention layers. Generally speaking, self attention focuses on relationships within the same sequence, in Stable Diffusion, it allows the communication between intermediate and context features to create globally coherent structures. Whereas cross attention is conditional on textual features and largely considers the relationship between the textual features and the to be outputted image features (aka the intermediate image features).

# Background - DenseDiffusion

DenseDiffusion is a training free modulation which can be applied to text-to-image models to improve accuracy when given a dense caption. A dense caption is one which includes a lot of specific details. DenseDiffusion works by modulating the attention layers of a text-to-image model. Specifically in the case that is covered in CMU’s paper applies there technique to stable diffusion.