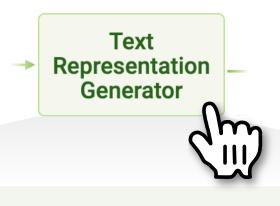
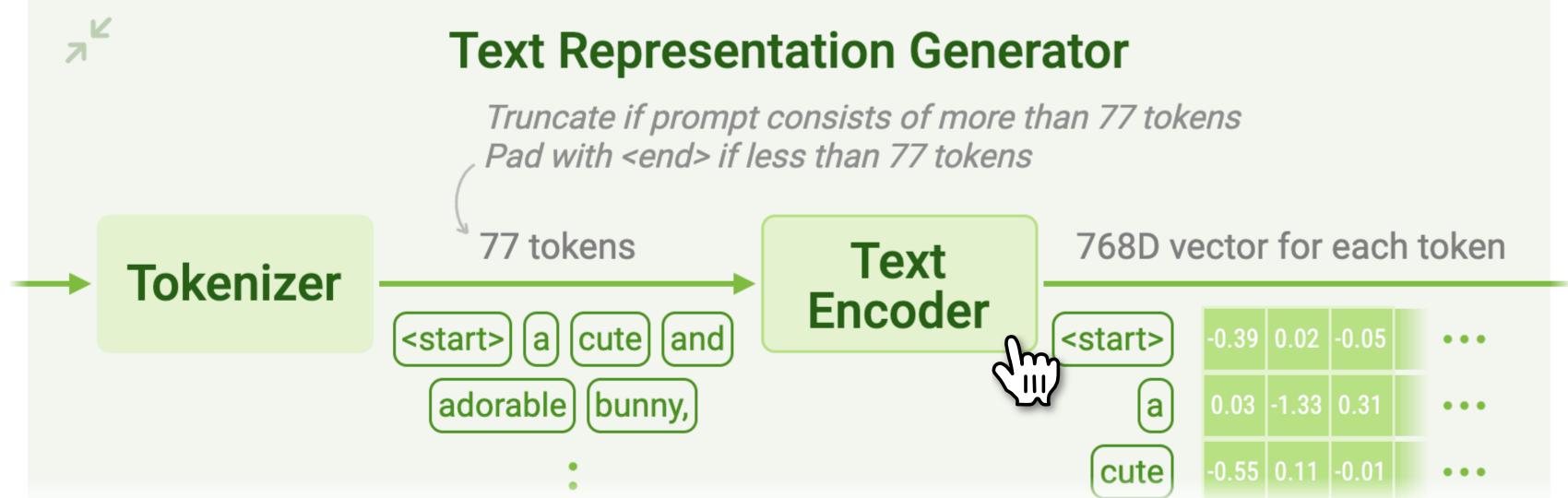
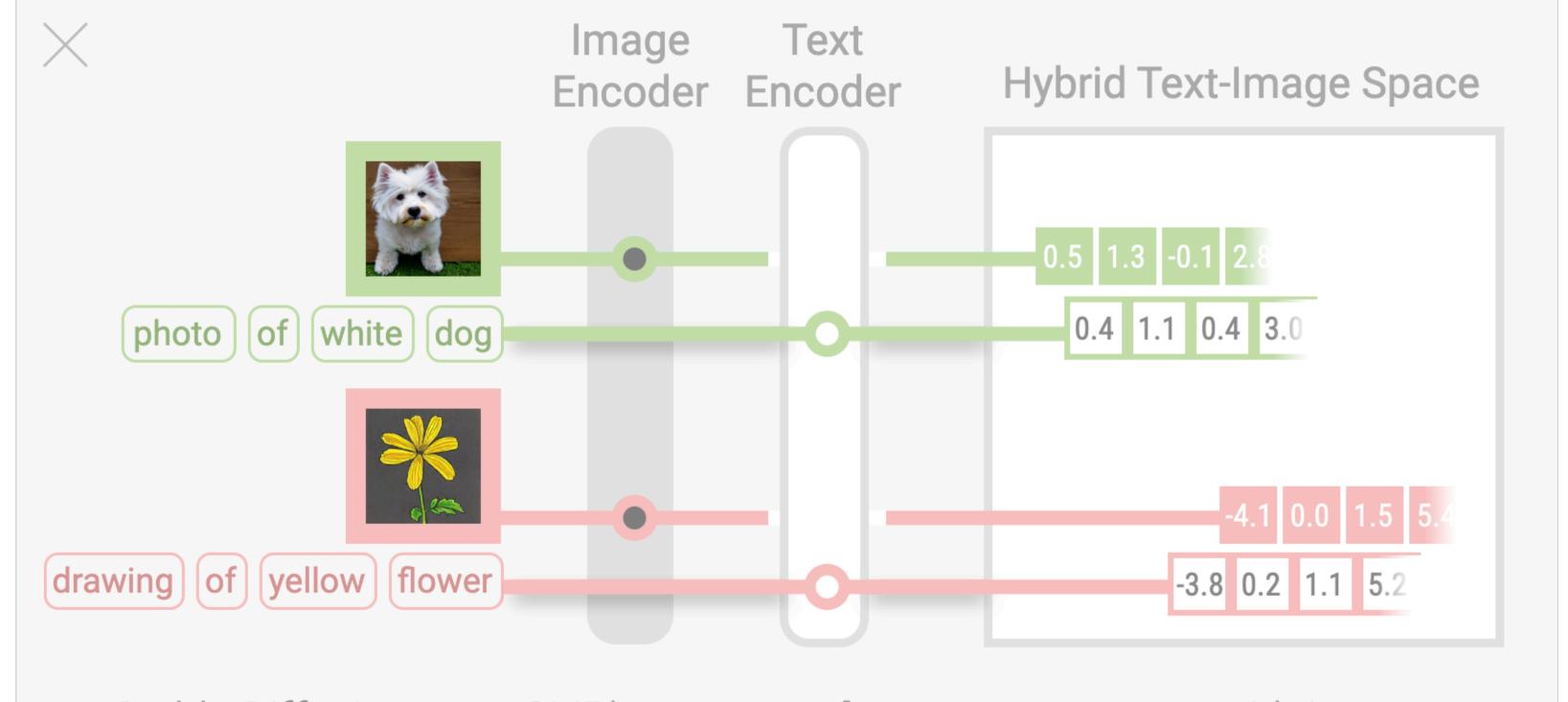
Text Operation View





Text-image Linkage Explanation



Stable Diffusion uses CLIP's **text encoder** to connect texts with images. For an image-caption pair, CLIP's text encoder converts the caption's tokens into a vector that is close to the image's vector (produced by CLIP's image encoder) in hybrid text-image space.