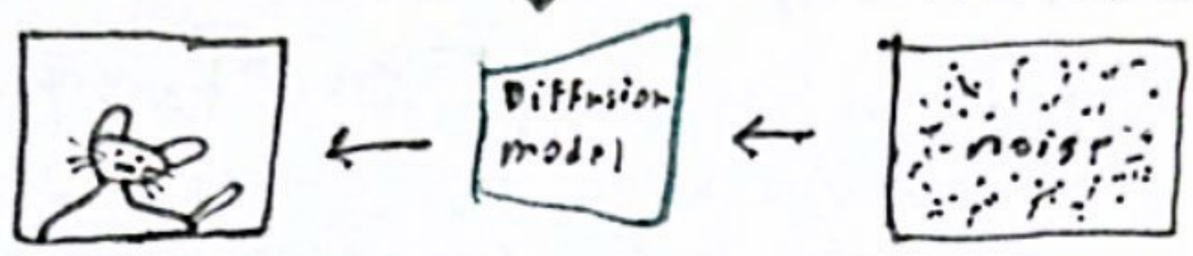
Conditioning (Gen ai notn)

- But how do we use the embedding vectors from CLIPs text encoder and img encoder to steer the diffusion process?

- One option is to pass our text vector as another input into our diffusion model and train as we would to remove noise.



- if we use img and caption pairs to train the diffusion model the model will learn to use the text information to more accurately remove noise from images since it now has more context about the image that its trying to denoise. This method is Conditioning

- There are many ways to pass in text to the diffusion model which would be a Transformer. One way is to use cross attention Like in LLM, another is to add or append the embedding text vector to the diffusion model input. and etc

- In practice it turns out conditioning is not enough to achieve DALLE 2

- Lets return to the spiral like mentioned before different sections in the spiral may correspond with Different types of imgs so when training the diffusion model When passing the cordinates and time pass in points class of img like "cat" this can help steer points to the right section of our spiral based on the points class. But still its not perfect we see confusion between dog and person for ex, this is because we are trying to get a realistic img by landing our points on the spiral and towards specific classes on the spiral.