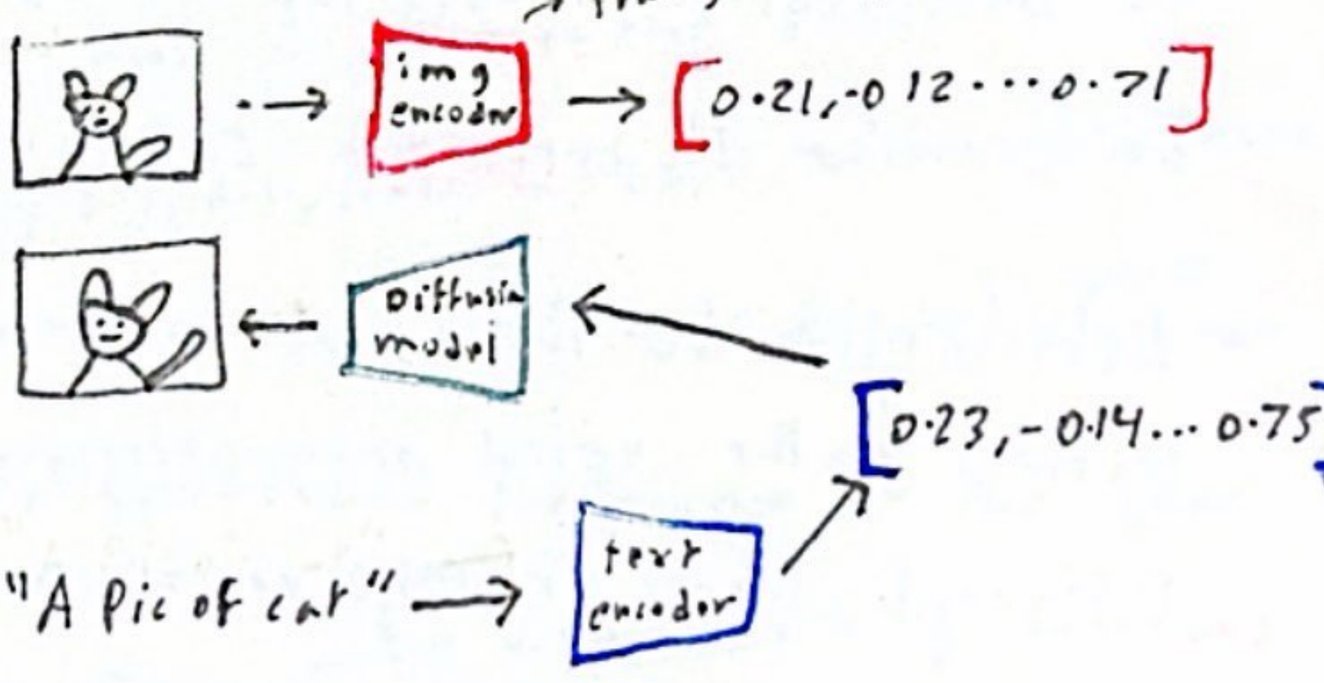
Dall E 2 (unCLIP) (generators)

- So far we have discussed how to generate images but the ability to steer the diffusion process (ie add detail to imgs and get more accurate results of what we want) using text prompts is very limited

- in clip we saw how it was able to learn a shared representation of imgs and text by concurrently training img and text encoders models. BUT these models only go one way converting text or imgs into embedding vectors.

- Diffusion models are potentially able to reverse the CLIP img encoder, generating high quality imgs. and the output vector of the CLIP text encoder could be used to guide the diffusion model towards the img or vidros that we want

- So the idea is we can pass a prompt into CLIP text encoder to get a embedding vector and use it to steer the diffusion process towards the img or vid that our prompt discribes



- A team at OpenAI did just this in 22, using pairs of img and caption pairs to train a diffusion model to invert the CLIP img encoder there approach got them a incredable level of prompt adherence (color or detail from input text) this model was called unCLIP and the model was called DALL E 2