Very high level



## Demonstration data



Refina (nAH PI-OT-OPT MT-OPT BC-Z-CnATO Auto regressive (onditioning on Action



## Data collection



Types of ablation studies

- 1. Capacity and expressiveness of the model
  - A. Changing number of parameters from from 35 M to 21 M
  - B. Removing transformer and using pre-trained efficient net instead
- 2. Action representation
  - A. Switching from discrete to continuous actions (normally distributed)
  - B. Ablating auto-regressive action representation
- 3. Image Net pre initialization
  - A. Instead of using image Net weight use random initialization
- 4. Access to short history
  - A. Instead of using past 6 images use only one image.