

Cross-environment & Cross-embodiment 7B VLA 7B VLA (LWM) (LWM) 7 DOF Action (256 x 6) + 1 (gripper) 1,537 action labels **Pseudo Action Labels** (seg len x codebook size) (e.g. $8 \times 8 = 64$ action labels)

Stage 1 : Latent Action Model (unsupervised) Training

Stage 2 : Supervised Fine-tuning VLA on pseudo latent action labels

Stage 3 : Fine-tuning VLA on Teleoperation data of 150 trajs (real action labels)