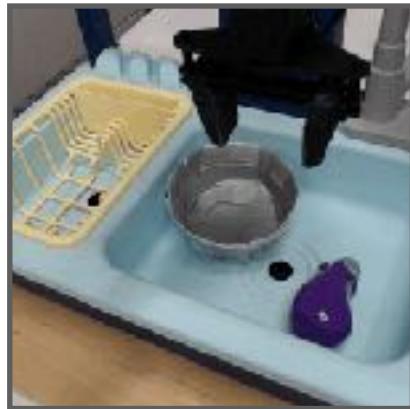


OpenVLA



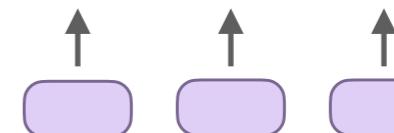
Input Image

“Put eggplant
in bowl”

Language Instruction

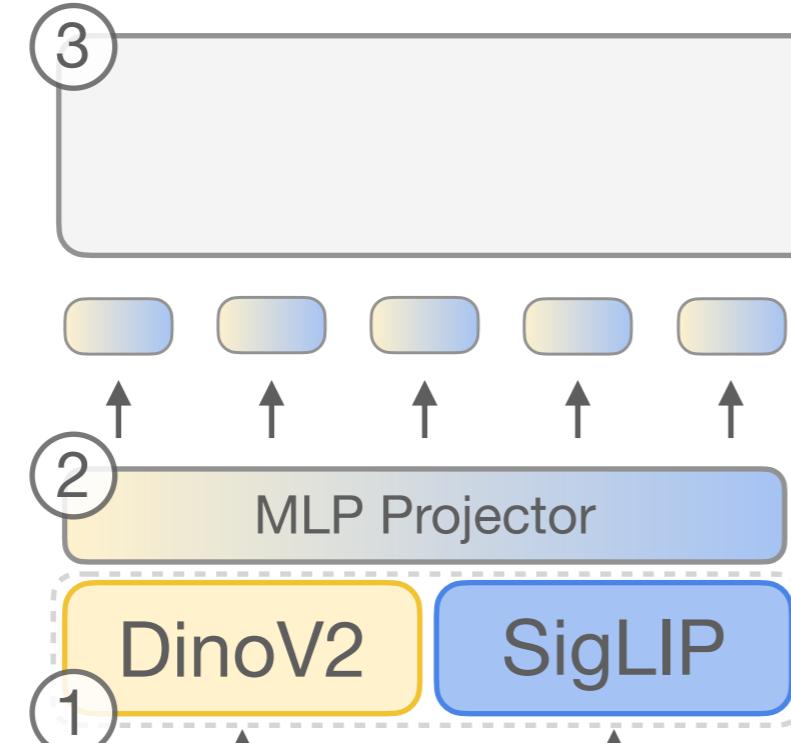
Llama 2 7B

Action De-Tokenizer



Δx
 $\Delta \theta$
 ΔGrip

7D Robot
Action



Llama Tokenizer

“What should the robot do to {task}? A:”