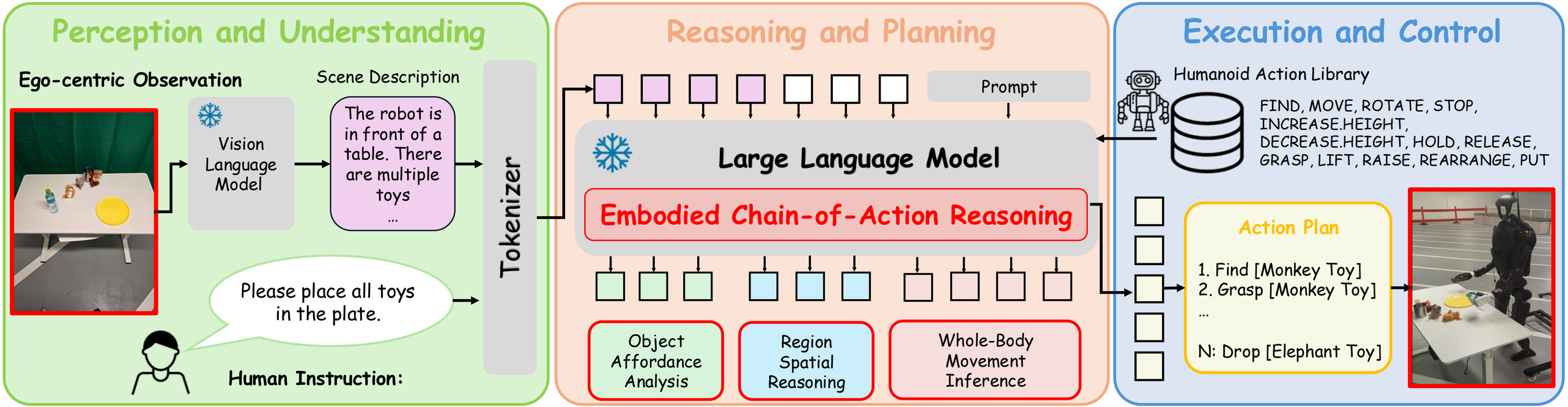
Perception and Understanding
Ego-centric Observation
Vision Language Model
Scene Description
The robot is in front of a table. There are multiple toys ...
Please place all toys in the plate.
Human Instruction:
Tokenizer
Reasoning and Planning
Prompt
Large Language Model
Embodied Chain-of-Action Reasoning
Object Affordance Analysis
Region Spatial Reasoning
Whole-Body Movement Inference
Execution and Control
Humanoid Action Library
FIND, MOVE, ROTATE, STOP, INCREASE.HEIGHT, DECREASE.HEIGHT, HOLD, RELEASE, GRASP, LIFT, RAISE, REARRANGE, PUT
Action Plan
1. Find [Monkey Toy]
2. Grasp [Monkey Toy]
...
N: Drop [Elephant Toy]