Answer
ID/bbox

Code

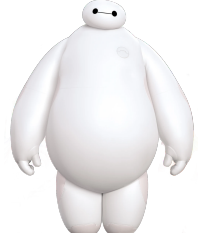
LLM

Zero-shot prompting

Fine-tune by self-correction

Reason

"Hey Baymax, can you help grab me ***the chair in the corner of the room, between the white and yellow desks***?"



3D Object Detector

Spatial-Semantic info of *all* objects

Spatial-Semantic info of *relevant* objects

Object Filter