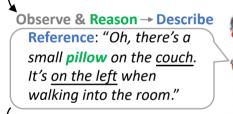
## 3D Visual Grounding

Intention: "I want something to support my back to relieve the pressure."



## **Observe & Understand Reference**

→ Detect *pillow* 

**Response**: "It is at [x = 1.3,y=2.0, z=0.5, w=0.6, h=0.5, d=0.2]."



## 3D Intention Grounding





 $\bar{\ \ }$  satisfy the intention."

Response: "Maybe you can use the object at [x = 1.3, y=2.0,

z=0.5, w=0.6, h=0.5, d=0.2]."