

# LVLM Reasoning

"Can you please find the blue rug in the bathroom?"

Targets = [A blue rug, a bathroom]

a)

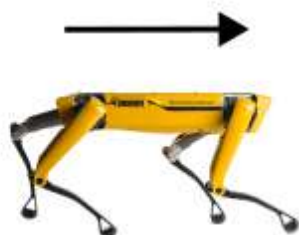


## Semantic Navigation



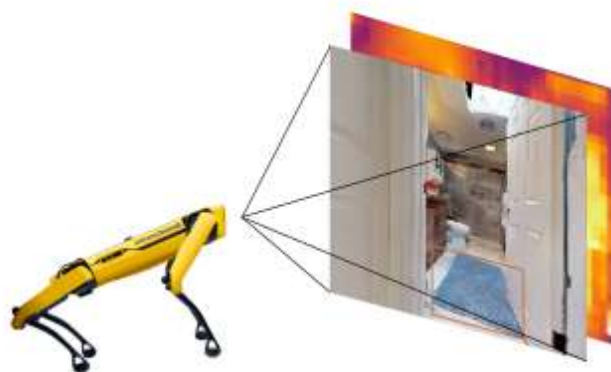
Environment observations

Map building  
and  
Robot navigation



b) Combined semantic similarity map

## Object Detection



A blue rug detected!

c)

## Validation

LVLM validation output =  
"The image shows a bathroom.  
There is a small blue rug  
positioned on the floor. The blue  
rug is inside the bathroom"

Validation SUCCESS, target  
FOUND!

d)