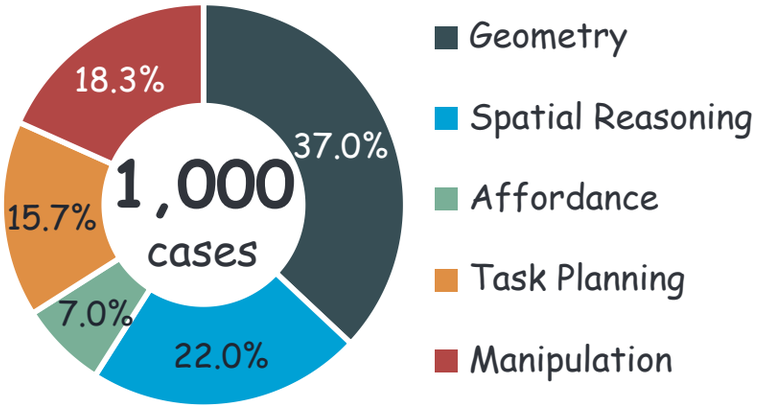
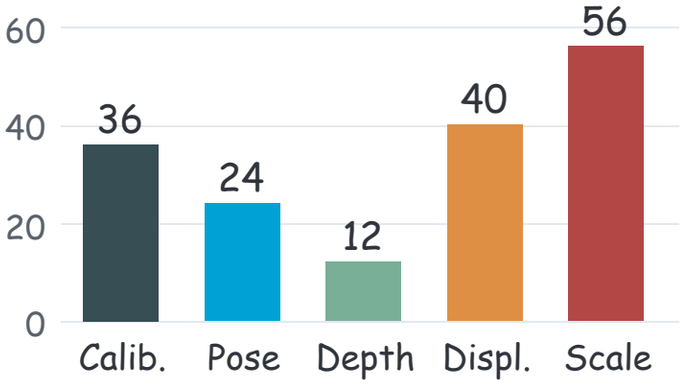
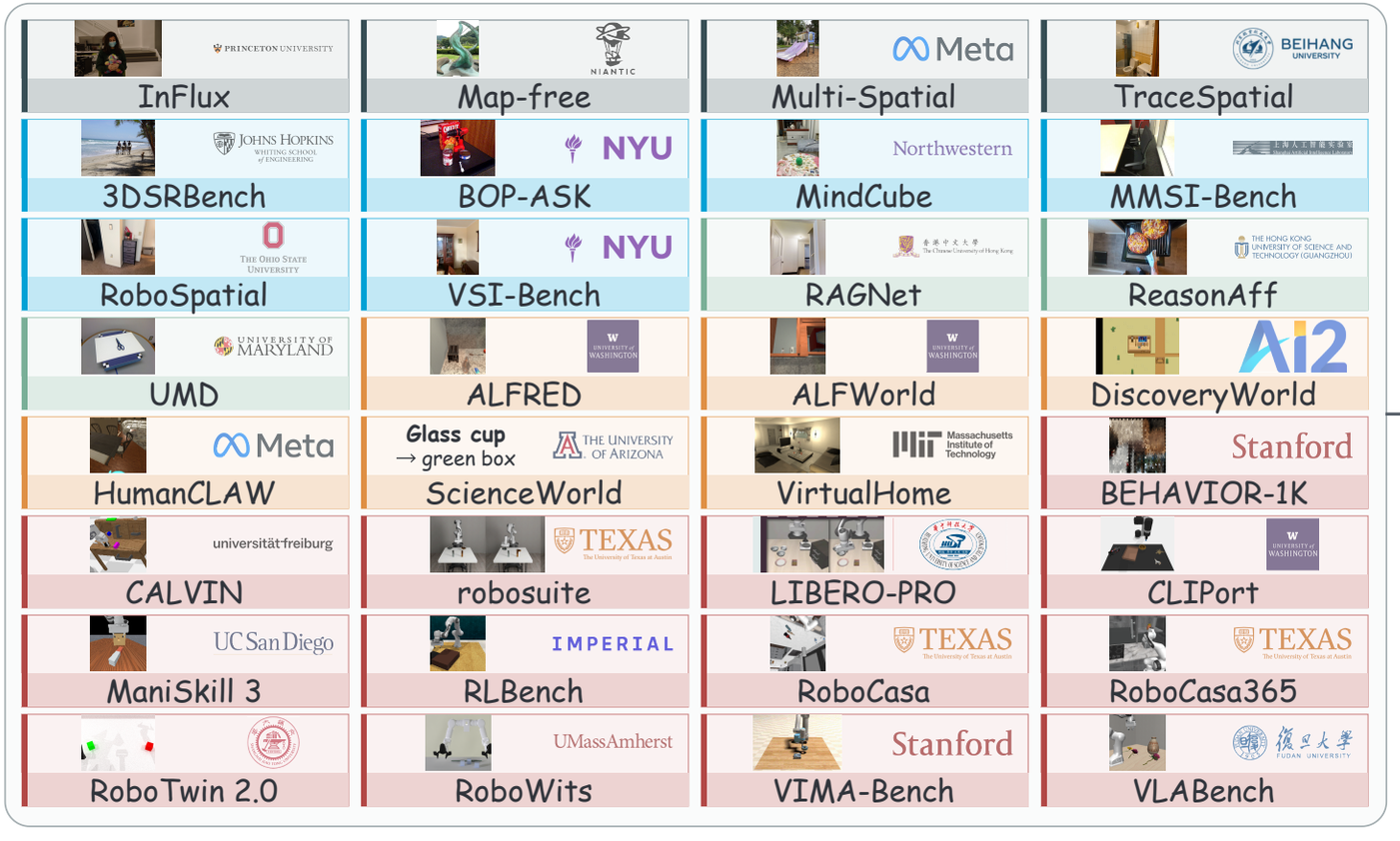

object
left
camera
right
person
hand
center
forward
find
touch
move
place
front
back
put
table
above
box
room
pull
turn
push
sit
(a)

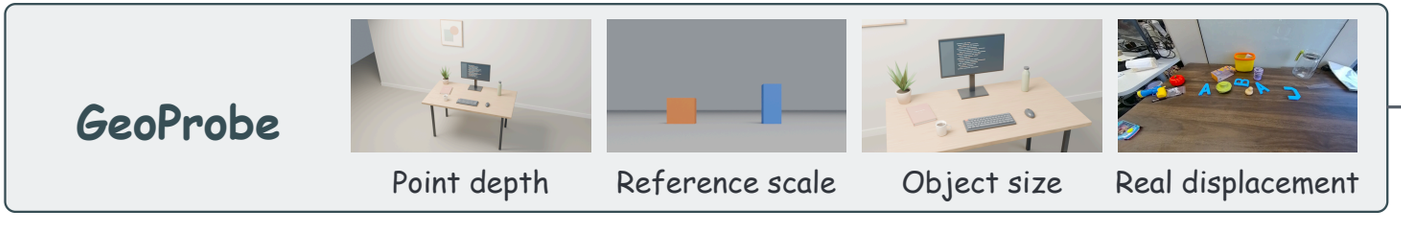

1,000
cases
37.0%
22.0%
7.0%
15.7%
18.3%
Geometry
Spatial Reasoning
Affordance
Task Planning
Manipulation
(b)

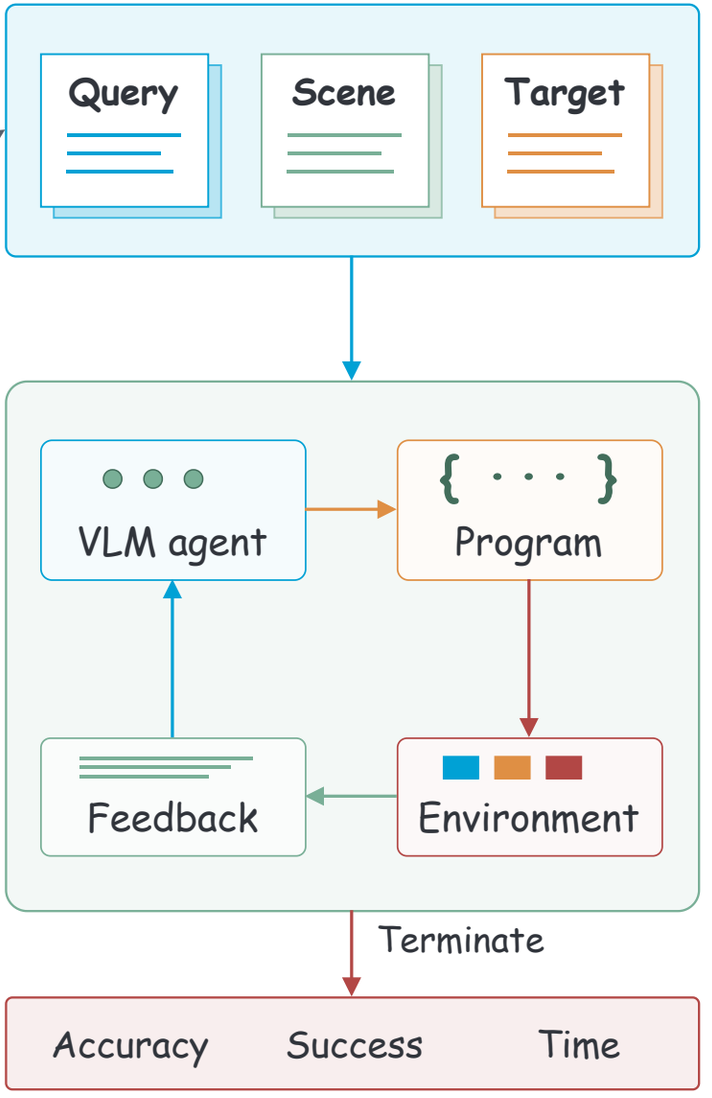

60
40
20
0
36
24
12
40
56
Calib.
Pose
Depth
Displ.
Scale
(c)

(d)


InFlux
Map-free
Multi-Spatial
TraceSpatial
3DSRBench
BOP-ASK
MindCube
MMSI-Bench
RoboSpatial
VSI-Bench
RAGNet
ReasonAff
UMD
ALFRED
ALFWorld
DiscoveryWorld
HumanCLAW
ScienceWorld
VirtualHome
BEHAVIOR-1K
CALVIN
robosuite
LIBERO-PRO
CLIPort
ManiSkill 3
RLBench
RoboCasa
RoboCasa365
RoboTwin 2.0
RoboWits
VIMA-Bench
VLABench

+


GeoProbe
Point depth
Reference scale
Object size
Real displacement


Query
Scene
Target
VLM agent
Program
Feedback
Environment
Terminate
Accuracy
Success
Time