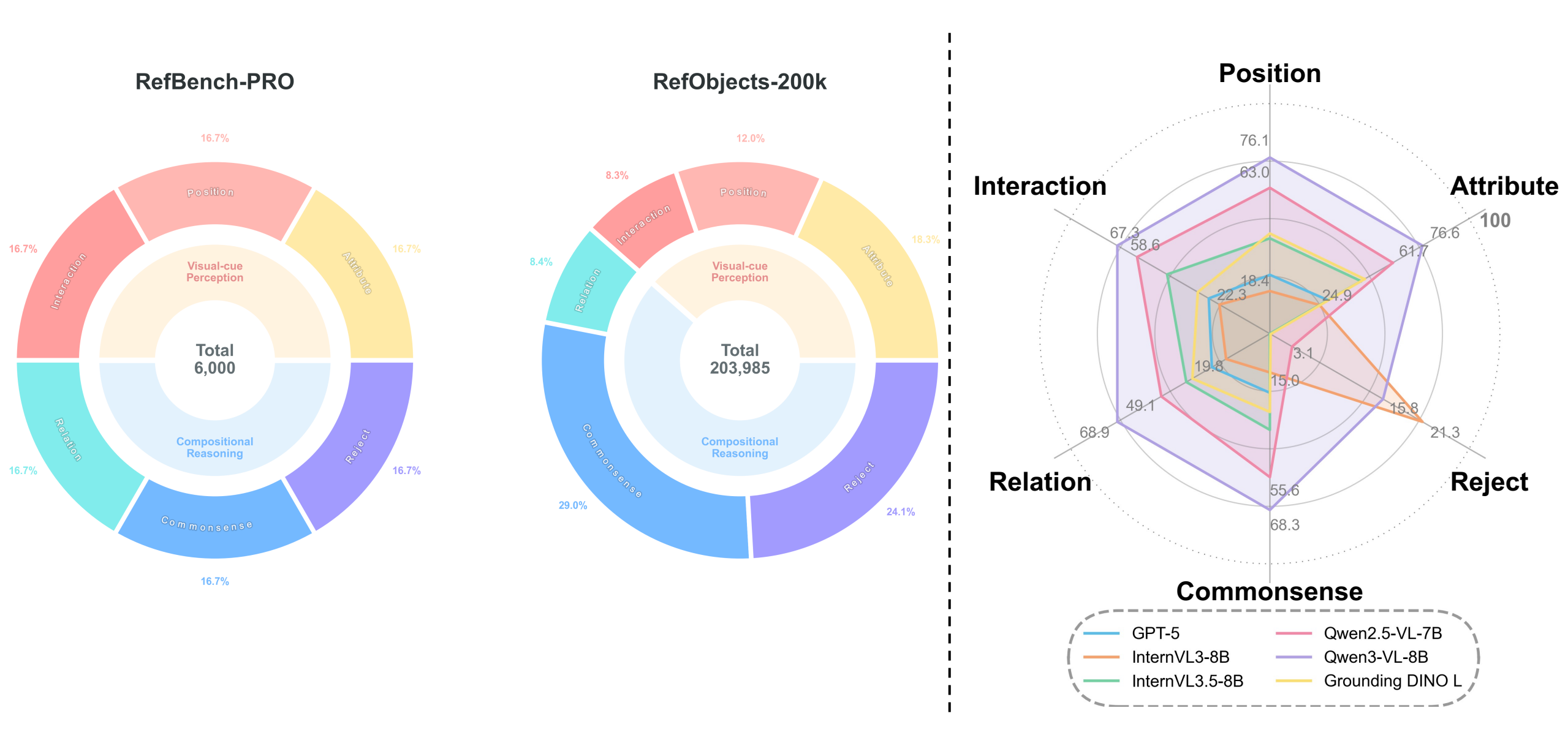
RefBench-PRO
16.7%
Position
16.7%
Attribute
16.7%
Interaction
Visual-cue Perception
Total 6,000
Compositional Reasoning
16.7%
Relation
16.7%
Commonsense
16.7%
Reject
RefObjects-200k
12.0%
Position
8.3%
Interaction
18.3%
Attribute
8.4%
Relation
Visual-cue Perception
Total 203,985
Compositional Reasoning
29.0%
Commonsense
24.1%
Reject
Position
Attribute
Reject
Commonsense
Relation
Interaction
76.1
63.0
76.6
61.7
24.9
18.4
22.3
3.1
15.0
15.8
21.3
55.6
68.3
49.1
68.9
19.8
67.3
58.6
100
GPT-5
InternVL3-8B
InternVL3.5-8B
Qwen2.5-VL-7B
Qwen3-VL-8B
Grounding DINO L