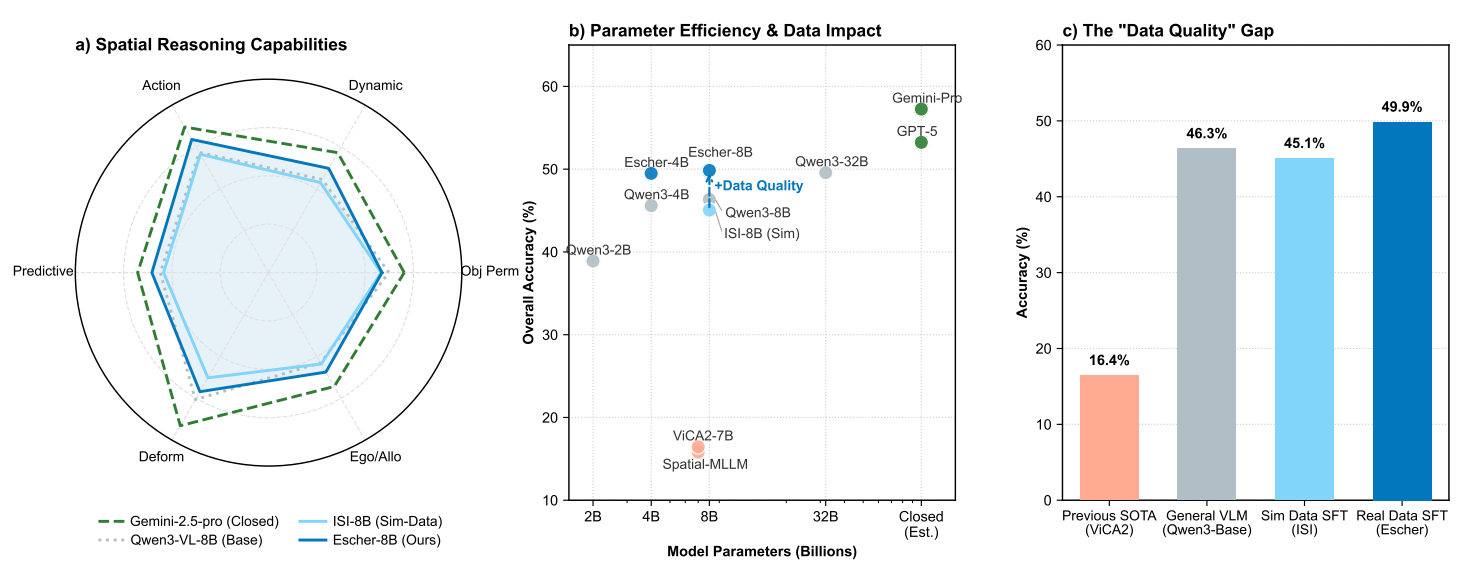

a) Spatial Reasoning Capabilities
Action
Dynamic
Obj Perm
Ego/Allo
Deform
Predictive
Gemini-2.5-pro (Closed)
Qwen3-VL-8B (Base)
ISI-8B (Sim-Data)
Escher-8B (Ours)
b) Parameter Efficiency & Data Impact
Overall Accuracy (%)
Model Parameters (Billions)
2B
4B
8B
32B
Closed (Est.)
Gemini-Pro
GPT-5
Escher-4B
Escher-8B
Qwen3-32B
+Data Quality
Qwen3-4B
Qwen3-8B
ISI-8B (Sim)
Qwen3-2B
ViCA2-7B
Spatial-MLLM
c) The "Data Quality" Gap
Accuracy (%)
16.4%
46.3%
45.1%
49.9%
Previous SOTA (ViCA2)
General VLM (Qwen3-Base)
Sim Data SFT (ISI)
Real Data SFT (Escher)