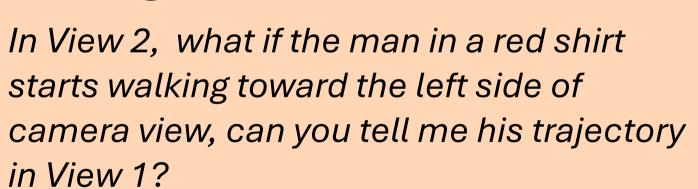**Can MLLMs identify the same attribute across views?**

Who stands closest to the spot where camera View 2 was positioned, and can we find the same person in camera View 1?

Can MLLMs track the object moving from another view?

In View 2, what if the man in a red shirt starts walking toward the left side of camera view, can you tell me his trajectory in View 1?

Can MLLMs estimate coarse camera poses?

If I randomly shuffle View 1 to View 4, can you sort them in clockwise order from top-down view?

View 3 · View 4 · View 1 · View 2

Counting · Camera Pose Estimation · Attribute Identification · Relative Distance · Relative Direction · Object Manipulation

Human Level · Gemini-2.0-Flash · Claude-3.7-Sonnet · GPT-4o · Qwen2.5-VL-72B · InternVL2.5-38B