a

Action & Intent-Driven Spatial Reasoning

TRUE/FALSE

After catching the ball, the boy's first step is taken in the opposite direction of the man in the red cap walking in the background.

VLM: True
Human: False
Case 01 | Original video: 05:29–05:42

Egocentric vs. Allocentric Reference Frames

SINGLE-CHOICE

The video pans from a wide city view to a viewing platform. From the perspective of the people sitting on the curved stone benches at the end, in which direction is the prominent golden-domed building?

VLM: C) To their far left.
Human: D) Completely behind them, out of sight.
Case 02 | Original video: 07:23–07:46

Predictive & Counterfactual Spatial Reasoning

TRUE/FALSE

If the bamboo pole were 3 feet shorter, it would still be long enough to lean securely against the corrugated metal roof as shown.

VLM: False
Human: True
Case 03 | Original video: 01:45–02:04

Dynamic Spatial Relationships

SINGLE-CHOICE

During which jump does the person land with both feet fully on the opposite side of the white line for the first time? [Options] A) The second jump. B) The third jump. C) The fourth jump. D) The fifth jump.

VLM: C) The fourth jump.
Human: B) The third jump.
Case 04 | Original video: 00:16–00:38

Object Deformation & State Transition

TRUE/FALSE

The fishing rod's bend is most pronounced when the man is actively turning the reel handle.

VLM: True
Human: False
Case 05 | Original video: 00:47–01:03

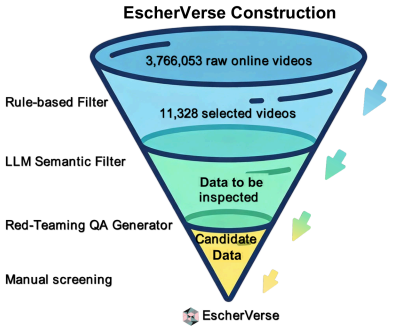
Object Permanence & Occlusion Tracking

FILL-IN-THE-BLANK

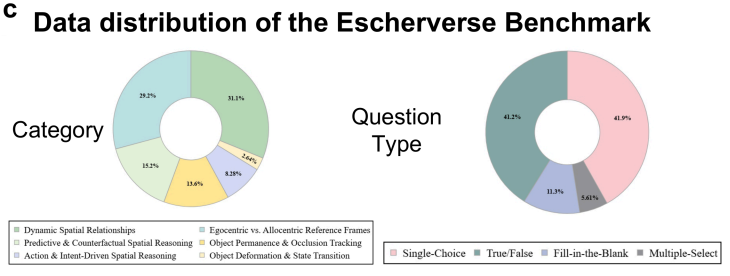
At 0:19, a small black puppy goes under the picnic table. It reappears on the boy's _____ side.

VLM: left
Human: right
Case 06 | Original video: 00:39–01:25

b



c



d

