Supplementary material for GRASP: A novel benchmark for evaluating language GRounding And Situated Physics understanding in multimodal language models

Contents

1 Level 1 Prompts

1.1 Binary Questions

Shape A . . . is on the table. Is this true? Answer only with yes or no.

Color The ball on the table is . . . Is this true? Answer only with yes or no.

Directionality The ball is rolling . . . . Is this true? Answer only with yes or no.

Movement The ball is . . . . Is this true? Answer only with yes or no.

Object Ordering From left to right, the following objects are on the table: *ob1*, *obj2*, . . . , *objN*. Is this true? Answer only with yes or no.

Relational Position The ball is on the . . . side of the black barrier. Is this true? Answer only with yes or no.

1.2 Open Questions

Shape What shape is the object on the table?

Color What color is the object on the table?

Directionality Which direction is the ball rolling?

Relational Position Which side of the black barrier is the ball on?

2 Level 2 Prompts

The tests are numbered following Section 5.

Test 1 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the outcome of the experiment plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 2 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the outcome of the experiment plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 3 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the final position of the ball plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 4 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the final position of the ball plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 5 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the location of the cube plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 6 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the trajectory of the ball plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 7 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the trajectory of the rotating plank plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 8 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the trajectory of the ball plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 9 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the trajectory of the ball plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 10 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the final position of the ball plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 11 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the trajectory of the ball plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 12 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the final position of the ball plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 13 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the final position of the ball plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 14 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the interaction between the balls plausible, assuming they are of the same mass? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 15 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the final position of the top cube plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

Test 16 The video you're seeing was generated by a simulator. Given how objects behave on earth, is the final position of the ball plausible? Your answer should be based on the events in the video and ignore the quality of the simulation. Answer only with yes or no.

3 Test-Specific Randomizations (Level 2)

The tests are numbered following Section 5.

Test 1

- Object positions
- Number of objects (1-3)

Test 3

- Horizontal position of both barriers

Test 4

- Gap size in the platform

Test 8

- Angle of deflection

Test 9

- Angle of the plank (10-30 degrees)

Test 11

- In the implausible case, we randomize whether the ball keeps floating or drops down in a straight line

Test 12

- Corner towards which the ball rolls to

Test 14

- Angle of the plank (10 - 40 degrees)

Test 16

- Corner towards which the ball rolls to

4 Model Parameters (Level 1 and 2)

Generation parameters for each model are left at their default values. Where necessary, the system prompts were modified from their defaults so as not to clash with the instruction to give one-word replies:

Video-ChatGPT "You are Video-ChatGPT, a large vision-language assistant. You are able to understand the video content that the user provides, and answer their questions correctly. Follow the instructions carefully and give short answers based on the provided video."

Video-LLaMA "You will be able to see a video once I provide it to you. Please answer my questions."

VTimeLLM "A chat between a curious user and an artificial intelligence assistant. The assistant gives helpful and polite answers to the user's questions."

5 Intuitive Physics Tests (Level 2)

Figures 1–4 show selected frames from example videos for each of our Level 2 tests. We show a plausible (top row) and implausible (bottom row) event for each test. Please refer to the scene descriptions in Section 6 for more details.

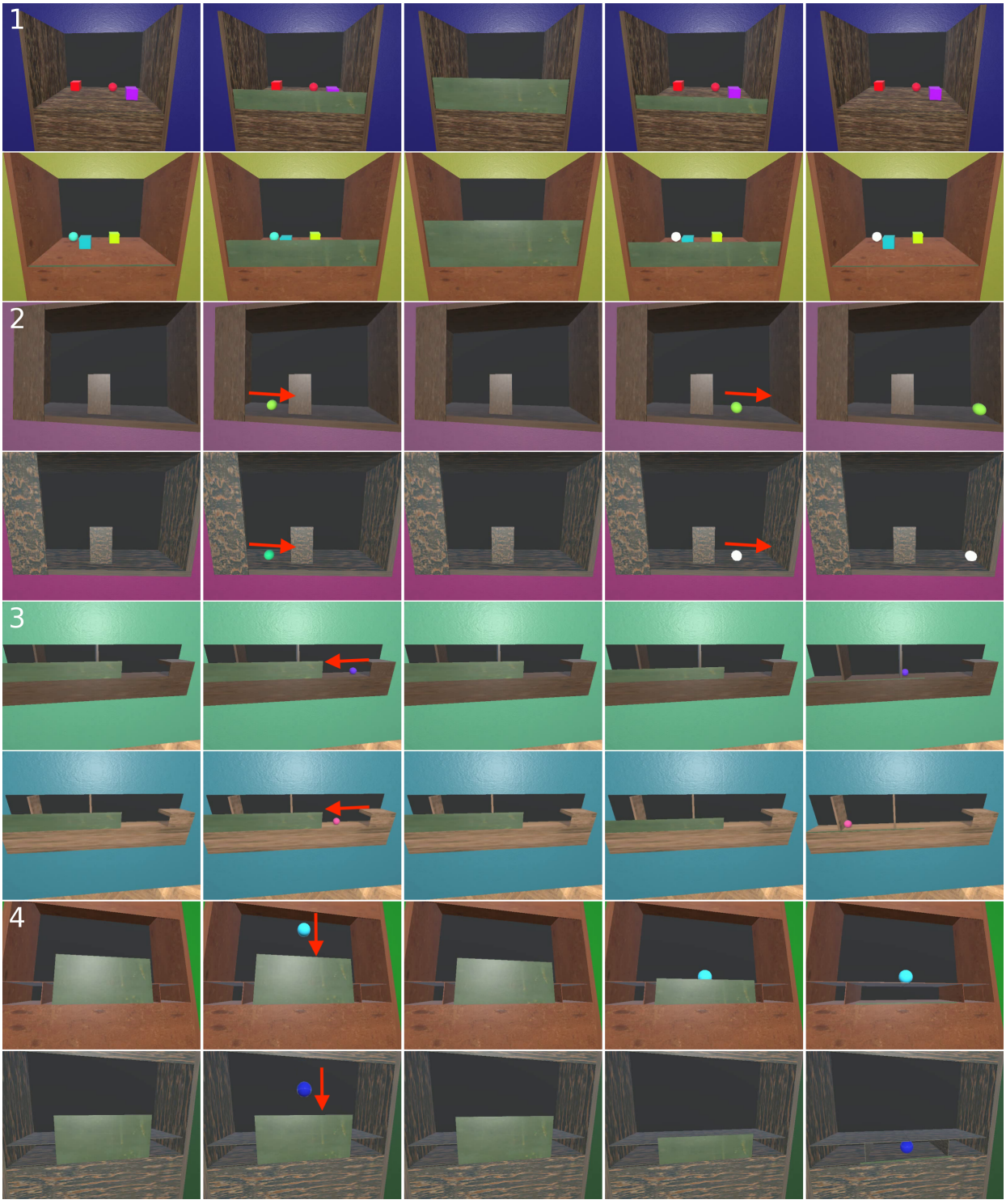

Figure 1: Level 2 (Intuitive Physics) tests: The first and second rows display plausible and implausible versions respectively for each experiment. The tests assess the understanding of (1) unchangeableness, (2) unchangeableness, (3) solidity & continuity, and (4) solidity & continuity. Red arrows indicate the movement of the ball.

Figure 2: Level 2 (Intuitive Physics) tests: The first and second rows display plausible and implausible versions respectively for each experiment. The tests assess the understanding of (5) object permanence, (6) object permanence, (7) object permanence, and (8) inertia. Red arrows indicate the movement of the ball. In (5), the two large blocks are moving from right to left until one of them covers the small block. Then the rod pushes out the small block from behind one of the large blocks.

Figure 3: Level 2 (Intuitive Physics) tests: The first and second rows display plausible and implausible versions respectively for each experiment. The tests assess the understanding of (9) gravity, (10) gravity & inertia, (11) gravity & inertia, and (12) continuity. In (9), the ball accelerates downwards in the plausible condition and upwards in the implausible condition.

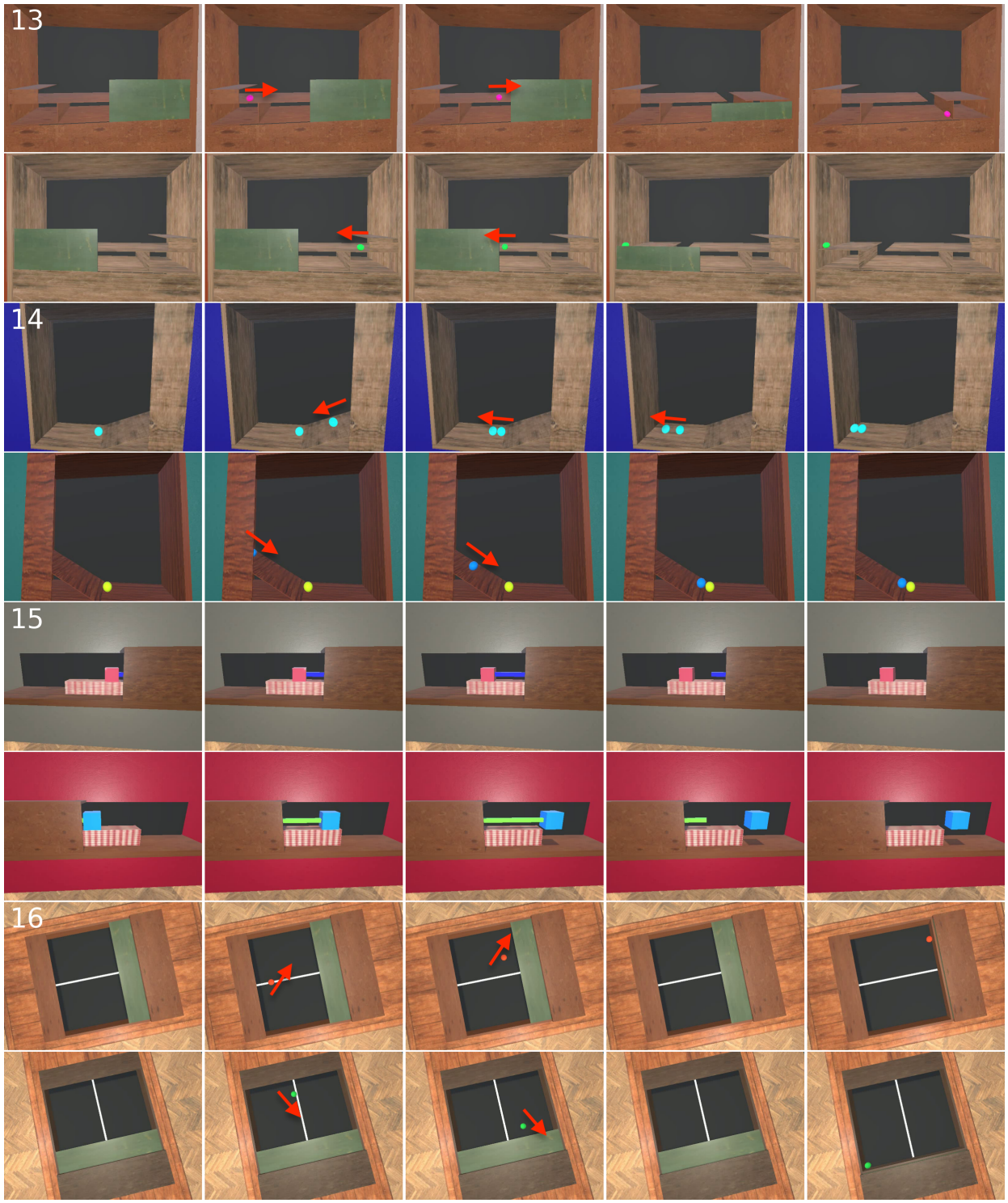

Figure 4: Level 2 (Intuitive Physics) tests: The first and second rows display plausible and implausible versions respectively for each experiment. The tests assess the understanding of (13) gravity & continuity, (14) collisions, (15) support & gravity, and (16) inertia. Red arrows indicate the movement of the ball(s). In (15), the yellow rod pushes the cube and then retracts.

6 Scene Descriptions (Level 2)

The tests are numbered following Section 5.

Test 1

- **Plausible**: Several objects sit on a flat surface at different positions. The objects have different shapes and colors. A mechanism moves an occluder such that the objects cannot be seen for a short while and then the occluder is removed again. All objects are still in the same position as before and have the same shape and color.
- **Implausible**: Several objects sit on a flat surface at different positions. The objects have different shapes and colors. A mechanism moves an occluder such that the objects cannot be seen for a short while and then the occluder is removed again. All objects are still in the same position as before but one has changed their shape or color.

Test 2

- **Plausible**: A sphere is rolling into the field of view. It rolls from one side to another, slowing down on the way. There is an occluding object, larger than the sphere, in the foreground. The sphere rolls past the occluding object. When the sphere rolls behind that object it disappears from the field of view and then reappears on the other side of the object. When the sphere reappears it looks exactly the same as before. The sphere continues to roll until it reaches a wall where it comes to rest.
- **Implausible**: A sphere is rolling into the field of view. It rolls from one side to another, slowing down on the way. There is an occluding object, larger than the sphere, in the foreground. The sphere rolls past the occluding object. When the sphere rolls behind that object it disappears from the field of view and then reappears on the other side of the object. When the sphere reappears it has a different color. The sphere continues to roll until it reaches a wall where it comes to rest.

Test 3

- **Plausible**: A sphere is rolling on a flat, solid plank. At the back of the plank is a wall. The sphere comes into the field of view on one side of the plank and rolls towards the other side. The occluder is standing at the frontal edge of the plank, parallel to the wall at the back. It is higher than the sphere but does not extend all the way to the ceiling. Two walls are visible above the occluder. The walls have a certain lateral distance from each other and extend all the way from the front to the back of the plank. The sphere rolls along the plank until it disappears behind the occluder. After a short while, the occluder is removed by a mechanism. The removal reveals the bottom parts of the two walls, which reach all the way to the plank at the bottom. The sphere has come to rest at the first wall along its trajectory.
- **Implausible**: A sphere is rolling on a flat, solid plank. At the back of the plank is a wall. The sphere comes into the field of view on one side of the plank and rolls towards the other side. The occluder is standing at the frontal edge of the plank, parallel to the wall at the back. It is higher than the sphere but does not extend all the way to the ceiling. Two walls are visible above the occluder. The walls have a certain lateral distance from each other and extend all the way from the front to the back of the plank. The sphere rolls along the plank until it disappears behind the occluder. After a short while, the occluder is removed by a mechanism. The removal reveals the bottom parts of the two walls, which reach all the way to the plank at the bottom. The sphere has come to rest at the second wall along its trajectory.

Test 4

- **Plausible**: There is a room enclosed by walls at the sides and at the back. The room has a lower level and an upper level, defined by a lower floor at the very bottom and an upper floor that is parallel to the lower floor but a certain height above it. Both floors are solid and extend all the way to the wall at the back and the walls at the side. An occluder blocks the field of view at the front of the room. It starts at the lower floor and extends over the upper floor. A sphere falls into the room from above and disappears behind the occluder. The part of the occluder that reaches over the upper floor is higher and wider than the diameter of the sphere. The occluder is removed by a mechanism. Both floors are completely visible. The sphere lies on the upper floor.
- **Implausible**: There is a room enclosed by walls at the sides and at the back. The room has a lower level and an upper level, defined by a lower floor at the very bottom and an upper floor that is parallel to the lower floor but a certain height above it. Both floors are solid and extend all the way to the wall at the back and the walls at the side. An occluder blocks the field of view at the front of the room. It starts at the lower floor and extends over the upper floor. A sphere falls into the room from above and disappears behind the occluder. The part of the occluder that reaches over the upper floor is higher and wider than the diameter of the sphere. The occluder is removed by a mechanism. Both floors are completely visible. The sphere lies on the lower floor.

Test 5

- **Plausible**: There is a mechanism moving two occluders in parallel. There is a little cube on the left side of the two occluders. The cube is smaller than the occluders. The occluders are moved sideways from right to left until the left occluder covers the cube. Another mechanism is used to move a pin. The pin is moved horizontally from left to right behind the occluders. It pushes out the cube from behind the left occluder to a position between the occluders. After pushing out the cube, the pin is retracted.

- **Implausible**: There is a mechanism moving two occluders in parallel. There is a little cube between the two occluders. The cube is smaller than the occluders. The occluders are moved sideways from right to left until the right occluder reaches and then covers the cube. Another mechanism is used to move a pin. The pin is moved horizontally from left to right behind the occluders. It pushes out the cube from behind the left occluder to a position between the two occluders. After pushing out the cube, the pin is retracted.

Test 6

- **Plausible**: A sphere is rolling from one side to another, passing behind an occluder on the way. In the middle of the surface is an occluder that is higher on the left and the right side and lower in the middle. At the low part in the middle, the height of the occluder exceeds the diameter of the sphere. The sphere disappears behind the occluder on one side and reappears on the other side. While it is behind the occluder, it remains entirely out of sight. The scene is viewed right from the front, with a line of sight exactly at the top of the lower part of the occluder.
- **Implausible**: A sphere is rolling from one side to another, passing behind an occluder on the way. In the middle of the surface is an occluder that is higher on the left and the right side and lower in the middle. At the low part in the middle, the height of the occluder is less than the diameter of the sphere. The sphere disappears behind the occluder on one side and reappears on the other side. While it is behind the occluder, it remains entirely out of sight. The scene is viewed right from the front, with a line of sight exactly at the top of the lower part of the occluder.

Test 7

- **Plausible**: A solid object is standing on a solid and flat surface. A certain mechanism is used to move a plank in a rotating motion. The plank starts out flat on the surface in front of the object. It is rotated upwards until it stands vertically on the surface and occludes the object, and rotated further until it is above the object. Its movement stops and then it is rotated back until it lies flat in its initial position again. While the plank is rotating back the object is revealed in its initial position.
- **Implausible**: A solid object is standing on a solid and flat surface. A certain mechanism is used to move a plank in a rotating motion. The plank starts out flat on the surface in front of the object. It is rotated upwards until it stands vertically on the surface, and rotated further until it lies flat on the surface again on the other side, where the object was standing before. Its movement stops and then it is rotated back until it lies flat in its initial position again. While the plank is rotating back the object is revealed in its initial position.

Test 8

- **Plausible**: A sphere is rolling onto a flat plane that is enclosed by a straight and rigid wall on the other side. There is a white line that runs through the middle of the plane. The sphere rolls onto the plane with a lateral offset from the line. It rolls diagonally towards the wall and reaches the wall exactly at the white line in the middle. When it reaches the wall it bounces off and rolls back on the other side of the white line. The angle between the sphere's trajectory and the white line is the same as on the way towards the wall. When the sphere disappears from view at the back of the plane, it is now on the opposite side of the white line compared to when it entered the plane, while the lateral offset is the same.
- **Implausible**: A sphere is rolling onto a flat plane that is enclosed by a straight and rigid wall on the other side. There is a white line that runs through the middle of the plane. The sphere rolls onto the plane with a lateral offset from the line. It rolls diagonally towards the wall and reaches the wall exactly at the white line in the middle. When it reaches the wall it bounces off and rolls back on the same side of the white line. The angle between the sphere's trajectory and the white line is the same as on the way towards the wall. When the sphere disappears from view at the back of the plane, it is still on the same side of the white line compared to when it entered the plane and the lateral offset is the same.

Test 9

- **Plausible**: There is a diagonally lying plank. A sphere rolls down the plank from one side to the other without slowing down.
- **Implausible**: There is a diagonally lying plank. A sphere rolls up the plank from one side to the other without slowing down.

Test 10

- **Plausible**: A sphere is falling down into the field of view. The sphere is falling towards a solid, flat surface. It falls until it disappears behind an occluder. When the occluder is removed, the sphere is lying behind it on the surface.
- **Implausible**: A sphere is falling down into the field of view. The sphere is falling towards a solid, flat surface. It falls until it disappears behind an occluder. When the occluder is removed, the sphere is hovering above the surface behind it.

Test 11

- **Plausible**: A sphere is rolling sideways on a flat solid surface until it reaches the edge of that surface. It drops down continuing to move in the same sideways direction while it is falling. When it hits the ground it continues rolling in that direction.
- **Implausible**: A sphere is rolling sideways on a flat solid surface until it reaches the edge of that surface. It drops down exactly vertically along the edge and stops moving upon reaching the ground.

Test 12

- **Plausible**: A sphere is rolling onto a flat plane from one side. The plane is enclosed by rigid walls to the left, right, and front of the sphere (in a U-shape). The end of the plane, where the sphere is rolling towards, is covered. The sphere rolls towards one of the corners, slowing down, until it disappears under the cover. The cover is removed automatically by an inbuilt mechanism. The removal of the cover reveals the frontal wall and the two corners. The sphere has come to rest in the corner it was rolling towards. The other corner is enclosed by an additional plank that runs diagonally from the wall at the front to the wall on the side, such that the walls and the plank form a triangle. There is no object inside the triangle.
- **Implausible**: A sphere is rolling onto a flat plane from one side. The plane is enclosed by rigid walls to the left, right, and front of the sphere (in a U-shape). The end of the plane, where the sphere is rolling towards, is covered. The sphere rolls towards one of the corners, slowing down, until it disappears under the cover. The cover is removed automatically by an inbuilt mechanism. The removal of the cover reveals the frontal wall and the two corners. The corner on the other side of the sphere is empty. The corner that the sphere was rolling towards is enclosed by an additional plank that runs diagonally from the wall at the front to the wall on the side, such that the walls and the plank form a triangle. The sphere has come to rest inside that triangle.

Test 13

- **Plausible**: A sphere is rolling slowly on a flat, solid plank. Below that plank is another plank that runs in parallel. The sphere comes into the field of view from one side and rolls towards the other side until it disappears behind an occluder. A mechanism removes the occluder. The removal reveals a gap in the upper plank on which the sphere was rolling. The gap is wider than the diameter of the sphere. The sphere is now on the lower plank and is at rest.
- **Implausible**: A sphere is rolling slowly on a flat, solid plank. Below that surface is another plank that runs in parallel. The sphere comes into the field of view from one side and rolls towards the other side until it disappears behind an occluder. A mechanism removes the occluder. The removal reveals a gap in the upper plank on which the sphere was rolling. The gap is wider than the diameter of the sphere. The sphere is now on the other side of the gap, still on the upper plank.

Test 14

- **Plausible**: A sphere is rolling down an incline. At the bottom of the incline, directly in the trajectory of the rolling sphere is another sphere of the same size and weight as the moving sphere. The rolling sphere is relatively fast. When it collides with the sphere at the bottom, it pushes that sphere away in the direction of movement. Both spheres continue to roll in that direction until they reach a wall. The first sphere (that pushed the second sphere) comes to rest right behind the sphere that it pushed. The sphere that was pushed comes to rest at the wall.
- **Implausible**: A sphere is rolling down an incline. At the bottom of the incline, directly in the trajectory of the rolling sphere is another sphere of the same size and weight as the moving sphere. The rolling sphere is relatively fast. When it collides with the sphere at the bottom, the sphere at the bottom remains at its exact original location. The sphere that was rolling comes to rest next to the sphere at the bottom.

Test 15

- **Plausible**: A box rests on an object that has a flat and solid surface. A pin is used to push the box sideways along that surface. Before the box reaches the edge of the surface the pin is retracted. The box rests at its new position. It has a different horizontal but the same vertical position as before.
- **Implausible**: A box rests on an object that has a flat and solid surface. A pin is used to push the box sideways along that surface. After the box has been pushed more than halfway over the edge of the surface, the pin is retracted. The box rests at its new position. It has a different horizontal but the same vertical position as before.

Test 16

- **Plausible**: A sphere is rolling onto a flat plane from one side. The plane is enclosed by rigid walls to the left, right, and front of the sphere (in a U-shape). The end of the plane, where the sphere is rolling towards, is covered. The sphere rolls towards one of the corners, slowing down, until it disappears under the cover. The cover is removed automatically by an inbuilt mechanism. The removal of the cover reveals that the sphere has come to rest in the corner it was rolling towards. The other corner is empty.
- **Implausible**: A sphere is rolling onto a flat plane from one side. The plane is enclosed by rigid walls to the left, right, and front of the sphere (in a U-shape). The end of the plane, where the sphere is rolling towards, is covered. The sphere rolls towards one of the corners, slowing down, until it disappears under the cover. The cover is removed automatically by an inbuilt mechanism. The removal of the cover reveals that the corner, which the sphere was rolling towards, is empty. The sphere has come to rest in the other corner.

7 Response Parsing (Level 1 and 2)

Responses to yes/no questions (Level 1 and 2) are only counted as valid if they begin with the word "yes" or "no"; the rest of the response is considered irrelevant. Responses that do not adhere to this are counted as incorrect.

Responses to open questions (Level 1) are parsed for certain keywords depending on the scenario. If a response does not contain any of the keywords, or contains multiple keywords that are associated with different answers, it is considered invalid. These are the keywords associated with each test:

Shape

- ball, sphere, round
- cube, box, rectangular

Color

- black
- blue
- green
- red

Directionality

- forward, forwards, up, upward, upwards
- backward, backwards, down, downward, downwards
- left, right to left, right to the left
- right, left to right, left to the right

Relational Position

- left
- right

8 Level 1 (Grounding) Results

Task	Test	Video-LLaMA (7B)		Video-LLaMA (13B)		Video-LLaMA2 (7B)		PandaGPT (7B)		PandaGPT (13B)		VTimeLLM (7B)		Video-ChatGPT (7B)	
		neg.	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.	pos.
Binary Classification Zero-Shot	Shape	1.3	96.9	15.6	82.6	0.0	96.6	0.0	100.0	0.0	100.0	0.0	100.0	2.1	97.7
	Color	0.0	96.9	13.0	82.6	3.1	96.6	0.0	100.0	0.0	100.0	0.0	100.0	0.5	97.7
	Movement	0.0	99.2	9.4	83.9	0.3	99.0	4.7	65.6	0.0	100.0	0.0	100.0	47.9	52.9
	Directionality	0.0	97.4	8.9	82.8	0.0	99.7	6.2	88.3	0.0	100.0	0.0	100.0	7.0	92.2
	Relational Position	0.3	100.0	9.1	86.7	0.3	97.7	0.0	100.0	0.0	100.0	0.0	100.0	2.3	97.1
	Ordering (avg.)	0.4	98.7	4.2	95.1	0.0	99.9	0.0	100.0	0.0	100.0	0.0	100.0	74.0	26.5
Binary Classification CoT	Shape	65.4	73.4	43.8	95.3	33.9	93.8	54.7	12.5	100.0	61.7	0.0	100.0	59.1	81.0
	Color	71.1	73.4	49.2	95.3	74.5	93.8	65.6	12.5	68.0	61.7	0.0	100.0	31.5	81.0
	Movement	23.4	74.2	9.9	87.8	4.2	92.4	25.0	63.3	37.5	34.4	0.0	100.0	1.3	84.9
	Directionality	30.7	59.9	29.9	65.1	34.4	65.9	50.8	51.6	77.3	25.0	0.0	100.0	26.3	65.9
	Relational Position	44.5	49.2	29.4	68.0	47.7	55.2	0.0	100.0	50.0	50.0	0.0	100.0	47.1	47.7
	Ordering (avg.)	25.7	71.4	3.6	94.4	0.4	99.6	1.3	98.4	70.8	30.4	0.0	100.0	1.6	99.6
Binary Classification One-Shot	Shape	32.0	50.8	21.9	25.3	39.1	52.1	85.9	15.1	80.2	12.0	0.0	83.9	51.3	45.6
	Color	45.1	50.8	24.5	25.3	70.8	52.1	49.5	15.1	36.5	12.0	0.0	83.9	84.6	45.6
	Movement	40.6	33.6	18.2	34.4	48.2	27.9	54.7	1.3	37.0	41.1	0.0	75.0	72.4	29.2
	Directionality	37.0	36.2	14.3	25.8	69.0	16.7	64.6	10.2	49.0	36.7	0.0	66.9	57.6	43.8
	Relational Position	35.7	44.8	18.8	23.2	52.6	22.1	65.4	41.4	44.8	26.0	0.0	51.3	74.2	28.6
	Ordering (avg.)	34.5	50.0	16.9	27.7	12.0	32.3	60.7	40.5	90.5	10.5	0.0	41.0	89.0	10.4

Table 1: Accuracy (%) for all models on GRASP's Level 1 using binary question prompts (inducing binary classification) with zero-shot, one-shot, and chain-of-thought (CoT) prompting strategies. Results are displayed for positive and negative samples. Correct answers are "yes" for pos. samples and "no" for neg. samples.