

# **1. "Create a realistic image of a seed germinating in a pot."**

Testing:

- Basic growth stage of a plant and simple biological stages
- Spatial understanding

Expected:

- A small seed with small root partially covered in soil growing downwards
- Root and seed should be correctly positioned
- Root and seed should be connected
- Not expected a sprout or much upper growth



Dalle3, 14.2.2025

- Seed still somewhat intact (Expected)
- Seed outside of the pot (Not expected but can be accepted)
- Sprout and root look as of early growth stage, however, the sprout looks much later than just newly sprouted (Not expected)
- Multiple small roots (Not expected, germinating is early stage so seed should be sprouting small root or few weak roots)



Gemini Flash 2.5, 31.07.2025

- A small sprout from moist soil in a pot (adheres to the prompt)
- Seed covering still visible
- Small water droplets (unrealistic placements as all are perfect rounds)
- Somewhat unrealistic is the perfect posture of the sprout



Stable Diffusion 3 Medium, 25.07.2025

- A small plant with very few leaves in a pot of soil
- Unrealistic is the mud/background around the pot
- Doesn't show "seed germinating"

## 2. “Create a realistic image of a seed that has just germinated in a pot.”

Testing:

- Basic growth stage of a plant and simple biological stages
- Spatial understanding

Expected:

- A small seed partially covered in soil, with a small shoot/sprout coming out from the soil.
- Seed outer covering may no longer be intact.
- Root, sprout and seed should be correctly positioned, whatever visible
- Root, sprout and seed all be connected, whatever visible



GPT-4o, 03.7.2025

- Seed still somewhat intact
- Seed outside of the pot (Not expected)
- Sprout looks as expected of early growth stage
- Seed and sprout connected (not disjoint)
- Seed coming as if fruit of small plant (Not expected, seed should be connected below where the sprout starts)



Gemini Flash 2.5, 31.07.2025

- An early-stage germinating plant on top of moist soil in a pot (adheres to the prompt)
- Seed covering still visible
- Small water droplets (unrealistic placements as all are perfectly spherical)
- Somewhat unrealistic is the perfect posture of the sprout
- Unrealistic is the frost (white in colour) around the pot



Stable Diffusion 3 Medium, 25.07.2025

- A small plant with very a few more leaves in a pot of soil
- Stem base also look strong
- All leaves are horizontal to ground
- Doesn't show "seed just germinated", i.e., no seed at all visible

**3. “Create a realistic image of a seed germinating in a pot, with the seed completely immersed in the soil and only a small shoot emerging above the surface.”**

Testing:

- Basic growth stage of a plant and simple biological stages
- Spatial understanding and occlusion

Expected: a small sprout coming out of the soil, no root or seed visible.



GPT-4o, 03.7.2025

- A small green element coming out of soil is visible but not able to make out if it's a sprout



Gemini Flash 2.5, 31.07.2025

- An early-stage sprout on top of moist soil in a pot (adheres to the prompt)
- Seed covering still visible
- Sprout is very new
- Seed is somewhat immersed in the soil, not completely



Stable Diffusion 3 Medium, 25.07.2025

- A small plant with very a few leaves in a pot of soil
- Pot looks unrealistic (not regular pot)
- The seed and the sprout are unrealistic (it's not a sprout, but a plant)
- Stem and leaf are disjoint

#### 4. "Give me a picture of a cup of hot tea turned upside down."

Testing:

- Physical plausibility
- Gravity
- Liquidity
- Positioning

Expected: A cup turned upside down completely with tea flowing from the cup, downwards



Dalle3, 14.2.2025



GPT-4o, 03.7.2025

- Some of the generated images show tea somehow outside the cup, with the cups still intact
- Movement of tea unnatural in some of the generated images
- Earlier generations show tea flowing in from somewhere and tea spilling (in motion) but not from cup
- Earlier generations show cups unnaturally positioned (slanting and off the ground)
- Latest generations show cups all positioned horizontally on saucers
- One of the later images (3<sup>rd</sup>) show something unnatural inside the cup that has the tea flowing out





Gemini Flash 2.5, 31.07.2025

- An upside-down cup is rendered with perfectly round tea/coffee/water stains on a wooden surface
- The stains on the cup and around the cup are unrealistic (bottom top)
- The steam coming from the middle of the cup is unrealistic (closed cup)
- No flowing of tea from the cup which is unrealistic for a cup of tea (maybe no tea in the cup? – Didn't adhere to the prompt)



Stable Diffusion 3 Medium, 25.07.2025

- Cup with some tea, no steam to show it might be hot
- Cup is not upside-down
- Cup is broken unrealistically



**5. “Create a realistic image of a cup of hot tea, where the cup is flipped upside down on top of the saucer.”**

Testing:

- Physical plausibility
- Gravity
- Liquidity and containment
- Positioning

Expected:

- A cup turned upside down completely on top of saucer
- Can expect that the cup no longer has any tea as it can't contain it anymore
- Can expect overflowing of tea and pouring/spilling out from the saucer to the surface



GPT-4o, 03.7.2025

- Correctly generated that cup is turned upside down
- Features a saucer
- Some of the tea is inside the saucer, while some outside the saucer
- No mark or signs of overflowing from the saucer so no plausibility for the tea outside the saucer to have spilled outside because of the change in position of the cup



Gemini Flash 2.5, 31.07.2025

- An upside-down cup is rendered on top of a saucer with some steam coming out, kept on a wooden surface
- The tea is pouring out from the cup, but it is unrealistic given the angle of the cup
- The steam originating from the middle of the cup is also unrealistic
- The resting angle of the cup is unrealistic as it looks like one edge of the cup is inside the saucer



Stable Diffusion 3 Medium, 25.07.2025

- One cup filled with tea on a saucer on a wooden surface. Another cup behind it, partially turned to one side (Didn't adhere to the prompt)

**6. "Show a realistic image of five chairs stacked together."**

Testing:

- Counting
- Alignment of objects
- Physical plausibility

Expected: Five similar in shape and size chairs, stacked neatly on top of one another vertically



Dalle3, 14.2.2025; GPT-4o, 08.7.2025

- Earlier image shows incorrect number of chairs while later image shows 4 seats, 3 back rests, 3 legs visible behind (visible from LHS) and 4 legs visible on each side in the front
- Earlier image some of the chairs are blurred
- Later image shows neatly stacked vertically
- Earlier image is kept randomly, can't make out the legs and seats (difficult to say that it will be a physically balanced structure/placement)



Gemini Flash 2.5, 31.07.2025

- Five chairs, all different types/colours on top of each other
- Legs of the chair at the bottom is unrealistic in terms of balance
- 2<sup>nd</sup> and 3<sup>rd</sup> as well as 3<sup>rd</sup> and 4<sup>th</sup> chairs seem to be intersecting (one inside another)
- There's an extra seat between the legs of the 2<sup>nd</sup> chair
- There's an extra rod/leg below the front legs of the 2<sup>nd</sup> chair resting on top of the bottom chair
- This arrangement may not be balanced



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic rendering- no distinct 5 chairs, no proper stacking behavior

## 7. “Create a realistic image of five chairs stacked on top of one another.”

Testing:

- Counting
- Alignment of objects and stacking
- Physical plausibility of balance

Expected: Five similar in shape and size chairs, stacked neatly on top of one another vertically



GPT-4o, 08.7.2025

- Correctly generated 5 seats, 5 back rests, 5 legs on both sides in front is visible
- Only 4 legs visible in back LHS and 2 visible in RHS
- Only 4 connectors to seat and back rest visible from LHS
- Rods at bottom connecting legs doesn't seem connected
- Neatly stacked vertically and a balanced structure



Gemini Flash 2.5, 31.07.2025

- Five chairs, all similar, on top of each other
- This arrangement is not balanced
- The legs, seat and backrest are disjoint which doesn't help in counting



Gemini Flash 2.5, 31.07.2025

- Prompt: "Create a realistic image of five similar chairs stacked neatly on top of one another."
- All similar and neatly stacked
- Number of handrests and seats is 5; Number of backrests is 6; Number of legs is 4 on each side in front and 3 in the back (only one side visible)



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic rendering- no distinct 5 chairs, no proper stacking behavior



**8. “Create an image where one person is standing behind another, but both faces are fully visible from the front.”**

Testing:

- Spatial understanding
- Occlusion
- Plausibility

Expected:

- Two people, one behind the other
- Both faces fully visible from the front view
- The rear one standing at either side of the front one, with a little distance, if same heights



Dalle3, 14.2.2025; GPT-4o, 03.7.2025

- Earlier image shows maybe body on one person but somewhat different facial features on either half. One half of the face shows features of a man (beard, shorter hair, slightly bigger nose) and the other half shows features of a woman (longer hair, make up/longer lashes, earrings, thinner and trimmed eyebrows)
- Earlier image only physically plausible if it's a make-up, otherwise a very natural image
- Later image is expected result, faces are fully visible as taller one is behind



Gemini Flash 2.5, 31.07.2025

- Two women either of different heights or one sitting while other standing behind
- Both faces are visible (2<sup>nd</sup> woman not fully) from front
- Followed the prompt correctly



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic rendering- neither face fully visible
- Unrealistic as the faces seem to be of different sizes/scales and two bodies blended into each other
- Neither is standing behind another

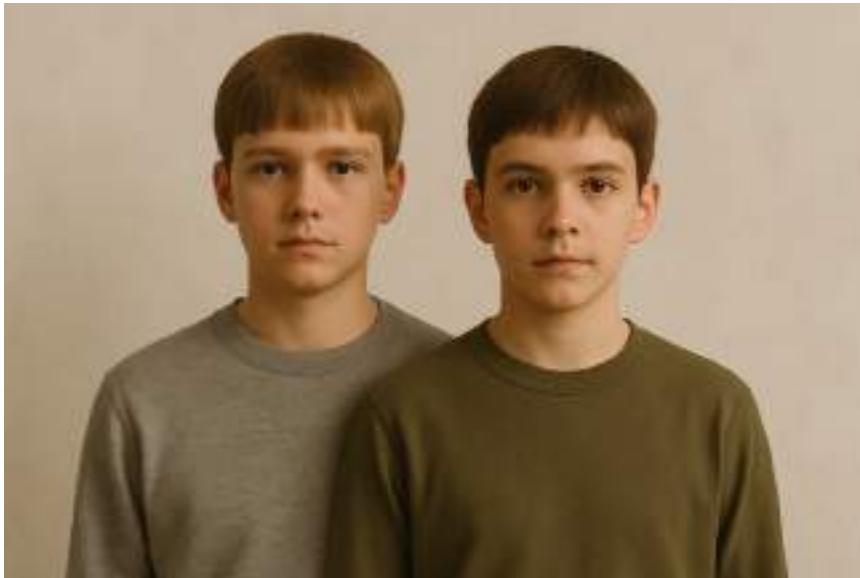
**9. “Create a realistic image where one boy is standing behind another, both same heights, and both faces are fully visible from the front. Faces can be generic.”**

Testing:

- Spatial understanding
- Occlusion
- Plausibility

Expected:

- Two people, one behind the other
- Both faces fully visible from the front view
- The rear one standing at either side of the front one, with a little distance
- Both of same heights



GPT-4o, 03.7.2025 - Generated image is as expected



Gemini Flash 2.5, 31.07.2025

- Two boys either of different heights or one sitting while other standing behind
- Both faces fully visible from front



Stable Diffusion 3 Medium, 25.07.2025

- Adhered to the prompt – Same heights, One standing behind another, Faces fully visible, Two boys

# 10. "Create a realistic image of two mirrors facing each other."

Testing:

- Spatial understanding
- Reflections
- Plausibility
- Shadows

Expected:

- Two similar mirrors, places next to each other at an angle, facing each other
- Mirrors should reflect partial surrounding and the other mirror



Dalle3, 14.2.2025; GPT-4o, 03.7.2025

- Both images show two mirrors placed at an angle, facing each other. This is as expected.
- Earlier image, background of the mirrors is not visible, the mirrors are not reflecting anything
- Later image, background is visible (wall with a wallpaper and wooden floor); the mirrors are reflecting just the background
- Later image, the mirrors are not reflecting each other; they're also showing somewhat a mirrored image of the reflection in the other
- No shadow in the later image



Gemini Flash 2.5, 31.07.2025

- Three mirrors probably side by side- 2 on one wall, the other on the other wall
- Viewed very close to the third mirror
- Reflections aren't accurate but otherwise the whole image looks very realistic
- No mirrors facing each other, or can't make it out if there's a mirror facing each other
- (reasoning it gave that if two mirrors are facing other then there would be infinite reflections)



Stable Diffusion 3 Medium, 25.07.2025

- Two mirrors on the wall side-by-side
- Reflections might be realistic on one mirror, can't make out the other
- However, the mirrors are not facing each other as instructed in the prompt

**11. “Create a realistic image of two mirrors facing each other, showing correct reflections.”**

Testing:

- Spatial understanding
- Reflections
- Plausibility
- Shadows

Expected:

- Two similar mirrors, placed next to each other at an angle, facing each other
- Mirrors should reflect partial surrounding and multiple diminishing reflections of each mirror in the other (recursive reflections)



GPT-4o, 03.7.2025

- Background is visible (wall with a wallpaper and wooden floor). The mirrors are reflecting the background partially.
- Mirrors seem to be reflecting each other. They're also showing somewhat a mirrored image of the reflection in the other.
- The reflections show the other mirror with inconsistencies (the support behind the frame is visible, the foot support of the frame is not visible, the back support seems to be also distorted as it leaning forward). [Physically impossible]





Gemini Flash 2.5, 31.07.2025

- Two mirrors side by side on the other wall
- Can't make out if there's 2 mirrors facing each other
- Reflections aren't accurate, mirrors in the reflection are different from the ones on the wall
- No mirrors facing each other, or can't make it out if there's a mirror facing each other
- It does try to show infinite reflections



Stable Diffusion 3 Medium, 25.07.2025

- Two mirrors, side-by-side, not of rectangle shape or hanged at an angle on the wall
- Reflections are not realistic
- Mirrors are not facing each other as expected out from the prompt

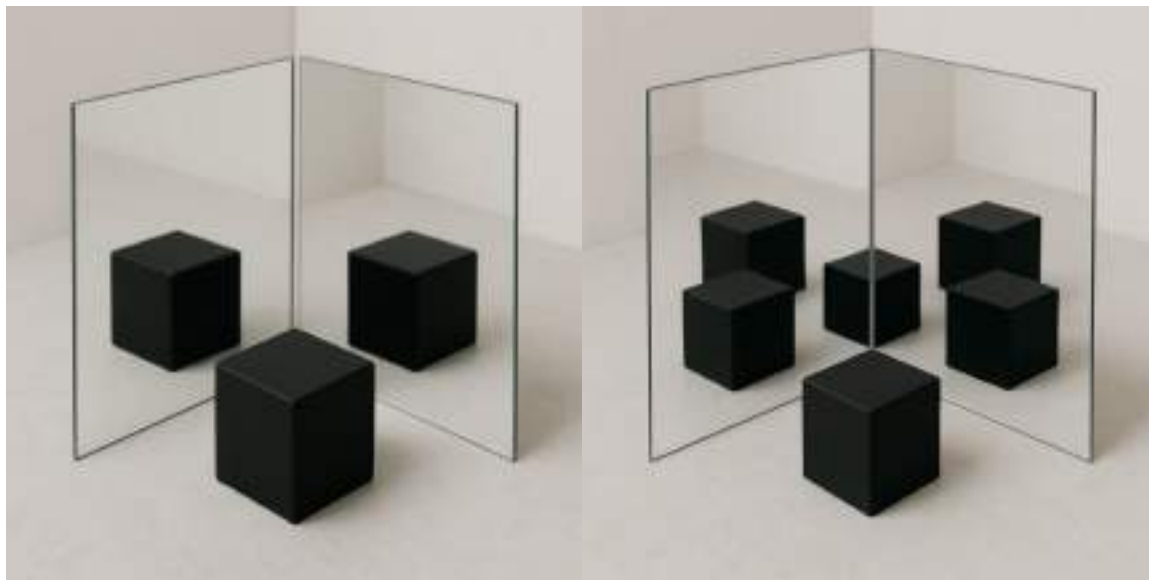
**12. “Create a realistic image of two standalone, borderless plane mirrors positioned at a 90-degree angle to each other, with a tiny black box placed between them.”**

Testing:

- Complex Spatial reasoning and angles
- Multiple reflections
- Plausibility

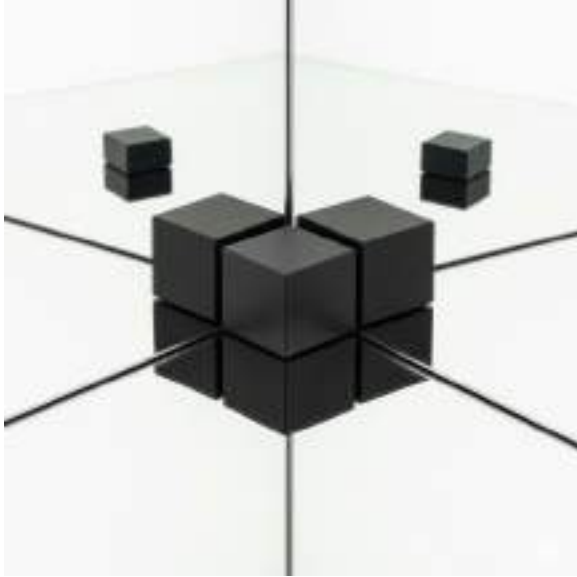
Expected:

- Two similar, plane, borderless, standalone mirrors, placed next to each other at a 90-degree angle, facing each other
- Small black box placed in between the two mirrors
- Mirrors should reflect the box, surroundings and each other mirror
- Multiple reflections expected ([Reference](#))



GPT-4o, 03.7.2025

- Mirrors and object placements are as expected
- First image shows a reflection of the black box, without accounting for the other mirror's placement
- Reflections of the background can't be evaluated in both the pictures
- Second image shows 5 views of the object which can be somewhat expected based on the angle of view
- Second image still however doesn't reflect the other mirror's boundary with the floor (we can expect there to be some presence even if there's no frame) so the views of the box can't be reasoned



Gemini Flash 2.5, 31.07.2025

- Mirrors side by side and the black box between them so positioning is right
- There might be more than 3 mirrors
- Reflections aren't accurate or can't really say about the black rectangle box shown in the reflections to be bit further away



Stable Diffusion 3 Medium, 25.07.2025

- Two mirrors, side-by-side, not of rectangle shape, not standalone as well with a black box further away from the mirrors
- No proper reflections (black)
- Mirrors are not facing each other as expected out from the prompt

**13. “Create a realistic image of the situation. I was drinking a glass of water and I heard a shocking news. The glass fell from my hands.”**

Testing:

- Cause and Effect understanding of dynamic events
- Realism of the motion
- Realism of Human reaction

Expected:

- A glass falling from a person’s hand onto a surface, if in motion otherwise shattered glass with spilled water on a surface
- Person in surprise/shock in the background



Dalle3, 14.2.2025

- Perfectly resting glass on a table’s surface
- Water flying upwards with a wave motion of water in the glass (only possible if the glass was kept down with a force)
- Right hand with palms showing a weird placement between the jumping water and glass
- Right hand’s fingers weirdly spaced, with impossible length (all five of similar lengths)
- Human reaction in background shown with a wide-open mouth, expected to depict shock



Gemini Flash 2.5, 31.07.2025

- Glass of water mid air with water splash all over, TV in background with very bad spelling for “SHOCKING NEWS”, hand below the glass
- Hand position doesn’t show for “hand fell from my hands”
- Water flowing from the bottom of the glass is unrealistic
- In the screen it’s written “shocking news”, however, prompt said “I heard” not “read” or “saw”



Stable Diffusion 3 Medium, 25.07.2025

- Glass with water on a wooden surface, glass has two top surfaces, hand is rendered incorrectly and unrealistically
- Nothing to indicate the situation described in the prompt

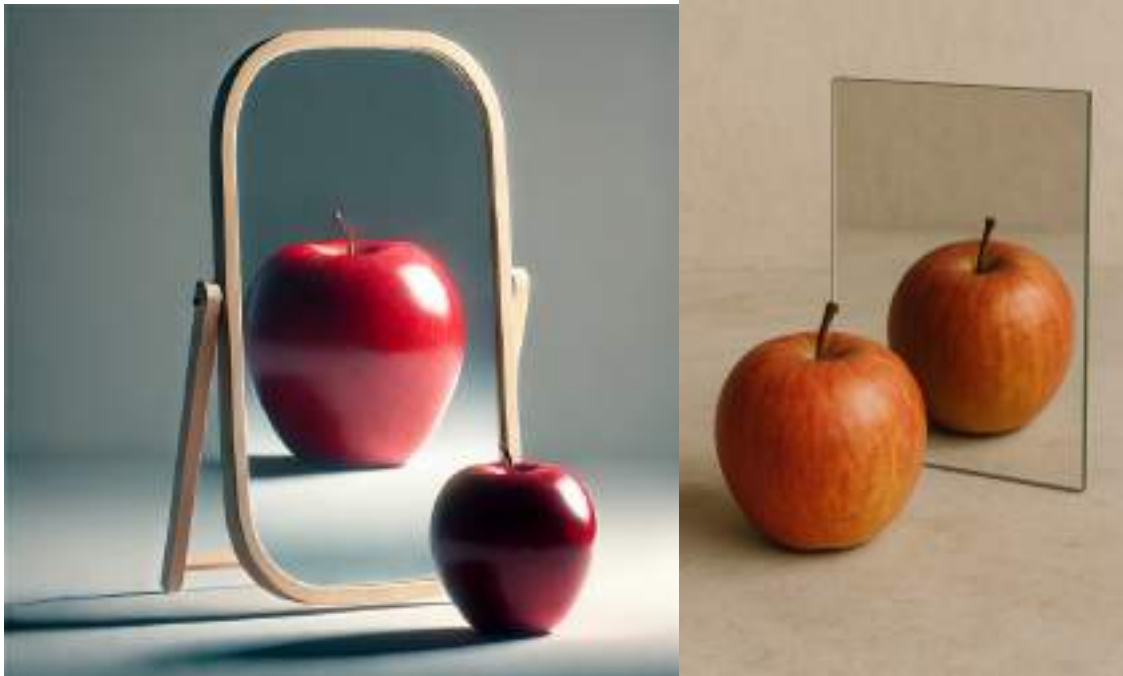
#### 14. “Create a realistic image of an apple behind a mirror.”

Testing:

- Occlusion
- Reflection

Expected:

- A red apple in front of a mirror
- An accurate reflection with apples of same size



Dalle3, 17.2.2025; GPT-4o, 03.7.2025

- Both images show a red apple in front of a mirror
- Earlier image show reflection as to same apple but bigger as if it's an image not a reflection (see the light hitting on apple to show exactly on the mirror)
- Later image shows a better reflection with same of size, however the light spot seems to be similarly placed here (towards the front half of the apple)
- If the light spot is front of the apple, the reflection shouldn't be showing it as the mirror sitting behind the apple wouldn't know of it



Gemini Flash 2.5, 31.07.2025

- A bright red apple is shown on a place mirror
- However, no apple in front of mirror to show that it's a reflection instead of a view from glass



Stable Diffusion 3 Medium, 25.07.2025

- Apple shown in a mirror but nothing indicative of it being reflection instead of a view from glass
- Mirror at the edge of table, however, apple in the view is away from the table edge (so not a reflection)



**15. “Create a realistic image of water boiling.”**

Testing:

- States of matter with phase change

Expected:

- A pot with water with slight bubbles or fully bubbling
- Steam from the pot



Dalle3, 17.2.2025

- Steam from the pot is shown
- Bubbles of water overflowing and moving outside the pot is not expected



Gemini Flash 2.5, 31.07.2025

- A glass vessel full of water on a burning stovetop, with water boiling, bubbles in the water (various sizes) and steam on top
- Very realistic rendering



Stable Diffusion 3 Medium, 25.07.2025

- Steel vessel with some steam on top, bubbles in the vessel
- Color of the content looks like milk (white) rather than water
- The soldering on the vessel is not against the handle which is again unrealistic

**16. “Create a realistic image of a big trophy fitting into a small suitcase.”**

Testing:

- Comparative sizing and proportion understanding
- Realistic containment

Expected:

- A big trophy and a small suitcase that's open
- Trophy bigger than the suitcase
- Trophy not fitting the suitcase (lengthwise or breadthwise)



Dalle3, 17.2.2025

- A big trophy in a suitcase that's smaller in size comparatively
- Trophy is bigger
- Trophy is not fitting
- However, the image also shows some's hand. The holding of trophy and fingers is incorrect
- Handles on both sides of trophy seems to be inconsistent with each other



Gemini Flash 2.5, 31.07.2025

- A big trophy and a smaller suitcase are rendered
- Angle of the suitcase and how its standalone is unrealistic
- Hand on top of the suitcase indicate the action “fitting”



Stable Diffusion 3 Medium, 25.07.2025

- A big trophy on top of a small case
- Unrealistic- how the trophy is floating, nothing showing to see what’s between the trophy and the case to hold the weight (the paper can’t hold it)
- Nothing indicating the action of “fitting”

**17. “Can you show me a realistic image of a big trophy struggling to fit into a small trunk?”**

Testing:

- Comparative sizing and proportion understanding
- Realistic containment

Expected:

- A big trophy and an open trunk of a car
- Trophy bigger than the trunk of the car (length/breadth)
- Trophy not fitting the trunk



Dalle3, 17.2.2025

- A big trophy and an open trunk that's smaller in size comparatively generated
- Trophy is bigger
- Trophy is not fitting
- Trophy handle on the bottom shows two fingers (randomly generated, away from the hand, position where normal trophies are held)
- Trophy handle on the top seems to have blended with the hand of the person inside the trunk (fingers also weird, incorrect numbers)
- The feet are pointing to opposite direction that the body is facing



Gemini Flash 2.5, 31.07.2025

- A big trophy and a smaller trunk/chest are rendered
- Hand on top of the suitcase indicate the action “struggling to fit”
- Expected car trunk as previous prompt rendered suitcase given it’s a similar prompt



Stable Diffusion 3 Medium, 25.07.2025

- A trophy on top of a chest box, with a dinosaur like gold body touching the trophy
- Trophy is leaning towards one side and it’s rendering is unrealistic as the base looks like it’s melting and there’s no symmetry to the trophy
- Unrealistic- how the trophy is floating, nothing showing to see what’s between the trophy and the case to hold the weight (the paper can’t hold it)
- Nothing indicating the action of “struggling to fit”

**18. “Can you show me a realistic image of a vacuum cleaner cleaning the room?”**

Testing:

- Correct depiction and use of the object
- Realistic interaction vacuum cleaner and the floor
- Avoidance of floating or disconnected objects

Expected:

- A fully connected and realistic vacuum cleaner
- A room in the background and a person operating the vacuum cleaner
- Vacuum cleaner in operation, cleaning while the cord connected to a plug point



Dalle3, 17.2.2025

- Disconnected parts and incomplete vacuum cleaner
- No cord visible (even though it looks like a traditional vacuum cleaner)
- Room background correctly generated however with a lamp that is balanced just on a thin pole (no base)
- Shadows of the cleaner is incorrect





Gemini Flash 2.5, 31.07.2025

- A very realistic looking vacuum cleaner in action in a room
- Flying stones and dust to indicate the action “cleaning”
- The stones however are unrealistic in a room setting, especially of this size
- The shadows are inconsistent and unrealistic
- The leg of the table in the back is floating



Stable Diffusion 3 Medium, 25.07.2025

- A traditional vacuum cleaner with the pipes broken, inconsistent and nothing indicative of the action
- No one's holding or supporting the pipe (unrealistic)

**19. “Can you show me a realistic image of a vacuum cleaner?”**

Testing:

- Correct depiction of the object
- Avoidance of floating or disconnected objects

Expected:

- A fully connected and realistic vacuum cleaner
- A room in the background



Dalle3, 17.2.2025

- Disconnected parts and incomplete vacuum cleaner (the cleaner brush connected to the engine and the tube for waste collection connected to the other end of the engine box)
- No cord visible (even though it looks like a traditional vacuum cleaner)
- Shadows of the cleaner is incorrect



Gemini Flash 2.5, 31.07.2025

- A very realistic looking vacuum cleaner in a room with all its parts visible and connected



Stable Diffusion 3 Medium, 25.07.2025

- A vacuum cleaner with the pipe broken, it's a hybrid of modern and traditional one
- Legs at the suction area in the front is unrealistic, as usually the wheels are enough (unrealistic)
- Pipe is connected to the wheel, instead of the suction area, and to the top of the motor which is unrealistic and inefficient for the design
- Shadows are inconsistent

**20. “Create a realistic image of a vacuum cleaner plugged into an outlet, with its cord coiled neatly on the floor.”**

Testing:

- Correct depiction of the object
- Avoidance of floating or disconnected objects
- Neat depiction of coiled structure

Expected:

- A fully connected and realistic vacuum cleaner
- The cord of vacuum cleaner plugged in correctly



GPT-4o, 10.7.2025

- Disconnected parts and incomplete vacuum cleaner - no cleaner brush, tube connected to cord and another plug cord coiled neatly on floor
- Plug on ground is incorrectly rendered
- No switch button on the board and a closed/burnt inlet above



Gemini Flash 2.5, 31.07.2025

- A body of vacuum cleaner in a room looks very realistic
- The pipe/cord seems to be blended (unrealistic)
- The cord is floating/hovering vertically over the switch plug point (unrealistic)
- Legs of sofa in the background is floating and not connected



Stable Diffusion 3 Medium, 25.07.2025

- A vacuum cleaner with the pipe broken, while some bits of pipe towards top is rendered
- The cord connection is not shown, just it being stretched out is shown
- Very unrealistic rendering of vacuum cleaner and no prompt adherence

## 21. “Create a realistic image of showing a pinch of sea salt.”

Testing:

- Correct understanding and depiction of the measurement of “pinch”
- Salt - it’s properties and positioning
- Realistic

Expected:

- Index finger and thumb holding some small amount of salt
- Salt in white colour



Dalle3, 17.2.2025; GPT-4o, 03.7.2025

- Both images show salt in white colour and solid crystals
- Earlier image shows salt in heap with some white powder hovering above and background outside the palm
- Later image shows salt in crystals but in different white (not regular iodized maybe) and as a lump of crystals which is not common for salt but is common for sugar
- Earlier image shows the salt in palm while later image shows salt between index and thumb
- Measurement of salt in the earlier image is incorrect (clearly not a pinch) and later image shows a pinch



Gemini Flash 2.5, 31.07.2025

- Some white crystals in between an index finger and thumb to indicate “pinch”
- Scale is unrealistic against the background and the crystals in between the two fingers



Stable Diffusion 3 Medium, 25.07.2025

- A bowl of solid white crystals on a wooden surface
- Nothing indicative of “pinch”

## 22. “Create a realistic image of sugar dissolved in a glass of milk.”

Testing:

- Understanding of what “dissolved” means vs “dissolving”
- Milk’s opacity and colour

Expected:

- A glass of milk with no visible sugar crystals in the glass
- Some sugar somewhere nearby the glass or some sugar in a spoon (maybe)
- Milk to be uniform visually



Dalle3, 17.2.2025

Notes:

- A glass of milk with some foam/bubbles
- Sugar falling from spoon to the glass
- Some sugar crystals and sugar cubes outside the glass
- Milk also flowing from the spoon to glass (quantity, strength and origination of the flow is questionable)
- Reflections on the glass is questionable
- Shadow of the glass is also incorrect





Gemini Flash 2.5, 31.07.2025

- A glass of white milk on a table is rendered
- As expected, there's no remaining of any solid particles that clearly indicates "dissolved"



Stable Diffusion 3 Medium, 25.07.2025

- An unrealistic rendering of glass of milk with a heap of white powder emerging from the middle
- Nothing indicative of "dissolved"

### 23. “Create a realistic image of sugar dissolving slowly in a glass of hot milk.”

Testing:

- Understanding of what “dissolved” means vs “dissolving”
- Understanding of effect of temperature
- Visualization of “dissolving”, a dynamic process
- Accurate rendering of hot
- Realistic interaction between sugar and milk

Expected:

- A glass of milk with maybe some steam (To show hot)
- Sugar crystals settling in bottom in small quantity or swirling of sugar
- Some sugar somewhere nearby the glass or some sugar in a spoon (maybe)



GPT-4o, 03.7.2025

Notes:

- A glass of milk (no steam, nothing to indicate that the milk is hot)
- Brown sugar (maybe to show the dissolution clearly?) crystals settling at the bottom
- Some colour change to the milk (to show dissolution of brown sugar?) however brown sugar doesn't dissolve this way and doesn't change the colour of milk in any way



Gemini Flash 2.5, 31.07.2025

- A glass of white milk, steam to indicate it's hot, kept on a table is rendered
- Rendering is unrealistic as there's bubbles on top and some color change in the middle with crystals of various shapes and sizes settling down
- The steam also seems to be originating from middle of the glass of milk (unrealistic)



Stable Diffusion 3 Medium, 25.07.2025

- A glass of milk with some foam around the glass to indicate it's hot (realistic)
- Nothing indicative of "slowly dissolving" which is okay assuming that sugar is white and is not easy to be captured/seen to naked eye
- Logically unrealistic is the white ceramic plate underneath the glass of milk

**24. “Create a realistic image of spoonful of sugar dissolving slowly in a glass of hot milk.”**

Testing:

- Placement of spoon and sugar
- Accurate rendering of hot and understanding of effect of temperature
- Measurement (spoonful)

Expected:

- A glass of milk with maybe some steam (To show hot)
- Spoon with sugar (To show spoonful) or some interaction with a spoon and the glass
- Sugar crystals settling in bottom in small quantity or swirling of sugar
- Some sugar somewhere nearby the glass or some sugar in a spoon (maybe)



GPT-4o, 03.7.2025

Notes (Same as above):

- Same as previous prompt
- A heap of brown sugar either floating or deposited from below (Incorrect measurement and sugar doesn't float above a glass of hot milk and in a heap state)



Gemini Flash 2.5, 31.07.2025

- A glass of white milk, steam to indicate it's hot, is rendered
- Sugar is falling down to the glass from the spoon- sugar on top of the milk is unrealistic
- Swirl and the reflection in the glass is unrealistic
- The steam is originating from the spoon which is unrealistic as it should be the milk and so should be originating from the glass



Stable Diffusion 3 Medium, 25.07.2025

- A glass of milk in the background while a spoonful of sugar is the focus
- Nothing indicative of “slowly dissolving”

## 25. “Create a realistic image of a river flowing into the sea.”

Testing:

- Connected stream and ocean
- River and sea realistic depictions and realistic meetings of the two water bodies
- Waterflow direction
- Landscape around the water bodies

Expected:

- A small stream of river flowing
- The river joining a sea
- The sea calm or small waves but endless at the other end
- Some movement where the stream of river is joining the sea



Dalle3, 17.2.2025

Notes:

- 1<sup>st</sup> image show huge waves in the stream and increased height in the middle (rivers usually slows down before meeting the sea)
- 2<sup>nd</sup> image show calm flow of river with some foam in the middle of the stream
- 1<sup>st</sup> image show waves in the sea and sea looks calm and endless but the stream and sea is disjoint. The sea also to be divided in the middle
- 2<sup>nd</sup> image show small waves across the stream and no wave in the sea (where is river and where is sea?)
- Reflections and shadows look unreal in both the pictures
- 1<sup>st</sup> image's landscape is unreal with the sand on either side of the stream
- 2<sup>nd</sup> image the trees on either side of stream looks mountainous (pine trees) with strange ending towards the bigger water body
- 2<sup>nd</sup> image the bigger water body looks more like a big lake than a sea



Gemini Flash 2.5, 31.07.2025

- A small stream joining a sea is correctly rendered
- Boat by the side of stream to indicate river
- Slowness of the movement of stream from the texture which is realistic of a river
- Waves by the vast water body and distant hills/mountains to indicate sea



Stable Diffusion 3 Medium, 25.07.2025

- Nothing indicate to if the water body is river or sea
- There's no "river flowing into the sea" as the generated image only shows one water body



**26. “Create a realistic image of a 24-hour clock.”**

Testing:

- Correct labelling of hours
- Avoidance of standard 12-hour format in the clock
- Realistic clock

Expected:

- A digital clock showing time with no am or pm, in 24-hour format
- OR a traditional clock with 24-hour dial (1-24 labelled) with correct number of hands



Dalle3, 17.2.2025

Notes:

- Numbers are illegible or incorrect (12, 4, 12, 33, 44...)
- Not a 24-hour dial (only 12 markings)
- Incorrect number of hands (Only hour and minute hand) and the head/tail of the hands are not of same styles
- 24 hours written on the face to show it's a 24-hour clock
- All markings incorrect





**27. “Create a realistic image of a ball rolling down the floor that has a black box.”**

Testing:

- Relative positioning of box and ball (spatial relationships)
- All touching the floor (gravity, no impossible positioning, floating)
- Rendering of the ball's rolling motion

Expected:

- A ball and a black box placed away from each other on a flat surface
- The ball rolling or about to roll down the floor towards the box



Dalle3, 17.2.2025

- A black box and a black ball correctly positioned relatively, i.e., not colliding etc.
- Box is resting on the floor
- Ball is not touching the floor and is floating above the floor



Gemini Flash 2.5, 31.07.2025

- Prompt is clearly adhered and the ball is touching the floor
- Some halo around the ball probably to indicate the movement
- Shadows and reflections look realistic



Stable Diffusion 3 Medium, 25.07.2025

- Prompt is clearly adhered and the ball is touching the floor
- Nothing to indicate the movement so it could be static as well but then if there are no other disturbances it should have assumed that the box would be the stopping point then the current position is unrealistic
- Shadows and reflections look realistic, but the white area adjacent to the box is unrealistic

## 28. “Create a realistic image of a ball colliding with a box.”

Testing:

- Correct depiction of contact of ball and box
- Understanding of collision with correct deformation on the objects
- No overlap of objects and physical plausibility

Expected:

- A ball and a box in contact with the ball maybe touching the box
- The ball or box may show some small deformities



Dalle3, 17.2.2025

Notes:

- A ball of some visually hard material above a cardboard box
- Small particles shattered and flying above the box- particles are of same colour as the box but of different material so where is the particles coming from?
- Some motion lines are generated to show the direction of ball's movement before the collision
- The direction of motion lines and of the shattered particles movement indicates that the ball to have hit hard on the top of the box- however, the top of the cardboard box is open so collision at this position is not possible physically.
- The ball seems to have collided hard, but the effect of a hard ball and cardboard box seems to be that the box might get deformed inwards at the point of collision depending on the force
- The ball seems to have some deformations instead on top left side



Gemini Flash 2.5, 31.07.2025

- Similar to the results of previous prompt, however the shadow is unrealistic
- The depiction collision is unrealistic (dramatic) – the box and the ball are flying
- The damage on a football from a box most likely made from paper (assumed from the look) is unrealistic



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic image, can't make out the boundaries of the box; shadows are unrealistic; can't make out any effect of the contact on either the ball or the cardboard box

**29. “Create a realistic image of a rubber ball colliding with cardboard box.”**

Testing:

- Material specific deformation and it's properties (rubber compresses and cardboard damage)
- Realistic depiction

Expected:

- A rubber ball with smooth surface and a cardboard box in contact
- The ball or box may show some small deformities



Dalle3, 17.2.2025

Notes:

- A ball and a cardboard box in contact
- Texture of ball seems to be rough
- The deformation on ball and the box is unreal as the ball seems to have a big impact even away from point of impact while the box it's the area of contact that is impacted
- The sticker or tape on the side is unrealistic as there's gaps



Gemini Flash 2.5, 31.07.2025

- The depiction collision is unrealistic (dramatic) with some white lines to indicate contact
- The damage on a rubber ball from a cardboard box is unrealistic
- The textures of the two materials are done right though
- Both are also on the floor
- Shadows are unrealistic



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic, top of the box isn't well defined; shadows are unrealistic; ball is floating; no contact of the two; texture of the ball isn't realistic but the box is acceptable for its colour

**30. “Create a realistic image of rubber ball rolling down the floor toward a glass of milk and hitting it.”**

Testing:

- Correct sequence of action of ball and glass
- Understanding of collision with correct impact on the objects of specific materials

Expected:

- A smooth ball and a glass of milk on a flat surface
- Some cue to show impact of the action like flying glass of milk, shattered glass and splattered milk on the ball or ball touching the glass and glass tilting as if about to fall



Dalle3, 17.2.2025

Notes:

- A smooth ball like a basketball and a glass partially white on a flat surface
- Glass top portion doesn't look like liquid
- Milk flowing down from top to the glass
- No contact between the ball and glass
- Milk wrapped around some parts of the ball however not really in contact but is physically not possible
- Ball seems to be floating
- Drops of milk around that cannot be again physically possible





Gemini Flash 2.5, 31.07.2025

- A rubber ball rolling down the floor wouldn't be floating, so this is unrealistic
- The direction of glass's fall is unrealistic if it's due to the impact of contact with the ball
- The very few milk droplets on the ground are unrealistic



Stable Diffusion 3 Medium, 25.07.2025

- A rubber ball rolling down the floor wouldn't be floating, so this is unrealistic
- The glass is still upright and no contact between the two so "hitting" is not depicted
- The very few milk splashes on the ground are unrealistic
- Shadows are unrealistic

**31. “Can u show realistic image of a thermometer showing 4 degrees Celsius?”**

Testing:

- Accurate scale and marking
- No ambiguity or mislabelling
- Realistic depiction of thermometer

Expected:

- A thermometer with Celsius markings or a digital thermometer showing 4-degree Celsius



Dalle3, 17.2.2025

- Assumed to look like a mercury thermometer with a bulb and markings on the sides
- Liquid spilling out of the bulb
- Incorrect markings and measurements



Gemini Flash 2.5, 31.07.2025

- A mercury thermometer with bulb, some marking and °C marking on the top right side
- Markings are incorrect
- Not indicating 4°C but 40°C



Stable Diffusion 3 Medium, 25.07.2025

- A mercury thermometer from vein's red color and some markings
- Markings are incorrect and illegible numbers
- 4 shown on the bulb or base but rather than being measured it looks like a print

**32. “Create a realistic image of a hot air balloon flying across the sky.”**

Testing:

- Balloon realistically above ground, straight
- Physically possible

Expected:

- A hot air balloon with some people in the basket, in the sky with some clouds



Dalle3, 17.2.2025

Notes:

- Depiction of balloon was as expected, big and upright
- The texture of the balloon also looks realistic
- Cloud looks unrealistic because of the size, density, scale and proximity around the balloon
- No visible burner flame even though the balloon is in flight
- Suspended ropes in the sky or between clouds visible
- The ropes and basket connections should be correctly attached



Gemini Flash 2.5, 31.07.2025

- Depiction of balloon was as expected, big and upright, the texture is realistic as well
- The backdrop is very realistic
- Balloon is inflated indicating flight state however no visible burner flame
- There is extra fabric behind/other side of the balloon where the ropes should be in place
- The ropes and basket connections should be attached but there are some loose ones



Stable Diffusion 3 Medium, 25.07.2025

- Depiction of balloon was as expected, big and upright, the texture is realistic as well
- The backdrop is very realistic
- Balloon is in inflated state indicating flight state but no visible burner flame
- The ropes and basket connections should be attached but there are some loose ones

**“Create a realistic image of ...”**

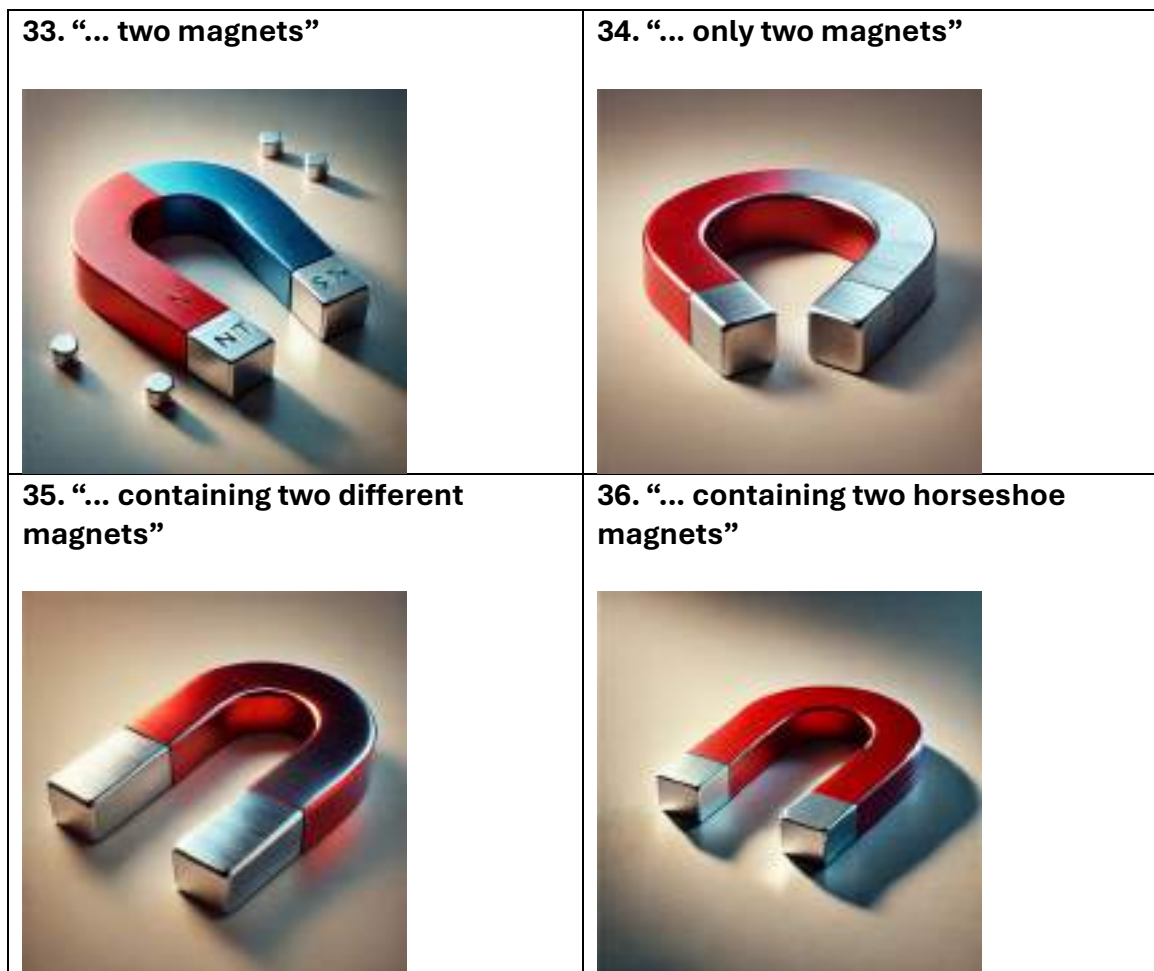
Testing:

- Realistic magnets
- No mislabelling or confusion with magnet types
- Accurate number/count of objects
- No impossible configurations (merged etc.)

Expected:

- Two different magnets to be generated with or without the poles marked

*Generated images using Dalle3, 17.2.2025*



- None of the images show two magnets
- No realistic shadows
- All prompts generated horseshoe/U-shaped magnets, and they all have metallic ends
- Prompt 33’s image shows two types of magnets (horseshoe and circular) instead of two magnets
- Prompt 33’s image shows incorrect markings for North and South (N, NT, STH ambiguous)

- Prompt 35's image shows a U-shaped magnet with metallic ends (same as all other images) except the metallic ends are longer and a clearer boundary between the horseshoe and the ends. This gives the impression that the metallic end might be an unlabeled bar or a simple iron bar
- Prompt 36's image contains only one U-shaped magnet





*Generated images using Stable Diffusion 3 Medium, 25.07.2025*

<p><b>33. "... two magnets"</b></p> 	<p><b>34. "... only two magnets"</b></p> 
<p><b>35. "... containing two different magnets"</b></p> 	<p><b>36. "... containing two horseshoe magnets"</b></p> 

- Prompt 33 - as expected (but no reaction between each other)
- Prompt 34 - looks like fridge magnets with the text printed on them
- Prompt 35 - Two bar magnets (but no reaction between each other)
- Prompt 36 - only one horseshoe not sure if it's magnetic
- All the items look like metal objects



Generated images using Gemini Flash 2.5, 31.07.2025 , 31.07.2025

<b>33. "... two magnets"</b> 	<b>34. "... only two magnets"</b> 
<b>35. "... containing two different magnets"</b> 	<b>36. "... containing two horseshoe magnets"</b> 

- Prompt 33 – Two shapes in blue and red; marking on 2<sup>nd</sup> one doesn't have "N"
- Prompt 34 – Two U-shaped magnets of red and blue color kept side by side
- Prompt 35 – One U-shaped magnet and one bar kept side by side, but the bar's magnetic properties can't be clearly identified (nothing indicating that it's a bar magnet)
- Prompt 36 – Two horseshoe magnets correctly generated in red and blue colors

*Conclusion – MLLMs generated images with red-blue indicative of magnets. However, there's no evidence of its magnetic properties like opposites attract, repelling of likes, or attraction of metal items towards itself.*



### 37. “Create a realistic image of the moon revolving around the Earth.”

Testing:

- Correct depiction of revolution (moon’s path around Earth)
- No confusion with rotation
- Realistic scale and arrangement

Expected:

- Earth and moon, with the moon depicted in an orbital path around the Earth



Dalle3, 17.2.2025

Notes:

- Two moons visible closely to the Earth
- Closer moon is smaller than the Earth and the other moon is much smaller comparatively
- Accurate depiction of the Earth and the moon’s surfaces



Gemini Flash 2.5, 31.07.2025

- Earth and moon are rendered, are partially dark
- Exact half portions with sun-facing on both
- American continent is visible
- Position of sun can be identified with the small light in the top-left corner of the image



Stable Diffusion 3 Medium, 25.07.2025

- A realistic picture of earth, moon is not visible, cannot make out the continent

**38. “Create a realistic image of the moon rotating around the Earth.”**

Testing:

- Model’s understanding of “rotation” vs. “revolution”
- Accurate depiction of motion
- Avoidance of static or misrepresented positions

Expected:

- Ideally, the moon shown both spinning on its axis and orbiting Earth, or at least not confusing the two



Dalle3, 17.2.2025

- Two moons visible and two earths also visible
- The smaller moon ‘s surface isn’t as expected



Gemini Flash 2.5, 31.07.2025

- Earth and moon are rendered, are partially dark
- Position of sun can be identified with the parts of the two visible
- African continent is visible



Stable Diffusion 3 Medium, 25.07.2025

- A picture of blue ball with white patches all over, cannot even say if it's the earth

**39. “Can you create a realistic image to show how the Statue of Liberty would look like from the Moon?”**

Testing:

- Realistic portrayal of the Earth from the moon’s perspective
- Avoidance of impossible detail (Statue of Liberty visible from the moon)
- Understanding of physical limits of visibility and scale

Expected:

- Earth as seen from the Moon, with the Statue of Liberty not visible (due to extreme distance and size disparity)



Dalle3, 17.2.2025

Notes:

- Statue of Liberty on the moon’s surface (physically possible but not expected)
- Earth seems closer to the moon
- Another moon or planet visible far away, gives off the impression of being another moon (physically impossible)



Gemini Flash 2.5, 31.07.2025

- Earth and moon are rendered, earth is partially dark
- To show perception, only moon crater is visible and viewed it from here
- African continent is visible (not American, Statue of Liberty is in USA)



Stable Diffusion 3 Medium, 25.07.2025

- A picture of statue of liberty as island
- No perception of moon

**40. “Create a realistic image of a table and chair beside each other.”**

Testing:

- Proper alignment and proximity of the two objects
- Realistic scale and support (no floating or intersecting)
- Commonsense arrangement

Expected:

- A table and a chair positioned side by side, both upright and on the same surface



Dalle3, 17.2.2025

Notes:

- A table and chair were rendered correctly
- They are properly placed with a reasonable distance
- No intersection or floating of the two
- Shadows of the two are incomplete and incorrectly rendered
- Rods connecting the legs of the chair not correctly connected and irregular shapes



Gemini Flash 2.5, 31.07.2025

- Prompt adhered, realistic



Stable Diffusion 3 Medium, 25.07.2025

- 2 or more chairs around a table (prompt was only 1)
- Chair legs are disconnected/incomplete
- No shadows



#### 41. “Create a realistic image of fish swimming in a glass.”

Testing:

- Understanding of object size (fish should fit in the glass)
- Realistic depiction of water and fish interaction
- Avoidance of impossible scenarios (fish out of water or glass overflowing)

Expected: One or more fish inside a drinking glass, in water



Dalle3, 17.2.2025

- A glass of water on a plane surface
- A fish inside the glass
- Small bubbles in the water
- Some reflection on the glass but the black shade that suddenly stops at the fish's height is unrealistic



Gemini Flash 2.5, 31.07.2025

- Prompt adhered
- Fish inside the glass is realistic
- Ripple in the middle is unrealistic



Stable Diffusion 3 Medium, 25.07.2025

- Realistic fish inside a glass of water

#### 42. “Create a realistic image of fish swimming in an aquarium.”

Testing:

- Accurate representation of an aquarium setting
- Realistic depiction of water and fish interaction
- Proper containment (no fish outside the tank)

Expected:

- Fish swimming inside a glass aquarium, with water, possibly plants or stones



Dalle3, 17.2.2025

Notes:

- A tank/aquarium with multiple small fishes, plants, sand, stones and water filter
- Some fishes not fully rendered with all parts
- Filter is not connected properly
- Scale of the aquarium against the background room's ceiling – not realistically possible given the size of the fish
- Rendering of the reflections (top in the water and both sides)
- Shadow of the aquarium doesn't look realistic



Gemini Flash 2.5, 31.07.2025

- Prompt adhered
- Small fish inside the aquarium is realistic
- No effect on water for the movement (including breathing) is not rendered and so is unrealistic (feels too still)



Stable Diffusion 3 Medium, 25.07.2025

- Realistic fish inside an aquarium
- Unrealistic is that some fish have another eye in place of mouth
- The two fish in focus seems to out of the aquarium

#### 43. “Show me a picture of a half-empty, slanted glass.”

Testing:

- Correct orientation of the liquid surface (not matching the glass slant)
- Realistic liquid volume and containment
- Avoidance of any impossible liquid behaviours

Expected:

- A glass tilted at an angle
- Water remaining horizontal due to gravity



Dalle3, 19.2.2025

Notes:

- A half-empty glass of water that's slanted (as expected)
- Shadows not looking realistic
- Bubbles and wave reflect movement of the water upon slanting the glass but the rendering is unrealistic especially the wave/swirl
- The swirl is unrealistic as the glass somehow doesn't have anything to show action being done



Gemini Flash 2.5, 31.07.2025

- Prompt adhered
- Shadow/reflection seem to be slightly inconsistent



Stable Diffusion 3 Medium, 25.07.2025

- Glass is of unusual shape, no water, no slanting position

**44. “Show a realistic image of a nut being tightened onto a bolt using a wrench.”**

Testing:

- Correct generation, placement and orientation of the objects
- Realistic depiction of tightening action
- Avoidance of floating or misaligned components

Expected:

- A wrench locked onto a nut
- The wrench turning the nut onto a threaded bolt
- All parts in correct alignment



GPT-4o, 13.6.2025

Notes:

- Nut, bolt, and wrench is rendered (as expected)
- Placement and orientation of the nut and bolt is correctly rendered (as expected)
- Nut is partially being tightened on the bolt
- Wrench not locked around the nut and so doesn't show tightening action correctly
- The lower part of the wrench head is incorrectly rendered as if passing through the bolt



Gemini Flash 2.5, 31.07.2025

- Nut is rightly threaded, however, the position of the wrench is unrealistic as it doesn't help with the tightening process
- Unrealistic is that the bolt doesn't have any hold for the threading/tightening to happen while it's only the wrench



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic rendering- multiple bots, can't make out what's actually being tightened
- Fingers are unrealistic and incomplete
- The threaded end looks like it was soldered into the metal or that the end is melting while it's being inserted



**45. "Create a realistic image of the sun rising and a compass with the needle pointing to the direction of sunrise."**

Testing:

- Correct orientation of compass (needle points east)
- Realistic sunrise depiction

Expected:

- The sun low on the horizon (east), with a compass needle pointing toward it
- Compass needle pointing to east



GPT-4o, 11.6.2025

Notes:

- Sun and a pocket compass with red head in the needle are rendered
- Sun is lower in the horizon, looks like a sunrise (as expected)
- Needle of the compass is pointed to North-East direction and not East
- Markings of direction are incorrect, North-West and South-West are not rendered



Gemini Flash 2.5, 31.07.2025

- Prompt is well adhered- compass and sunrise is rendered correctly
- Compass is correctly marked with all the directions clearly visible
- Direction looks realistic



Stable Diffusion 3 Medium, 25.07.2025

- Prompt is well adhered- compass and sunrise is rendered correctly
- Cannot make out the directions so “needle pointing direction of” can’t be evaluated

**46. "Create a realistic image of someone holding six plates of sandwiches on one hand."**

Testing:

- Correct number of objects
- Realistic hand posture and balance
- Avoidance of impossible stacking or floating plates

Expected:

- A person balancing six plates stacked or arranged on one hand



GPT-4o, 11.6.2025

- Incorrect number of plates
- Incorrect number of sandwiches
- Sandwiches doesn't look like halves of each other
- Random slices of bread in some of the plates at the bottom (sandwich on top of the slice of bread in the top two plates)



Gemini Flash 2.5, 31.07.2025

- Prompt is adhered- 6 plated in one hand
- Some of the plates contain wraps, mini burger buns etc. instead of sandwiches
- Also shows that it's wobbly as the top plate is not exactly vertically
- Looks realistic



Stable Diffusion 3 Medium, 25.07.2025

- A person is holding a plate of sandwiches in their two hands (not adhering to the prompt)
- Hands are not rendered correctly
- Looks like there are almost 6-7 sandwiches in the plate

#### 47. “Can you create a realistic image of a rocket falling into the sea?”

Testing:

- Realistic depiction of falling motion of the rocket (motion, gravity)
- Correct interaction between rocket and water
- Avoidance of static or floating

Expected:

- A rocket descending toward and falling into the sea, (maybe exploding) with clouds of smoke, huge water splashes and maybe hints of fire
- OR a crashed rocket with parts disintegrated and partially floating/sinking
- Reference: [https://www.youtube.com/watch?v=1vsKQ\\_wTMU4](https://www.youtube.com/watch?v=1vsKQ_wTMU4)



Dalle3, 19.02.2025; GPT-4o, 11.6.2025

Notes:

- Rocket and an endless sea are rendered as expected
- Image 1, the rocket seems to fly out of the sea with the direction of the head, the fire and the impact on the sea
- Image 1 doesn't look realistic as well, as the water displaced and the water nearby seems to have a boundary
- Image 2, looks like a safe landing rather than falling as the head is still upright and pointing upwards
- Image 2, also looks like a take-off minus the fire as there has been no contact yet with water unless the rocket has flied off from the sea



Gemini Flash 2.5, 31.07.2025

- A damaged rocket in the middle of the sea
- The top part is damaged
- Smoke and big splashes of water around the rocket
- Looks realistic
- The upright position for the fall is debatable



Stable Diffusion 3 Medium, 25.07.2025

- Rocket flying away from the sea, smoke and cloud around the sea
- Fire from the rocket and direction says it's flying away

**48. “Can you show me a glass of half filled with oil and half filled with water?”**

Testing:

- Correct order of liquids (density, layering)
- Visible separation between layers
- Realistic depiction of transparency and colour (liquid properties)

Expected:

- A glass with bottom half filled with water and top half filled with oil (yellow colour and thicker than water) due to density difference of the two liquids



Dalle3, 19.02.2025; GPT-4o, 11.06.2025

Notes:

- Images contain a glass filled with some liquids (not half-filled also)
- In Image 1, the two liquids are vertically separated
- Image 1- One part is transparent; the other part is yellow but looks just as dense as water
- Image 1- No clear boundary between the two liquids
- There are some small bubbles in oil and water along with the reflections, the image looks unrealistic
- In Image 2, the liquids seem to be visually different. The yellow liquid looks visually denser than water.
- The yellow liquid is above the water
- Glass is half-filled with water, however, not half-filled with oil as well
- There are some small bubbles in oil and not in water, which looks realistic
- There is clear boundary and horizontally separated between the two liquids (as expected)



Gemini Flash 2.5, 31.07.2025

- Glass is not half-filled with water and half-filled with oil
- Depicts density correctly, denser oil on top of water



Stable Diffusion 3 Medium, 25.07.2025

- Glass partially filled with yellow colour liquid only, no transparent liquid (water)
- Another glass in background with some transparent liquid and something yellow inside it



**49. “Create an image of three books stacked neatly on a table, with the middle book slightly larger than the others.”**

Testing:

- Proper stacking
- Clear depictions of size differences
- Correct number of books

Expected:

- Three books stacked, middle one larger, all edges visible, neatly aligned



Dalle3, 19.02.2025; GPT-4o, 11.06.2025

Notes:

- Images contain books neatly stacked on top of each other on a wooden surface
- Image 1 contains 4 books while Image 2 contains 3 books
- Images are arranged in the order of increasing size
- In both the images, the middle ones are not larger than the one at the bottom (clearly not as expected)
- In Image 2, the middle book's rear bottom edge seems to be oddly in blue color (as the bottom book) unlike rest of its body in beige color



Gemini Flash 2.5, 31.07.2025

- Three books are stacked on top of each with their own titles and colours clearly visible and legible
- The book in the middle is thicker than the other two



Stable Diffusion 3 Medium, 25.07.2025

- 4 or more books on top of a wooden surface
- Can't make out unrealistic books/covers
- 2<sup>nd</sup> book (light brown colour) is slightly thicker than others

**50. “Create a realistic image of a chess board at the starting position.”**

Testing:

- Accurate depiction of chessboard pattern (pattern recognition)
- Correct arrangement of pieces (colour, position)
- No missing or extra pieces

Expected:

- A chessboard with all pieces in their correct starting squares



Dalle3, 19.02.2025; GPT-4o, 11.06.2025

Notes:

- Images contain incorrect number of pieces
- Both the images show incorrect depiction of chessboard pattern- same color tiles close-by and checkboard size is not 8x8
- Black horse pieces and white horse pieces look different (similarly all others)
- Image 2 doesn't depict the starting position of a game as the pieces are not correctly placed inside the squares
- Also, the black side there are some white pieces instead of black, that means there are more white pieces on the board than black



Gemini Flash 2.5, 31.07.2025

- Arrangement of the chessboard pieces is correct
- The chessboard doesn't have 8x8 squares and some same colour squares are together



Stable Diffusion 3 Medium, 25.07.2025

- Doesn't look like starting position
- The board doesn't have proper squares
- The pieces are random and shapes are not well defined

**51. “Create a realistic image of falling dominoes interrupted by a rubber duck placed between two dominoes.”**

Testing:

- Proper placement of the duck between dominoes
- Realistic depiction of interrupted motion
- No dominoes passing through the duck

Expected:

- Dominoes in a line, some falling, with a rubber duck blocking the chain reaction



GPT-4o, 13.06.2025; GPT-4o, 7.07.2025

Notes:

- Images contain a yellow rubber duck and dominoes some vertically standing and some collapsed
- Image 1, some of the dominoes have collapsed onto the duck. The other domino is standing further away from the duck (not equally spaced, not expected)
- Image 2, the falling dominoes not properly resting on the duck (not past tense then?), two fallen dominoes isolated from the chain (unrealistic and not expected)
- Image 2, if it's still “falling” and not yet “interrupted” then it's still realistic on the front of why there's no contact



Gemini Flash 2.5, 31.07.2025

- Yellow duck is in between two dominoes, one section falling while other still untouched
- No contact between the falling domino and duck to say “interrupted” in past tense, but if it’s “falling” this make a suitable image



Stable Diffusion 3 Medium, 25.07.2025

- Random dices (some are not cubes) around a yellow duck
- The dices are floating, not touching the ground
- Prompt not adhered

**52. “Create a realistic image of a tennis ball in a swimming pool.”**

Testing:

- Correct floating position (buoyancy)
- Realistic water effects (motion, reflection)
- Proper scale of ball and pool

Expected:

- A tennis ball floating (partially submerged) on the surface of a pool
- Water surface may have some ripples



GPT-4o, 13.06.2025

Notes:

- Tennis ball and pool are rendered of correct size and color
- Ripple effects on pool's surface
- Tennis ball partially submerged and floating on water
- Some reflection in the pool
- The submerged portion is not wet (darker shade around the around of contact with water)





Gemini Flash 2.5, 31.07.2025

- Image is very realistic – partially floating/sinking, ripples around, correct reflections, wet section on the ball for the area of contact, distorted tiles
- Reflection area is though very still, almost looks like a glass



Stable Diffusion 3 Medium, 25.07.2025

- Tennis ball is floating in a swimming pool on top of a wave, no wetness, no partial sinking
- Reflection is good, not the shadow



### 53. “Create a realistic image of a rock thrown into a swimming pool.”

Testing:

- Correct submerged position (buoyancy)
- Realistic water effects (motion, reflection etc.) and rock textures (material depiction)
- Proper scale of rock and pool

Expected:

- A rock sinking down a pool
- Maybe be some effects on water like ripples to show action



GPT-4o, 13.06.2025

Notes:

- Rock and pool are rendered of correct size and color
- Ripple effects on pool's surface
- Rock partially submerged and floating on water (not expected- expected for the rock to sink)
- The submerged portion is wet (i.e., darker shade around the around of contact with water)



Gemini Flash 2.5, 31.07.2025

- Image is very realistic – rock just thrown into the water with the huge splash around it and the ripples as effect with the distorted shape of tiles when seen through water
- Unrealistic component is the water rising from the middle of the rock



Stable Diffusion 3 Medium, 25.07.2025

- Rock is floating in a swimming pool, slight wetness on the sides, no partial sinking
- The shadow of the rock is unrealistic
- If the rock was thrown, no effect or nothing on the rock to indicate such an action or effect

**54. “Create a realistic image of two blocks – red and green, with the red block in the middle of the green block. The red block is larger than the green one. The green block features a round hole in its center on the other side.”**

Testing:

- Accurate depiction of size and placement (Spatial reasoning)
- Realistic handling of holes
- No impossible containment

Expected:

- A green block with a hole, a larger red block partially visible through the hole, but not fully contained



GPT-4o, 13.06.2025

Notes:

- Red and green colored square blocks are correctly rendered
- Green colored block contains a hole
- Red block smaller than green block (not expected)
- Red block sitting on top of green block (not expected)



Gemini Flash 2.5, 31.07.2025

- Image is very realistic – understood the complexity in the prompt
- Red is bigger than green; green does have a hole; part of the red block is in middle of the green block
- Neither of the shapes have been defined to be cube so the shapes are very realistically generated to match the requirements of the prompt



Stable Diffusion 3 Medium, 25.07.2025

- Two green blocks and a red block; one green block has a hole
- One of the green blocks is incomplete

**55. “Create a realistic image of a football rolling down the floor and hitting a brick wall in front.”**

Testing:

- Realistic depiction of motion - rolling
- Realistic collision depiction (ball stops or deforms)
- No penetration or floating

Expected:

- A round black and white football, in motion, about to or just touching a brick wall



GPT-4o, 13.06.2025

Notes:

- An American football was rendered lying against a brick wall with no deformations on the ball or the wall (expected)
- American football was not expected, instead a soccer ball was expected (ambiguity of English)
- The American football's black stitches are inconsistent and incomplete (not expected)
- Shadows not present (unrealistic)



Gemini Flash 2.5, 31.07.2025

- Realistic impact on the football and wall on the ball hitting the brick wall
- Some halo around the ball to show the action of previously moving
- Ball is above the ground which is realistically not possible if it was previously rolling down the floor
- Shadow is unrealistic



Stable Diffusion 3 Medium, 25.07.2025

- An American football with a brick wall in the background
- Ball in static position, not rolling down and not showing hitting which would involve contact with the wall
- Shadow is unrealistic

**56. “Create a realistic image of the situation - water is thrown into a fireplace with a burning fire and logs of wood.”**

Testing:

- Realistic depiction of water-fire interaction (Dynamics and material interaction)
- Proper rendering of logs, flames, and water
- No impossible coexistence (fire unaffected by water)

Expected:

- Water visibly thrown onto burning logs in a fireplace
- Effect on fire - steam, smoke, flames diminishing, ashes turning black from grey



GPT-4o, 13.06.2025

- Water thrown from top with splashes and steam coming out from the logs going upwards
- Logs of wood in a brick fireplace with some grey ashes
- Logs turning black and ashes becoming darker showing water interaction
- Water flowing away from the fireplace as the water flows
- Diminishing fire at the top as it interacts with water
- The logs of wood floating without a balance (not resting on anything, not expected)
- Logs turned black however ashes nearby haven't changed color (not expected)



Gemini Flash 2.5, 31.07.2025

- Wood burning by a fireplace with ashes and soot around
- Water flowing from top to down, however drops of water are all over the place which is unrealistic given the direction of flow



Stable Diffusion 3 Medium, 25.07.2025

- Wood burning brightly in a fireplace, no sign of water
- No soot as well as no ash around the fireplace
- All the wood look freshly placed



**57. “Create a realistic image of a stool with only three legs, equally spaced and of equal height.”**

Testing:

- Stability of the stool- equal heights, angles and spacing (Physical and structural plausibility)
- No tipping (Physical and structural plausibility)
- Legs under the seat (Structural plausibility)
- Legs touching the ground, no floating (Physical plausibility)

Expected:

- A well-balanced stool seen from the side, three legs of equal heights and at equal angles under the seat, all touching the ground, seat level



GPT-4o, 05.07.2025

Notes:

- Wooden stool with three legs, all identical in height, spacing and angle
- All three legs touching the ground, no floating
- Legs correctly connected to the seat
- Shadows seem unrealistic



Gemini Flash 2.5, 31.07.2025

- Wooden stool with four legs and follows all expected rules
- Didn't follow prompt for it to have only 3 legs



Stable Diffusion 3 Medium, 25.07.2025

- Wooden stool with four legs and follows all expected rules
- Didn't follow prompt for it to have only 3 legs
- Legs also are not equally spaced, angle and heights

**58. “Create a realistic image of a one-legged stool.”**

Testing:

- Stability of the free-standing stool (Physical and structural plausibility)
- Geometric relationships
- Leg touching the ground, no floating

Expected:

- A well-balanced stool design with a single leg



GPT-4o, 05.07.2025 (*Different conversations*)

Notes:

- Wooden stool with one leg, at an angle (like three-legged stool minus two legs)
- Leg touching the ground, no floating in both the images
- Leg correctly connected to the seat in both the images
- Angle of the leg is unrealistic and can't help with the stability of the stool, given the heavy material too



Gemini Flash 2.5, 31.07.2025

- Wooden stool with one leg and standing against the wall
- Expected it to find a design with balance however there's no balance with the little contact and given however poorly balancing this design is
- Shadows are not right



Stable Diffusion 3 Medium, 25.07.2025

- Wooden stool with three legs that equally spaced and angled and may not be of equal heights
- Didn't follow prompt for it to have only 1 leg
- Shadows are not right
- Shape of the carpet is also unrealistic

**59. “Create a realistic image of a one-legged plastic stool”.**

Testing:

- Stability of the free-standing stool (Physical and structural plausibility)
- Geometric relationships
- Characteristics of material - Plastic

Expected:

- A well-balanced plastic stool with a single leg



GPT-4o, 05.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- Plastic stool with two legs (like a three-legged stool minus two legs)
- Leg touching the ground, no floating
- Leg correctly connected to the seat
- Balance of the design is not physically plausible



Gemini Flash 2.5, 31.07.2025

- Similar to the wooden stool with one leg and standing against the wall
- Expected it to find a design with balance however there's no balance with the little contact and given however poorly balancing this design is
- Shadows are not right
- Colour and texture look like plastic



Stable Diffusion 3 Medium, 25.07.2025

- Metal/plastic stool with four legs with a blue plastic seat
- Didn't follow prompt for it to have only 1 leg
- Shadows are not right

**60. “Create a realistic image of a one-legged plastic stool, where the leg is perpendicular to the seat and the end is slightly larger than the rest of the leg, but not larger than the seat itself.”**

Testing:

- Stability of the free-standing stool (Physical and structural plausibility)
- Geometric relationships, logical proportions and spatial reasoning of dimensions
- Characteristics of material - Plastic

Expected:

- A plastic stool with a single leg where base of the leg is larger but smaller than the seat
- Only one leg connected to the middle of the seat
- Leg is perpendicular to the seat



GPT-4o, 05.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- Image generated as expected
- Physically plausible with the given image and design of the stool



Gemini Flash 2.5, 31.07.2025

- A very common one-legged plastic stool design rendered correctly, followed the prompt well
- Shadows are not right



Stable Diffusion 3 Medium, 25.07.2025

- Can't make out the material from the texture
- Three legs are different from each other
- No shadows



**61. “Create a realistic image of someone looking into the rearview mirror while driving.”**

Testing:

- Depiction of the situation- a driver in a car, driving (context awareness)
- Realistic depiction of looking into mirror (geometric reasoning and reflection)

Expected:

- Person in a driver’s seat, eyes or face directed toward the rearview mirror, with a realistic car interior



GPT-4o, 05.07.2025

Notes:

- A male driver with hands on the steering wheel is rendered, left drive
- Interior of the car doesn't look right – spacing and sizing are odd
- His posture is not right and the way he's looking at the mirror is odd
- The mirror's angle is also odd



Gemini Flash 2.5, 31.07.2025

- Someone's driving a car with left side driving and right line policy which is correct
- The light beside the rearview mirror is usually not on so it's a bit unrealistic
- The person is looking above not at the rearview mirror
- The relative positioning of seat, the steering wheel, the mirror is not realistic



Stable Diffusion 3 Medium, 25.07.2025

- Both images not realistic (the right one "driving a car" was specified in the prompt)
- Position of the rearview mirror and the view of the mirror are not realistic
- Image is incomplete with no understanding of context of "rearview mirror"

**62. “Create a realistic image of a person carrying a large mirror across a busy street.”**

Testing:

- Depiction of a complex scenery- a road with traffic and pedestrians
- Realistic handling of reflective surfaces in complex scenery

Expected:

- Person gripping a large mirror (showing reflections), crossing a street with visible traffic or pedestrians



GPT-4o, 05.07.2025

Notes:

- A male holding a plane surface while crossing a road through zebra crossing
- No reflections on the mirror or the view on the surface looks unrealistic as the person holding itself is not there
- The angle he's walking also looks off
- The road also looks unrealistic as there's two sets of pedestrian marking where the person is crossing the road as well as there's one not very far away, unrealistic



Gemini Flash 2.5, 31.07.2025

- Someone's carrying a mirror across a very busy one-way three-lane street
- Reflection is not accurate on the mirror (eg: the taxi behind in the reflection is not realistically possible as understood from the street behind)
- The signal behind is also unrealistic as simultaneous on the same street there's red and green lights on
- All cars also seem to have their headlights on



Stable Diffusion 3 Medium, 25.07.2025

- Very unrealistic rendering, the person's going towards the street; the mirror and it's reflections is unrealistic; the black object in middle of the road is unrealistic; the markings on the road is unrealistic

**63. “Create a realistic image of a person walking backwards into a busy street while texting.”**

Testing:

- Proper body posture, orientation and movement depiction
- Realistic depiction of situation - busy environment and personal risk

Expected:

- Person facing away from street, walking backward, eyes on phone, possibly unaware of traffic



GPT-4o, 05.07.2025

Notes:

- A male holding a phone while crossing a road through zebra crossing
- Movement of his legs (back leg heels lifted and front leg heel touching the ground) look as if moving forward
- Background is blurred but few actions we can see – people crossing the road not on the pedestrian crossing
- Multiple traffic signals also visible, in close proximities to each other which is also unrealistic
- Headlights on for all the vehicles even if from the image it looks like day (unrealistic again)



Gemini Flash 2.5, 31.07.2025

- Someone's crossing across a very busy street on the pedestrian crossing with their phone in their hand – very realistic rendering with the signal lights, street markings, buildings etc.
- Can't make out if the person's walking back, it looks like walking forward because of the position of the foot that's off the ground
- All cars also seem to have their headlights on



Stable Diffusion 3 Medium, 25.07.2025

- Very unrealistic rendering- the cars, the street, the pedestrian crossing, the person, flatness of the street
- The person is just standing on their toes, not walking
- Unrealistic rendering of the hand holding the phone



**64. “Create a realistic image of a person pouring hot milk from a big pot into a small cup.”**

Testing:

- Correct liquid trajectory, no impossible flows or spills (scale and containment)
- Accurate depiction of liquid filling process and proportions

Expected:

- Person holding a large pot, milk flowing in a stream into a cup, possible steam above cup



GPT-4o, 05.07.2025

Notes:

- A person holding a big steel/aluminum saucepan (pot) of milk with froth on the sides and milk skin formed on top, with an orange/golden line at the bottom of the saucepan
- The golden line at the bottom of the saucepan might be to signify that the pot is hot but it's unrealistic for a metallic pot on a stovetop to look like this
- The froth and the interior walls of the pot look realistic
- The milk skin formed on top of the milk in the pot look realistic
- The movement of milk from the pot to the cup look realistic
- The reaction of the milk in the cup looks unrealistic to the pouring milk
- Neither the pot nor the cup has any steam which is expected from hot milk
- The holding position of the pot by the person is unrealistic - holding this way of hot milk with a hot pot base is risky and we normally pour at the side so we can see milk flow and the cup clearly, unless it's for the photo to show the milk in the pot
- The body posture is unrealistic - if the person is to pour this way, they would at least lean forward to see the amount and angle while pouring is correct but here it is a stoic posture



Gemini Flash 2.5, 31.07.2025

- Milk poured from a steel pot to a mug
- Holding is realistic
- Steam only from the cup, not from the pot which is unrealistic
- Reflection on the pot of the cup is incorrect



Stable Diffusion 3 Medium, 25.07.2025

- Milk poured from a steel small vessel to a big pot, big pot already has  $\frac{3}{4}$  filled with milk
- Some steam around the pots is realistic
- The amount in the big pot isn't what the small one can hold, which means it might have pre-filled
- The big pot, the fingers, hands are incompletely rendered and unrealistic



**65. “Create a realistic image of a person opening a refrigerator door.”**

Testing:

- Correct articulation of hand placement and position
- Correct depiction of object manipulation- door opening direction
- Realistic interior details

Expected:

- Person’s hand on fridge handle, door swinging open, interior visible



GPT-4o, 06.07.2025

Notes:

- A person standing beside with their hand on refrigerator’s door handle is seen
- Interior of the fridge looks realistic (ignoring the size)
- Position and angle of the person is unrealistic towards opening a refrigerator door
- The door looks unrealistic as there’s no gasket visible on the side of the door that helps in the insulation
- The light inside the refrigerator also looks unrealistic (upon opening the door, the entire refrigerator would be lit up at least the top portion)
- There are no shelves in the door side which isn’t the case ideally or realistically speaking (very rare case)
- The angle of door and size is unrealistic



Gemini Flash 2.5, 31.07.2025

- Hands on a fridge door open
- There's light inside the fridge too
- Can't understand the direction or mechanism for opening the door so unrealistic



Stable Diffusion 3 Medium, 25.07.2025

- The fridge door is open, there's someone in front of the fridge
- Fridge is empty and there's no lights inside so maybe the fridge isn't in a working condition
- Fingers are unrealistic and incomplete

66. “Create a realistic image of a bicycle leaning against a wall.”

Testing:

- Plausible contact between bicycle and wall
- No floating or unstable configuration (surface relationships)
- Balance

Expected:

- Bicycle angled, one side touching the wall, both wheels on ground, standing via support from the wall



GPT-4o, 06.07.2025

Notes:

- A bicycle touching the floor kept with the support of only the wall
- The only contact between the cycle and the wall is one of the sides of the cycle's handle
- The cycle is upright no leaning towards the wall which is unrealistic given the only support is the contact between the handle and wall (which is small)
- The rendering of the wheel is incorrect, especially of the rear end (some metallic rods interconnected and some are disjoint)
- Mudguard (the rear end) is merging with the tyre
- Only one pedal is visible (possibility that the other pedal is hidden based on the view angle) and the pedal incorrectly rendered



Gemini Flash 2.5, 31.07.2025

- A bicycle with incomplete parts standing next to a wall
- No support taken from the wall and is not leaning on the wall
- Shadows are unrealistic
- Handlebar design and paddle are unrealistic as well



Stable Diffusion 3 Medium, 25.07.2025

- Bicycle against a wall is rendered
- Parts are incomplete or unrealistically rendered – the handlebar and paddle
- Only contact is the handlebar which is unrealistic
- No shadows

**67. “Create a realistic image of two people shaking hands across a table.”**

Testing:

- Realistic human interaction and table placement
- Correct orientation and body positioning

Expected:

- Two people facing each other, arms extended, hands meeting above or across the table



GPT-4o, 06.07.2025

Notes:

- A table with papers and pen, with two people sitting across from each other at either side of the table, giving a handshake
- People wearing different color suits (scene representing a successful business deal in a meeting)
- Unrealistic component in the generated picture is the rendering of the hands- the length of the fingers of both and the number of fingers on the black suit person's hand)
- Bias that the business setting is with two males (understood from the hands)



Gemini Flash 2.5, 31.07.2025

- As expected, and very realistic
- Shadows are not correct
- Eye contact between the two people (socially it's an important element to handshake)



Stable Diffusion 3 Medium, 25.07.2025

- Two people shaking hands
- One of the hands of the male is blending into the table
- Shape of the table can't be determined

**68. “Create a realistic image of a sandwich cut in half on a plate.”**

Testing:

- Both halves clearly visible (partitioning)
- Proper plate and sandwich interaction (containment and placement)

Expected:

- Slicing of sandwich on a plate
- OR two halves side by side or overlapping, both on the plate



GPT-4o, 06.07.2025

Notes:

- A white plate on a table (wooden surface) with a sandwich rendered
- Unrealistic component - the sandwich doesn't look like halves instead a single sandwich (only three bread slices; tomato, leaves and cheese doesn't have the same arrangement)





Gemini Flash 2.5, 31.07.2025

- As expected, very realistic and symmetric
- Symmetry of the fillings show that's half



Stable Diffusion 3 Medium, 25.07.2025

- As expected, and somewhat realistic
- Symmetry is somewhat there between the piece on top and bottom
- Shape of the slices of bread can't be determined



**69. “Create a realistic image of a child playing on a playground swing.”**

Testing:

- Proper suspension points and chains/ropes
- Realistic action and child posture
- Scale of child, swing and playground

Expected:

- Child swinging, swing suspended from frame, feet above the ground
- Feet touching the ground if simply seated
- Playground in the background



GPT-4o, 06.07.2025

Notes:

- A happy child sitting on a swing with feet above the ground
- Swing is correctly rendered with joints in the chain
- Suspension can't be fully tested as it's just the chains and the child that the image has focused on
- Posture and placement of the child is correct
- Background looks like a playground
- Unrealistic component – only one foot visible



Gemini Flash 2.5, 31.07.2025

- As expected, and very realistic
- Shadows isn't there of the kid which is unrealistic



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic swing design, placement and shadows
- Swing is not firmly rooted in the ground
- Incomplete legs of the child is rendered
- The seat design is unrealistic for such a small child and will pose safety concerns
- There's a difference in colour of the grass where the kid is playing, and it gradually becomes darker and brighter towards the handrails

**70. “Create a realistic image of a person closing an umbrella before entering a building.”**

Testing:

- Umbrella orientation (tip down, handle up)
- Realistic scene progression (not open indoors)

Expected:

- Person just outside building entrance, umbrella partially or fully closed in hand



GPT-4o, 06.07.2025

Notes:

- A man clearly outside a brick-wall building in front of a door with reflections on the door
- Man holding the umbrella is unrealistic and incorrectly rendered as there's no handle on the umbrella
- Umbrella is disjoint as it has no handle, no tips, no proper ribs etc. (parts of the umbrella incorrectly generated)
- The frame of the door is incorrectly generated towards the lower end (no opening and frame boundaries visible in between the person's legs which was visible above, should have been visible)



Gemini Flash 2.5, 31.07.2025

- As expected, and very realistic
- Reflections on the road is too clear which is unrealistic
- Logically, why is the person closing the umbrella on the road when he could do it by the shelter, also poses safety concern



Stable Diffusion 3 Medium, 25.07.2025

- Doesn't adhere to the prompt, he's facing the wall; entrance door not in sight; the umbrella is clean; the handrail/black thing near the person is unrealistic; the shaft angle is unrealistic; the hand holding to the shaft is unrealistic

**71. “Create a realistic image of a person opening a window to let in fresh air.”**

Testing:

- Window panel and frame correctly aligned
- Realistic hand placement and window mechanism

Expected:

- Person opening a window open, with indication of fresh air (greenery outside, breeze)
- Opening of the window to be consistent with the design of the window door’s mechanism



GPT-4o, 06.07.2025

Notes:

- A man pulling open the door of a window with the handle turned down
- Greenery outside the window to indicate fresh air as expected
- Not clear is the position of the other door’s handle (looks closed)
- Frame of the window at the top is unclear
- Positioning of the other door which looks closed is far off the lower frame of the window (unrealistic containment)



Gemini Flash 2.5, 31.07.2025

- As expected, and very realistic



Stable Diffusion 3 Medium, 25.07.2025

- Very unrealistic rendering of the person opening a window
- The rendered action poses a safety concern as it looks like he's trying to get out through the window
- Reflection above the boy is unrealistic

**72. “Create a realistic image of a person looking into a bathroom mirror. Keep the faces generic.”**

Testing:

- Placement and visual coherence of reflection
- Consistency between real and reflected pose

Expected:

- Person facing a bathroom mirror, reflection matches pose



GPT-4o, 06.07.2025

Notes:

- A man inside a bathroom with a mirror placed above the bathroom sink and towel hanging beside
- Reflection of the person in the mirror is incorrect - posture, angle, eyes, background
- The towel ring has inconsistent ends
- Reflection of the person or nothing indicative of a person in front is visible in the tap



Gemini Flash 2.5, 31.07.2025

- As expected, and very realistic – including the reflections, shadows etc.
- Only unrealistic is the shower door looks incomplete and unrealistic



Stable Diffusion 3 Medium, 25.07.2025

- Very unrealistic rendering of the person looking at the bathroom mirror with incomplete rendering of a 2<sup>nd</sup> faucet
- Reflections aren't right
- The hands of the woman is unrealistic



### 73. “Create a realistic image of a person putting a key into a door lock.”

Testing:

- Accurate hand and key alignment with the door
- Lock and keyhole rendered distinctly and proportionally

Expected:

- Person holding a key, tip inserted or about to be inserted into the keyhole of a door lock



GPT-4o, 06.07.2025

Notes:

- A man standing in front of a door is correctly rendered
- Posture, body orientation and the look of the person is as expected realistically
- There's a key partially visible hanging down from the palm
- There's no key visible as being inserted to the keyhole, instead it's the thumb of the person inserted
- Position of the hole where the thumb is inserted is also incorrect



Gemini Flash 2.5, 31.07.2025

- As expected however the end of the key where the person is holding on is not a common design



Stable Diffusion 3 Medium, 25.07.2025

- Two hands, more than 3 keys, 2 keyholes which is unrealistic  
- The keys are all vertical, not “inserted” into the keyholes

**74. “I am placing my keys on the table. Please give me a picture of the keys.”**

Testing:

- Familiarity with common objects
- Proper orientation
- Correct spatial relationship of the objects

Expected: A set of keys (house keys or generic keys) lying flat on a table surface



GPT-4o, 06.07.2025

- Three keys and a keychain on a keyring placed on the table
- 1<sup>st</sup> and 2<sup>nd</sup> keys have some abnormality around the keyring hole



Gemini Flash 2.5, 31.07.2025

- 5 normal and 1 car key but none/some aren't connected to the keyring
- Shadow looks realistic



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic – common key design isn't there; some are blended into each other; keyring only partial

**75. “I now have a car. So, I have added my car key also to the keys. Please give me a picture of the keys on the table.”**

Testing:

- Familiarity with common objects
- Object addition
- Maintenance of context within the same conversation (ChatGPT)

Expected:

- A set of keys (house keys or generic keys) lying flat on a table surface (from previous conversation)
- Addition of a car key to the previous key stack



GPT-4o, 06.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- A key stack with car key and 2 generic keys placed on a wooden surface
- Wooden surface indicates table
- Car key is correctly rendered
- Two keys don't have any key ring (key ring disjoint here) – unrealistic
- Didn't maintain the conversation context- key chain missing, one key missing



Gemini Flash 2.5, 31.07.2025 (same conversation)

- 3 normal and 1 car key and all are connected to the keyring
- Shadow looks realistic
- Different texture of table
- Some of the keys from previous image is not there



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic – common key design isn't there; some are blended into each other; keyring only partial

**76. “I am peeling a banana in the kitchen. Please give me a picture of the banana.”**

Testing:

- Visualization of common food preparation and partial states
- Accurate depiction of the peeled portion (not fully peeled or unpeeled)

Expected:

- A banana partially peeled (skin open, fruit visible), on a kitchen surface or in hand



GPT-4o, 07.07.2025

Notes:

- A partially peeled banana on a marble surface that's likely to be a kitchen countertop
- Banana seems have been peeled from the bottom end
- Banana peel and the banana's texture have been realistically rendered
- Unrealistic component- the orientation in which the banana is standing, it needs a support but it's rendered as if it's freestanding in that position



Gemini Flash 2.5, 31.07.2025

- All peels are peeled partially so top portion of the banana is exposed
- Light and shadows are realistic
- Top of the banana that the person is holding is unrealistic – all peeled and the ends of the peels isn't what it is
- Tip portion of the banana is missing



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic – banana holding and peeling is done with same hand; the peel isn't coming from the banana visible; shadow isn't realistic
- Hand is incomplete/unrealistic



**77. “I am now cutting the banana into coins. Please give me a picture of the banana.”**

Testing:

- Representation of cut banana
- Correctness in shape/size
- Tracks transformation of an object through a sequence of actions (would the banana be peeled or not and then cutting)

Expected:

- Banana sliced into round pieces (“coins”), slices laid out on a cutting board or plate



GPT-4o, 07.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- A knife cutting through a banana with some small round pieces on a wooden cutting board
- Marble countertop similar pattern and color, same wall, with something black appliance in the background as the previously generated image
- Unrealistic component - cutting is happening with knife pointing away from the wall



Gemini Flash 2.5, 31.07.2025 (same conversation)

- Correctly understood what “coins” in this context means
- Cutting is done right (maybe with left hand)
- The residuals on the knife also look very realistic
- All the cut pieces are of same radius but different thicknesses



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic – banana is kept beside few gold coins of various sizes; coins symbol shows “B” symbol; didn’t understand the meaning of “cut into coins”; banana is not peeled

**78. “I am shopping at a grocery store and my trolley is filled with vegetables. Please give me a realistic picture of the trolley.”**

Testing:

- Tests recognition of retail settings, object containment, and quantity
- Variety in vegetables
- Plausible fullness

Expected:

- Grocery trolley/cart full of varied vegetables, within a store



GPT-4o, 07.07.2025

Notes:

- A grocery trolley with some vegetables (not really full) in a grocery store
- Indicates to be a grocery store as the context with retail shelves in the background
- Trolley is not overflowing with vegetables and only contain vegetables as per the prompt
- Variety of vegetables in the trolley – Cabbage, cucumbers, carrots, potatoes, bell peppers, broccolis, leek/spring onion, celery, lettuce, kale
- Yellow cucumber is an unrealistic component



Gemini Flash 2.5, 31.07.2025 (same conversation)

- Trolley is filled vegetables of various kinds, it's full, no space left for anything else
- All are vegetables
- The background shows a grocery store
- Unrealistic are the shelves in the background, the height changes randomly



Stable Diffusion 3 Medium, 25.07.2025

- Many vegetables but some blended into each other inside the trolley
- 5 wheels for the trolley
- As a whole background looks somewhat realistic yet unrealistic - it contains baskets of vegetables ready to be picked but the tomato in the front line is blended to become a long line same is with the red bell peppers

**79. “I am now carrying the vegetables in a grocery bag in my hand. Please give me a realistic picture of the bag.”**

Testing:

- Correct depiction of weight, hand grip
- Realistic bag shape/proportion
- Ability to maintain context of the conversation

Expected:

- Bag held in hand, visibly bulging or filled with vegetables



GPT-4o, 07.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- A man holding a cloth tote bag filled with vegetables in a grocery store
- Indicates to be a grocery store as the context with retail shelves in the background
- Was able to maintain the context of the conversation from previous prompt – the vegetables variety seem to be like those in the trolley previously
- The shape of the tote bag also looks realistic
- Only the placement of broccoli looks unrealistic- straight and balanced- even though it's a heap of vegetables likely inside the bag



Gemini Flash 2.5, 31.07.2025 (same conversation)

- Grocery bag is filled with vegetables, background looks like entrance/exit of a store/mall
- Some of the vegetables (broccoli etc.) that were in the trolley are not here, some new variety is also seen here (cherry tomato, beans, aubergine, carrots with leaves)
- Holding of the grocery bag is realistic but not realistic if the bag is heavy



Stable Diffusion 3 Medium, 25.07.2025

- Few vegetables, still in middle of the grocery store aisle with the grocery bag
- The bag and the holding are unrealistic



**80. “I have now reached home and I'm placing the groceries into the refrigerator. Please give me a realistic picture of the fridge.”**

Testing:

- Food items are plausible, properly arranged, not floating or misplaced
- Understanding of storage and spatial arrangement
- Ability to maintain context of the conversation

Expected:

- Refrigerator with doors open, vegetables and other groceries either held by a person or inside on shelves



GPT-4o, 07.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- An open refrigerator with vegetables and few other items across multiple shelves
- Vegetables that were previously seen in the trolley are somewhat still maintained through the conversation
- Vegetables – Cucumbers, broccolis, red bell peppers, carrots, yellow cucumber, cabbage
- However, there some vegetables missing that were seen in the trolley and the bag
- Missing vegetables – lettuce, leek, celery, potatoes
- Potatoes however also could be the diced ones in a bowl in the top shelf
- Additional items like the bright orange juice (likely to be mango) and another bottle on the middle shelf, zucchinis, another red bell pepper, and more leafy vegetables in top shelf
- No light inside the refrigerator which is ideally present when all working refrigerator's door is open



Gemini Flash 2.5, 31.07.2025

- Realistic as expected, can't make out the contents of the bag
- Fridge already has some contents that includes milk and maybe some juice dark orange/red colour



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic – no vegetables visible; the fridge itself is unrealistic with two doors for a realistically single door fridge; doesn't follow containment



**81. “I am holding a soccer ball on the field. Please give me a picture of the ball.”**

Testing:

- Realistic ball depiction (size, markings)
- Plausible interaction with human hand
- Ability to maintain context of the conversation (ChatGPT)

Expected:

- A soccer ball held in hand, background as a grassy field or pitch



GPT-4o, 07.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- A man dressed in soccer socks, trousers, t-shirt holding a soccer ball
- Background of grassy field
- The skin color of the person has changed from the previous prompt of holding a grocery bag of vegetables



Gemini Flash 2.5, 31.07.2025

- The football and the hand look realistic, however the goal posts in the background doesn't look logically realistic
- Also doesn't show "holding", how is the ball supported in this position?



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic – fingers, scale of the person and the background (legs against body)
- Shadows also incorrect
- Does show "holding"

**82. “I have now finished playing on the field and have reached back home. Opening the fridge, I now take only a mango juice from the refrigerator. Other items are untouched. Please show a picture of the refrigerator now.”**

Testing:

- Memory of prior contents (ChatGPT)
- Ability to maintain context of the conversation (ChatGPT)
- Accurate reflection of a single object removed

Expected:

- Person holding a bottle of mango juice/ with orange-coloured liquid with the refrigerator in the background
- If the door of the refrigerator is still open, all other items previously shown, should remain as it is



GPT-4o, 07.07.2025 (Continuation of the conversation from previous prompt)

Notes:

- A refrigerator with its door open
- There are some vegetables inside the refrigerator, however, the contents have changed from the previously generated image – for example there was a bowl of diced potatoes in the top shelf, now in its place is whole potatoes. Others are eggs, milk (white bottle), tomatoes, beetroots, cucumbers, carrots, cabbage, bell peppers (green now instead of red).
- A bottle with label “mango juice” is placed before the refrigerator (to indicate taking out)
- Bottle and its cap’s shapes have changed
- Vegetable tray is missing, there’s just vegetables instead in its place
- Logically looks like there’s two people – one holding the bottle and the other just standing in front of the refrigerator



Gemini Flash 2.5, 31.07.2025 (Same conversation)

- Design of the fridge has changes (metal in the middle) but is still realistic
- Fruits previously not there are also in the fridge
- Follows the prompt “mango juice” is taken out but the fridge didn’t have it in the previous prompts’ images



Stable Diffusion 3 Medium, 25.07.2025

- The empty freezer section is open but freezer shouldn’t have shelves on the door
- Doesn’t follow the prompt

**83. “I am pouring the juice from the bottle into a glass. Please give me a picture of the glass.”**

Testing:

- Realistic liquid flow
- Plausible positioning
- Containment of juice
- Ability to maintain context of the conversation (ChatGPT)

Expected:

- Glass with mango juice being poured in, or partially filled
- Previously shown bottle visible overhead or to the side



GPT-4o, 07.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- A glass sitting on a likely kitchen countertop
- The glass is partially filled with mango juice
- Mango juice is flowing down from the bottle
- The label on the bottle doesn't look like English or incorrigible (MIUNCO instead of MANGO)
- Shape and size of the bottle has changed again
- Positioning and posture of the person holding the bottle is realistic



Gemini Flash 2.5, 31.07.2025 (Same conversation)

- Bottle has changed
- The liquid is flowing down from the bottle to the glass and realistic flow
- The liquid seems to be thin when flowing but is thick in the bottle and thicker in the glass
- Shadow incorrect – shadow of glass but no shadow of bottle



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic 1<sup>st</sup> image - bottle design; the liquid flowing and one in glass are of different colors; there isn't any liquid in the bottle so the flow is unrealistic
- Unrealistic 2<sup>nd</sup> image (specified "mango" in the prompt) – thickness is different in glass, bottle and the flow; no liquid in the neck of bottle but there's thick flow

**84. “I am putting a steel straw into the glass. Please give me a picture of the glass.”**

Testing:

- Proper placement
- Correct depiction of material and scale (steel straw vs. glass)
- Incremental action and effect
- Ability to maintain context of the conversation

Expected:

- Glass with mango juice, a steel straw inserted or being inserted



GPT-4o, 07.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- A glass of mango juice on the kitchen countertop
- Maintained from the previous prompt's image – the countertop surface, the design of the glass, the amount of juice
- A steel straw is inserted in the glass
- Unrealistic component – the straw is not resting or making any visible contact with the glass



Gemini Flash 2.5, 31.07.2025 (Same conversation)

- Glass is half-filled with mango juice there's some residue on the sides of glass but it's perfect height across which is unrealistic
- Glass shape and height changed
- Steel straw is held between the fingers



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic – no shadow and fingers



**85. “I have finished drinking the juice and am placing the glass with straw into a sink basin for washing. Please give me a picture of the sink basin.”**

Testing:

- Plausible arrangement
- Correct sink context
- Objects at appropriate orientation
- Follows object lifecycle through use and placement in new environment
- Ability to maintain context of the conversation (ChatGPT)

Expected:

- Sink basin with empty glass and straw inside
- Water may be present on the sink basin



GPT-4o, 07.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- A glass with some residual mango juice and steel straw placed in an empty sink
- The steel sink is clean and empty with multiple types of textures to indicate maybe it's over the time use
- The glass has some textures in the insides to indicate the presence of mango juice previously
- The steel straw is now resting on the glass
- Unrealistic component - the straw doesn't have the texture that the glass has



Gemini Flash 2.5, 31.07.2025 (Same conversation)

- Glass with a steel straw in a kitchen sink
- Glass shape and height changed; colour of juice changed; lustre of the straw changed
- Straw has contact with the glass
- Droplets in the glass look unrealistic



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic – glass is broken; the glass still has content; the content is thick like milkshake or ice-cream
- Steel straw is resting on the glass

**86. “I am picking up an apple from a fruit bowl. Please give me a picture of the apple.”**

Testing:

- Object isolation
- Context understanding (apple vs. bowl)

Expected:

- Apple in hand (lifting from a bowl), or apple held with bowl in background
- Bowl to have few fruit



GPT-4o, 07.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- A realistic depiction of apple held in a hand lifting from a bowl of fruits
- Maintained the same kitchen countertop marble-like surface as previous images
- The color, texture etc. of the apple looks realistic
- The apple held is the only one bright red in color and has a stem, the other apples in pale red or not fully red and doesn't have a stem (not unrealistic but an observation)
- The fruits- 3 apples (one in hand), 1 orange and 3-4 bananas
- Odd component is the last banana – there's two stems but at the bottom it looks like one single banana



Gemini Flash 2.5, 31.07.2025 (Same conversation)

- Follows prompt correctly – apple picked, fruit basket in background
- Reflection and shadows look realistic
- Basket contains 4 oranges and 2 bananas
- Apple is held in their right hand



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic – incomplete rendering of cat and hand
- Doesn't show "picked"
- Can't make out the content of the white bowl beside the fruit bowl

**87. “I have now cut all the fruits from the fruit basket. Please give me a picture of the cutting board and the fruit basket.”**

Testing:

- Cutting results visible
- No uncut fruit left in basket
- Quantity and spatial reasoning
- Memory of prior contents (ChatGPT)

Expected:

- Full fruit bowl now empty somewhere near it
- Cutting board crowded with sliced fruit



GPT-4o, 07.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- A bowl and cutting board with cut fruits on a marble-like surface
- The bowl is same as the previously generated image
- The countertop is sustained with same pattern, color etc. throughout the conversation
- The cut fruits are of exact variety that was generated in the previous prompt – apples, banana, and orange
- Quantity of the cut fruits doesn't match the quantity of whole fruits we saw earlier – lesser cut bananas and apples
- Apple skin color point to only one of the apples that was previously in the bowl



Gemini Flash 2.5, 31.07.2025 (Same conversation)

- Follows prompt correctly – all fruits cut
- Cutting board also has 4 apple seeds (realistic and logical as only one apple was to be cut)
- Orange slices don't look like that of 4 oranges
- Bowl was previously white in colour, now it's transparent



Stable Diffusion 3 Medium, 25.07.2025

- Unrealistic – knife shape; cut fruit is unrecognizable; shadow
- All fruits are not cut

**88. “Generate an image of a potted plant sitting on a windowsill.”**

Testing:

- Proper scale
- Daylight context
- Object placement, support and stability

Expected:

- Potted plant stably positioned on top of a windowsill



GPT-4o, 07.07.2025

Notes:

- A healthy plant in a pot sitting on top of a windowsill
- Three branches stemming out from the pot
- There's sunlight falling on the pot creating a shadow
- Background outside the window also looks realistic





Gemini Flash 2.5, 31.07.2025 (Same conversation)

- A potted plant sits not too far and not too close to the window on the windowsill
- The window glass has some dust and there's a reflection of just the plant on the glass
- The shadow isn't realistic



Stable Diffusion 3 Medium, 26.07.2025

- Multiple potted plants by the windowsill
- Shadows aren't realistic
- Effect of sunshine on plants is realistic



**89. “A cat was walking on the windowsill, and the potted plant fell down. Please give me a picture of the potted plant.”**

Testing:

- Gravity effect and result of fall (broken pot, spilled soil)
- Dynamic process, consequence, and final state after event

Expected:

- Potted plant on the floor, possibly damaged/spilled, windowsill visible above
- Cat somewhere nearby or parts visible



GPT-4o, 07.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- A cat's tail is visible, a pot spilled over with mud lying around on the windowsill is rendered
- The pot and plant look smaller compared to the cat and windowsill
- The plant is not similar to the plant generated previously
- Pot saucer is missing here
- The position in which the plant is lying means the plant was closer to the window, however the position was different from the previous image
- Plant here has a single stem, with no major branches, whereas in the previous one, there were three branches from the soil, instead of a single stem



Gemini Flash 2.5, 31.07.2025 (Same conversation)

- The pot is broken, some parts are scattered on the floor
- The pot contains small plant still intact without any damage
- The mud and the broken section look unrealistic
- Cat looks realistic
- Doesn't look its windowsill beside but a balcony/door and the cat and pot is on the floor
- If the pot fell from the windowsill to the floor, the damage and the upright position is unrealistic
- The type of plant has remained



Stable Diffusion 3 Medium, 26.07.2025

- Unrealistic – pot and cat are floating; doesn't show “fell” and its effect

**90. “Show a picture of a bowl filled with cereal and milk on a breakfast table.”**

Testing:

- Containment
- Plausible object relations and arrangement
- Food depiction

Expected:

- Bowl with cereal and milk, on a table set for breakfast (may show spoon, placemat)



GPT-4o, 07.07.2025

Notes:

- A wooden table surface with a bowl of cereal and milk, spoon, placemat and a glass of milk
- Chair in the background partially visible suggesting it's a dining table
- Cereals in the bowl look realistic – heap of cereals, some floating, some partially submerged
- All cereals look crispy and fresh
- Some stain on the table is seen, even though the background is blurred out



Gemini Flash 2.5, 02.08.2025

- A bowl of milk and cereals floating on top
- Swirl in the middle that's like a boundary is unrealistic



Stable Diffusion 3 Medium, 26.07.2025

- A bowl of cereals on a table with a bottle of milk on the side
- The shape of spoons is unrealistic
- The milk inside the bottle is unrealistic (funnel shape or swirl inside)
- Reflection/shadow doesn't look realistic

**91. “I am adding sugar to the cereal bowl. Please give me a picture of the cereal bowl on the breakfast table.”**

Testing:

- Plausibility of changes to food surface/contents post-action
- Plausibility of size of sugar against cereals
- Plausibility of reaction of sugar, cereal and milk (no piling)
- Memory of previous arrangement

Expected:

- Sugar being sprinkled or freshly sprinkled on top of cereal bowl
- Spoon or hand visible likely to show sprinkling



GPT-4o, 07.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- Most of the elements – cereal, glass of milk, placemat, spoon, chair in the background, color of the table has remained same as the image generated for the previous prompt
- Followed the instruction and generated image as expected- spoon of sugar is sprinkled
- Color of the cereal has now changed from dark orange to a pale orange
- The table no longer has stains
- The chair seems to have moved slightly from the extreme end of the table of the table to slightly inwards from the previous one
- The gap between the rods in the chair also seems to have increased from the previous one



Gemini Flash 2.5, 02.08.2025 (Same conversation)

- A bowl of milk and cereals floating on top, sugar falling down from a spoon
- Table, bowl, cereals, and quantity of milk has changed from previous image
- The sugar is going down correctly but the middle portion is suddenly without any cereals and has a ripple, the sugar on the milk also seems to be floating (unrealistic)



Stable Diffusion 3 Medium, 26.07.2025

- No milk in the bowl
- The cereals don't look realistic
- The heap of sugar in the bowl look like a big foam/cotton
- Hands are incomplete, but the holding of the sugar bowl looks realistic

**92. “I am adding more cereal to the bowl. Please give me a picture of the cereal bowl on the breakfast table.”**

Testing:

- Containment of contents with the increase (no overflow)
- Size visualisation (sugar vs cereals)
- Memory of previous arrangement (ChatGPT)

Expected:

- Bowl now more filled or cereal being poured, likely both cereal and sugar still visible



GPT-4o, 07.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- Most of the elements – cereal, glass of milk, placemat, spoon, chair in the background, color of the table has remained same as the image generated for the previous prompt
- Followed the instruction and generated image as expected- cereals being added to the bowl
- Cereals is now a heap, not unrealistic
- Sugar crystals are no longer visible against the cereals
- No unrealistic overflow due to the action
- Color of the cereal has changed again
- Chair is also changed (smaller and fewer rods)

*[was able to maintain context for this one and follow commonsense]*





Gemini Flash 2.5, 02.08.2025 (Same conversation)

- A bowl of milk and cereals floating on top, cereals falling down from the box of cereals to the bowl
- Table, bowl, cereals, and quantity of milk has changed from previous image



Stable Diffusion 3 Medium, 26.07.2025

- No milk in the bowl
- The cereals don't look realistic
- Hand/fingers are incomplete



### 93. “Create a realistic image of a person watering a plant with a watering can.”

Testing:

- Action plausibility and liquidity
- Hand-object interaction
- Gravity effects - liquid flow, plant growth direction

Expected:

- Person holding a watering can, water poured onto soil of potted/bedded plant



GPT-4o, 08.07.2025

Notes:

- A man holding a watering can, watering a potted plant
- Stem is yellow in color, brown only at the bottom
- Water flowing the leaf (unrealistic)
- No impact of water on the leaf (unrealistic effect of water)
- The leaves all are overly glossy (unnatural) and of same color (no yellow/brown leaves)
- Holes on the watering can isn't uniform, i.e., it is random which is unrealistic for a watering can
- The watering can's handle isn't well defined – only one end visible – unrealistic design as a person can't hold on to a metallic watering that can carry such an amount of water (so the water can is heavy) with a handle of that design where there's no support for the user
- There's an unnatural symmetry to the plant
- However, the plant follows gravity effects – leaves pointing upwards, growth of the plant upwards, water flowing down, plant firmly rooted in soil



Gemini Flash 2.5, 02.08.2025

- A potted plant is being watered using a watering can
- Water flowing on the look realistic
- Mud is wet and there's some depth at the point of water flow
- The watering is unrealistically small for the water it can hold (no functionality except aesthetics)
- Light effect on plant's leaves and can is realistic



Stable Diffusion 3 Medium, 26.07.2025

- Holding of water can and its design is unrealistic
- No water

**94. “The plant has been overwatered. Show the picture of the plant.”**

Testing:

- Recognition of excess water effects with realistic wetness
- Understanding of plant health and signs of stress/rot
- Pot not overflowing unnaturally
- Gravity effects - liquid flow, plant growth direction
- Memory of previous image and arrangement

Expected:

- Plant with water-logged soil, drooping leaves, and some water pooling
- Plant same as previous



GPT-4o, 08.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- Potted plant with signs of overwatering with the drooping and dark leaves, as well as the dark and drooping stem
- Water in the pot (reflections as well) as well as around the pot
- Soil with water is slightly darker in color, otherwise soil looks dry
- Pot shows signs of overflowing in the body (water streaks) and at the base (darker in color) but not in the brim. No splashes or mud on the pot, unusually clean
- All leaves are of same color (no yellow/brown leaves)
- All leaves size comes down with the height and the leaves at each interval are of almost equal sizes
- Lighting has changed- before it was sunny now it's gloomy
- There's an unnatural symmetry to the plant
- Pot has changed (there's a line at the brim)



Gemini Flash 2.5, 02.08.2025 (Same conversation)

- A potted plant is overwatered and plant is dying from the black areas and drooping leaves
- Water logging and the wet mud look realistic
- Plant, pot and mud has changed
- The black colour at the bottom of the pot is unrealistic to show the effect of overwatering
- Light effect on plant's leaves is realistic however the shadows are not (just the pot)



Stable Diffusion 3 Medium, 26.07.2025

- Healthy potted plant
- No signs of overwatering
- Leaves look unrealistic – different veins, all sap green, no signs of pest or anything

**95. “The plant above looks like it's captured at least a day or two later. However, show the picture of the plant immediately after it was watered.”**

Testing:

- Temporal reasoning - immediate post-watering, not delayed effects

Expected:

- Plant standing upright
- Plant exactly same as first image with no drooping
- Soil with some water pooling
- Signs of overwatering and overflowing as the previous prompt clearly said so
- Memory of previous images and arrangement



GPT-4o, 08.07.2025 (*Continuation of the conversation from previous prompt*)

Notes:

- Potted plant with signs of overwatering with the slightly drooping leaves, as well as the dark green stem
- More water in the pot as well as around the pot (as expected) but the pot is not moist (unrealistic)
- Symmetry in the water pool around the pot (unrealistic)
- Pot shows only one (previously five) water streak (unrealistic as it doesn't explain the overflowing) and there's still no splashes or mud on the pot, unusually clean
- All leaves are of same color (no yellow/brown leaves)
- All leaves size comes down with the height and the leaves are all almost equal sizes
- Lighting has changed- before it was gloomy now it's slightly warmer
- There's an unnatural symmetry to the plant
- Pot looks like what it would have been in the first image (no line at the brim)
- Plant is smaller
- Drooping leaves but it was just watered (wouldn't expect an immediate reaction)



- Leaf veins are different



Gemini Flash 2.5, 02.08.2025 (Same conversation)

- A potted plant is just watered
- Water logging, the wet mud and the pot look realistic
- Plant might be the same looking at the stem
- Might be overwatered, can't say for sure



Stable Diffusion 3 Medium, 26.07.2025

- Healthy potted plant
- No signs of overwatering
- Leaves look unrealistic – different veins, all sap green, no signs of pest or anything

**96. “Create a realistic image of the sunny sky after a light rain.”**

Testing:

- Knowledge of weather sequence and atmospheric phenomena

Expected:

- Bright sky with some clouds, sunlight, and faint rainbow
- Ground with small puddles or just wet road



GPT-4o, 08.07.2025

Notes:

- A blue sky filled with clouds towards lower half of the image
- Unrealistic placement and size of the clouds- placement there's no cloud around the sun and the top half of the image; clouds are all similar size
- Unrealistic rain flow and angle
- The sun is bright and unobscured
- Unrealistic sun rays
- With this level of clear, intense sun, a uniformly heavy rain is rare (clouds would have obscured the sun or a rainbow might have been present)
- Raindrops are visible as white streaks, more like stylized motion blur than actual raindrops



Gemini Flash 2.5, 02.08.2025

- An open field with some grass in between the wet road is realistic
- Water puddles in multiple areas of the road
- Bright sky with unrealistic clouds (clouds doesn't look like it just rained)
- Reflection of the sky in the puddles look unrealistic



Stable Diffusion 3 Medium, 26.07.2025

- Bright sky with sun and clouds, green tall trees, droplets
- The image generated as if it's a view from a window or a glass surface as the droplets look similar to the ones on the window after a shower



**97. “Create a realistic image of a sandstorm in a city with tall skyscrapers.”**

Testing:

- Depiction of extreme weather and its effects on visibility
- Understanding of atmospheric density
- Realistic sand motion

Expected:

- Dusty atmosphere obscuring skyscrapers, sand particles in air, altered colours or faded outlines



GPT-4o, 06.07.2025

Notes:

- A dramatic image of a sandstorm where the skyscrapers are standing tall in a desert
- The skyscrapers and the haze around it look realistic
- The sand cloud in the foreground is highly stylized, with smooth, fluffy billows more typical of fog or smoke than heavy, granular sand
- The ground also is filled with sand completely as if to look like a desert but is unrealistic in a such city background with tall skyscrapers
- The color is perfectly even and uniform
- Sand appears to roll in thick waves but doesn't interact naturally with building bases (no visible accumulation or piled drifts)
- Except for the small accumulation of the buildings, there's nothing to indicate a city
- Reflections on the glass skyscrapers is unrealistic



Gemini Flash 2.5, 02.08.2025

- Tall skyscrapers, cityscape, dust and haziness as expected
- Unrealistic is the stones/black particles



Stable Diffusion 3 Medium, 26.07.2025

- Even though there are tall skyscrapers, the landscape isn't that of a cityscape rather a desert
- Such a huge desert close to human settlements is unrealistic given the current city planning advancements (and logically not possible for construction of a building without amenities and comfort around)
- The haziness though looks very realistic
- The sandstorm is unrealistic

**98. “Create a realistic image of a person tying their shoelaces while sitting on a bench.”**

Testing:

- Human posture
- Action depiction
- Accurate interaction with shoe, bench and laces
- Correct rendering of shoes

Expected:

- Person bent forward on bench, hands near shoe, laces partly tied, feet on ground



GPT-4o, 06.07.2025

Notes:

- Person bent forward on bench, hands near shoe
- One foot on the ground, while one is raised with the help of a small stool (legs look realistic and consistent with the position of the feet)
- Hands holding the shoelace look realistic
- The laces are partially inserted in the holes in both the shoes
- Position and angle of shoes is realistic
- Laces break at the holes of both the shoes (ends are missing, overlaps between the crisscrosses, not sure where the lace in the hands is coming from)



Gemini Flash 2.5, 02.08.2025

- Landscape in the background is of a park
- The person has placed their feet on top of a bench (similar to [this](#) might their posture)
- The end of the thread is unrealistic as it's suddenly too thin
- Shadow is unrealistic



Stable Diffusion 3 Medium, 26.07.2025

- Person sitting on a bench with a thread in their hands
- Shoes are already tied so the thread in hand is unrealistic
- Shadow is unrealistic

**99. “Create a realistic image of a person braking hard with the front brake while cycling fast on a straight, smooth road.”**

Testing:

- Representation of the immediate effect of suddenly applying the front brake—potential for tipping, shifted centre of mass, realistic reaction of the bicycle and cyclist
- Concept of “braking hard” with realistic body and cycle movement
- Details of the scenario
- Avoidance of Errors (unrealistic braking, person sitting upright and unaffected by the sudden deceleration)

Expected:

- Person riding a bicycle on a straight, smooth road, actively using the front brake lever
- Person may be pitched forward or in the early stages of tipping, as braking hard with the front brake at high speed typically causes the bike’s rear to lift or the rider to lunge forward
- The cycle on the road, the rear wheel maybe slightly off the ground



GPT-4o, 08.07.2025

Notes:

- The cyclist’s bent-forward position, grip, and attire align with real sports cycling
- Road and Grass: The background and road textures are reasonably natural.
- The dust at the front wheel is pronounced but doesn’t match the motion and the location
- Shadows are soft, lack direction, and aren’t fully formed to reflect the cyclist
- Face is unrealistic
- Bike handle is not of equal lengths on either side
- No mudguards visible, behind tyre smaller than the front one
- Some spokes are intertwined and disjoint
- Left hand only 3 fingers
- Left forearm is inconsistent with the size of wrist

- Size of the bike for the cyclist is unrealistic for a smooth biking experience
- The rod between the handlebar and the head tube is at an angle which is unrealistic to the angle of the cycle
- The cyclist however is not applying the brake (hands are just holding the handlebars)



Gemini Flash 2.5, 02.08.2025

- A cyclist on a bicycle on a smooth road
- Parts of cycle is incorrect – the extra items on back, the handlebar's symmetry, the elevated top and bottom tube, and the elevated front tyre (floating)
- Cycle isn't of the right size for the person and the legs doesn't look like cycling
- The cyclist isn't braking, hands not on either of the brakes
- The front tyre elevated might have been picked up from "front brake"
- Unrealistic shadows
- Cycling fast is indicated through the blurry background and wheel spikes



Stable Diffusion 3 Medium, 26.07.2025

- Person sitting on a bicycle
- Incomplete hand, leg, handlebar, top tube, bottom tube, pedals etc.
- Unrealistic – spikes on the wheels, seat in place of handlebar, handlebar in front of the wheel, floating (no contact between the wheels and the road)
- Blurry background
- Posture looks realistic
- Shadow is incorrect