



Stage 1 Filtering PIQA Scenarios

Input: Correct solutions from PIQA

Gemini LLM: Filter for Secondary Tool Usage and Affordances

Output:
Filtered Correct Solutions (Highlighting tool use/affordance)

Filtered Correct Solution



To separate egg whites from the yolk using a water bottle, you should: Squeeze the water bottle and press it against the yolk. Release, which creates suction and lifts the yolk.



Stage 2 Base Prompt Generation

Input: Filtered Correct Solutions

Gemini LLM: Convert solutions into Visual Video Prompts (Short, descriptive scenes)

Output:
383 Base Video Generation Prompts

Video Generation Prompt

A person squeezes a plastic water bottle and presses the opening against an egg yolk in a bowl of egg whites. They release the squeeze, and the yolk is sucked into the bottle.



Stage 3 Prompt Upsampling

Input: Base Video Generation Prompts

Gemini LLM: Enrich prompts with

- Detailed Physical Structure
- Object Affordance Logic
- Preserve original scene scope

Output:
383 Upsampled Video Generation Prompts

Upsampled Prompt

A person firmly squeezes a translucent, pliable plastic water bottle, reducing its internal volume. They carefully press the bottle's circular opening against the fragile, spherical egg yolk, which sits surrounded by viscous, clear egg whites in a ceramic bowl. The bottle's rim makes full contact with the yolk's surface. The person slowly releases their grip on the bottle, creating suction inside the bottle. The pressure difference causes the yolk to deform slightly as it is drawn upwards into the bottle's opening.



Stage 4 Video Generation

Input: Both Base and Upsampled Video Generation Prompts

Models: A wide range of open and closed-source video models (CogVideoX, Cosmos, HunyuanVideo, LTX, MAGI, VideoCrafter2, WanAi 2.1, Sora, Veo-2)

Output:
For each model, 766 videos were generated, targeting tool usage and affordance-based physical quality.

Generated Video



* Video above was generated using Cosmos-Predict1-14B-Text2World