

1. System Prompt for Edit Instruction Generation

To automatically generate edit instructions for each image, we employed Gemini-2.5-Flash with a carefully designed system prompt that guides the model to behave as a professional photo-editing prompt writer. The model is instructed to produce natural editing instructions that reflect plausible user intents.

The following system-level instruction was provided to Gemini 2.5 Flash:

System Prompt:

You are an expert photo editor prompt writer.

Given an image, write ONE concise, natural language instruction that a user might give to an image-editing model.

The instruction MUST be grounded in the visible content (objects, colors, positions) and be closely related to the image content.

Output Format

Return a JSON object with a "prompts" array of photorealistic prompts.

Example Output structure

```
{
  "prompts": [
    "<first prompt<",
    "<second prompt<"
  ]
}
```

2. System Prompt used by Gemini-2.5-Pro as a Judge

The system prompt we used to control editing quality is provided as follows:

System Prompt:

You are a professional image quality evaluator specializing in image editing assessment.

Your task is to evaluate edited images by analyzing the following items in sequence:

1. **Edited Image:** The final edited result (primary evaluation target)
2. **Input Image(s):** One or more reference images used for the edit operation (1–N images)
3. **Editing Instruction:** The specific editing prompt or instruction used

Multi-Image Evaluation Context: You will receive the edited result image first, followed by one or more input images (the reference images used for editing), and finally the editing instruction. Use all of these to make your assessment.

Evaluation Criteria (Weighted Scoring for Image Editing):

- **Edit Instruction Compliance (40% weight):** Does the edited image fulfill the specific instruction? Are the requested changes clearly visible and properly implemented? Does the result match the intended edit?
- **Editing Quality & Seamlessness (25% weight):** Are the edits natural and realistic? Are there visible artifacts, inconsistencies, or blending issues? Is lighting and perspective preserved?
- **Preservation vs. Change Balance (20% weight):** Are appropriate elements from the original preserved? Are unrelated regions unaffected? Is the editing focused and not overly destructive?
- **Technical Quality (15% weight):** Overall sharpness, color consistency, exposure, and absence of artifacts or distortions.

Comparative Analysis: Compare the edited result against the original image to assess:

- What changes were successfully made
- What elements were properly preserved
- Whether the instruction was accurately interpreted

Scoring: Provide a final weighted score from 0.0 to 1.0 based on the evaluation criteria above. The pipeline will automatically compare this score against a strictness threshold.

A. Editing Examples by Category

This appendix showcases examples from the Pico-Banana-400K dataset. Each figure illustrates how the model performs specific transformations such as object addition, removal, style change, or contextual outpainting.



Edit Type: Pose tweak (minor plausible change)

Edit Instruction: Slightly adjust the man's right arm and hand to show him lifting the coffee cup towards his mouth, as if taking a sip, while maintaining the newspaper's current position and the overall lighting, brushwork, and texture of the original painting.

Figure 1. **Pose tweak.** Adjusting the pose or orientation of humans.



Edit Type: Convert person to 2D anime/manga style (identity-preserving)

Edit Instruction: Transform the person in the image into a 2D anime/manga style, preserving their identity by maintaining facial features, glasses, and beard, rendering them with vibrant, cel-shaded colors, distinct black outlines, simplified textures, expressive anime-style eyes, and a clear, stylized 'MIKE MOZART' beanie, while keeping the background and the 'BRACHS Red Velvet CANDY CORN' bag in their original photographic style.

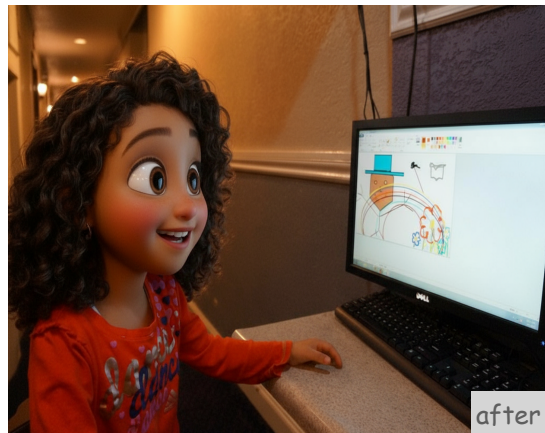
Figure 2. **Convert person to 2D anime/manga style.** Rendering human in anime style.



Edit Type: Simpsonize the person (yellow-skin cartoon style)

Edit Instruction: Transform the person in the image into a character from 'The Simpsons' by giving them a vibrant yellow skin tone, large, simple eyes, and a broad, friendly smile, while rendering their plaid shirt and hat in a simplified, cel-shaded cartoon style with clean lines and flat colors, ensuring they integrate seamlessly into the realistic outdoor background.

Figure 3. **Simpsonize the person.** Rendering human in “Simpsonized” style.



Edit Type: Convert person to Pixar/Disney-like 3D cartoon look

Edit Instruction: Transform the girl into a Pixar/Disney-style 3D animated character, giving her larger, expressive eyes with a joyful sparkle, smooth and slightly simplified skin with a warm glow, more defined and voluminous curly hair, and a cheerful, exaggerated smile, while maintaining her original pose, orange shirt with simplified textures, and seamless integration with the existing background lighting and computer screen.

Figure 4. **Convert person to Pixar/Disney-like 3D cartoon look.** Rendering human in “Pixar” style.



Edit Type: Add a new object to the scene

Edit Instruction: Add a tall, ornate vintage wrought-iron lamppost with a warm, glowing incandescent light source to the left side of the main entrance, positioned slightly behind the front row of bushes, ensuring its light realistically illuminates the foreground and casts soft shadows consistent with the building's existing nighttime illumination.

Figure 5. **Add a new object to the scene.** Adding new objects or elements seamlessly into the scene while preserving lighting and composition.



Edit Type: Change an object's attribute (e.g., color/material)

Edit Instruction: Change the white metal body and ornate details of the foreground lamppost to a dark, aged bronze finish, ensuring the new material realistically reflects the ambient light and maintains a cohesive aesthetic with the surrounding architecture, while preserving the existing yellow glass panels and the overall scene's lighting.

Figure 6. **Change an object's attribute.** Changing specific attributes such as color, texture, or material of an object.



Edit Type: Outpainting (extend canvas beyond boundaries)

Edit Instruction: Extend the canvas of the image outwards in all directions, seamlessly continuing the textured, light-colored stone wall, the dark metallic grid pattern of the window grates, and the vibrant green foliage with dappled sunlight, ensuring consistent warm lighting and natural shadows to maintain the overall realistic aesthetic.

Figure 7. **Outpainting.** Expanding the canvas to reveal more of the environment beyond the original image boundaries.



Edit Type: Change the size/shape/orientation of an object

Edit Instruction: Redraw the main flag to appear significantly larger and more dramatically unfurled, with its fabric catching more sunlight to distinctly highlight the yellow, black, and red sections and their heraldic lion patterns, while maintaining the strong backlighting and the silhouetted buildings in the background.

Figure 8. **Change the size/shape/orientation of an object.** Altering geometry, scale, or proportions of objects while maintaining realistic structure.



Edit Type: Remove an existing object

Edit Instruction: Seamlessly remove the vertical metal pole on the far right side of the image, ensuring the background elements like the seats, floor, and window view are realistically in-painted with matching textures, colors, and the existing warm ambient lighting, to create an unobstructed view of the train interior.

Figure 9. **Remove an existing object.** Removing undesired objects or distractions while filling in background details naturally.



Edit Type: Strong artistic style transfer (e.g., Van Gogh/anime/etc.)

Edit Instruction: Apply a vibrant, post-impressionistic style reminiscent of Van Gogh, transforming the entire image with swirling, visible brushstrokes and exaggerated, luminous colors, particularly in the sky, water, and distant cityscape, while rendering the boat's interior and exterior surfaces with a textured, impasto-like finish.

Figure 10. **Strong artistic style transfer.** Converting real-world photos into artistic renditions or stylized interpretations.



Edit Type: Apply seasonal transformation

Edit Instruction: Transform the summer scene into a serene winter landscape by replacing the lush green foliage with snow-covered branches and bare trees, covering the brown ground with a fresh blanket of pristine white snow, and adjusting the overall lighting to reflect a cold, soft, and diffused overcast winter day, while ensuring the green sign and its white arrow and telephone symbols remain clearly visible, potentially with a subtle dusting of snow on its top edge.

Figure 11. **Apply seasonal transformation.** Changing seasonal context, e.g., transforming summer landscapes into snowy winter scenes.



Edit Type: Photo to cartoon/sketch/comic

Edit Instruction: Transform the provided landscape photograph into a vibrant, stylized cartoon illustration, featuring simplified, smooth green hills, distinct yet rounded tree shapes with lush foliage, clean outlines for all elements, and a soft, slightly textured sky, while maintaining the original pastoral composition.

Figure 12. **Photo to cartoon/sketch/comic.** Applying artistic or photographic styles, such as watercolor, oil painting, or cinematic tone.



Edit Type: Change overall color tone

Edit Instruction: Adjust the overall color balance to introduce a subtle warmth, shifting the cool grays of the sky and pavement and the muted greens of the grass and trees towards a more golden, late afternoon light, while preserving the natural red brick tones and the vibrancy of the foliage.

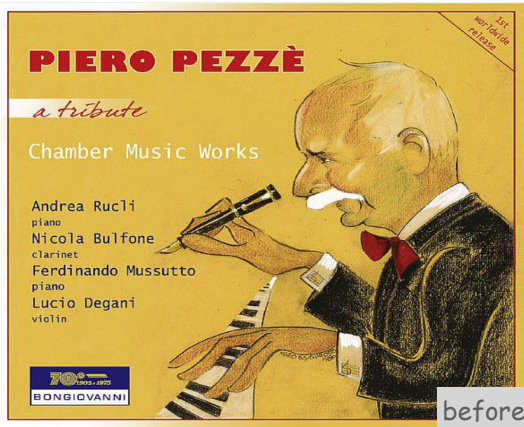
Figure 13. **Change overall color tone.** Adjusting global tonal properties.



Edit Type: Change weather conditions

Edit Instruction: Transform the overcast sky into a vibrant, clear blue sky with a few faint cirrus clouds, casting bright, warm sunlight from the upper right onto the entire scene, creating distinct, sharp shadows from the church spires, buildings, and people, and enhancing the saturation and contrast of the red roofs, stone facades, and distant greenery to evoke a brilliant, sunny summer day in Prague.

Figure 14. **Change weather conditions.** Changing environmental conditions such as turning a sunny scene into rain or fog.



Edit Type: Add new (handwritten/printed/etc) text

Edit Instruction: Add the new text "Special Edition" in a handwritten, elegant cursive script, similar in style and dark reddish-brown color to the existing "a tribute" text, positioned centrally below "Chamber Music Works" and above "Andrea Rucli", ensuring it integrates seamlessly with the album cover's aged paper texture and overall vintage aesthetic.

Figure 15. Add new (handwritten/printed/etc) text . Add text to the image where appropriate.



Edit Type: Change font style or color of visible text if there is text

Edit Instruction: Change the color of the text "MELBOURNE AUSTRALIA" on the overhead banner from its current white to a vibrant, electric yellow, ensuring the letters retain their sharp edges and bold, sans-serif font style for enhanced visibility and a modern aesthetic.

Figure 16. Replace text in signs/posters/billboards. Changing existing text in the image.



before



after

Edit Type: Replace text in signs/posters/billboards

Edit Instruction: Replace the prominent 3D-rendered text 'IDEAS' on the banner with the new text 'VISION', ensuring the replacement maintains the original's dark gray color, three-dimensional shadow/bevel effect, vintage paper texture, and natural lighting, while seamlessly integrating into the existing composition with the fingers pointing towards it.

Figure 17. **Change font style or color of visible text.** Changing the style of text in the image to a specific style.