(a) Thinking Reward Model (TRM)

Image Editing

Instruction: Replace the right bedside lamp with an alarm clock.

Source Image



Edited Image

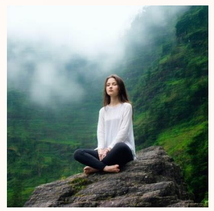


Image Generation

Prompt:

Serene woman with long brown hair wearing a white tunic and dark leggings meditates cross-legged on a boulder, surrounded by lush, misty green mountain slopes in an ethereal nature photograph.

Generated Image



Adaptive Rubric Generation

1. Prompt Alignment

R1: Is the bedside lamp on the right nightstand removed?

R2: Is there an alarm clock on the right nightstand?

...

2. Technical Quality

R1: Does the clock look naturally placed?

R2: Is the wall and nightstand free of visual artifacts?

...

3. Visual Consistency

R1: Is the lamp on the left side preserved in the edited image?

R2: Are the bed, wall picture preserved from the source image?

...

1. Prompt Alignment

R1: Is the woman sitting cross-legged on a boulder?

R2: Are there lush green mountain slopes in the background?

...

2. Technical Quality

R1: Is the composition well-balanced around the subject?

R2: Does the color palette convey an ethereal and serene mood?

...

3. Visual Aesthetics

R1: Is the image clear, natural, and free of visible artifacts?

R2: Are the woman's face, hands, and feet free of deformities?

...

Holistic Analysis & Final Score

Dimension-level Reasoning

Instruction Following	The bedside lamp on the right nightstand is successfully removed, and an alarm clock is added in its place.
Technical Quality	The alarm clock is clearly rendered and naturally blended with the nightstand, and the edited area is free of obvious artifacts.
Visual Consistency	Most scene elements, including the bed, desk area, and lighting are well preserved. However, the lamp on the left is removed.

Final Score

7.5/10

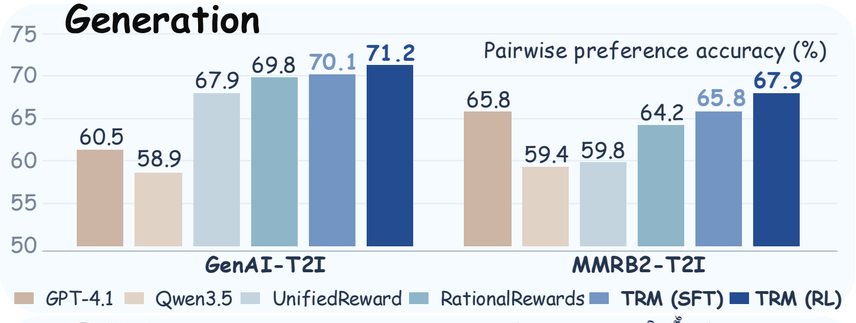
Dimension-level Reasoning

Instruction Following	The image accurately depicts the woman, her attire and meditative pose on a boulder, together with misty green mountain setting.
Technical Quality	The composition is well balanced, and the harmonious color successfully conveys the requested serene and ethereal mood.
Visual Aesthetic	The image is generally clear and without major artifacts, but the woman's hands are visibly deformed, with fused shapes and fingers.

Final Score

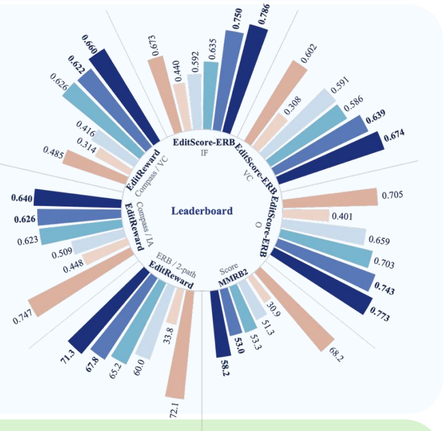
8.2/10

(b) RM Performance



Editing

Rank	Model	Overall
1	TRM (RL)	67.23
2	GPT-4.1	66.98
3	TRM (SFT)	63.57
4	EditScore-72B	61.27



(c) RL Performance

