

# NAMI: Efficient Image Generation via Bridged Progressive Rectified Flow Transformers

## Supplementary Material

### A. Detailed Description of NAMI-1K

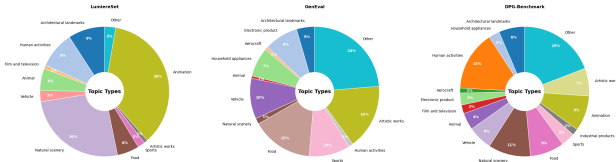


Figure 1. The distributions of topic types of GenEval, LumiereSet and DPG-Benchmark.

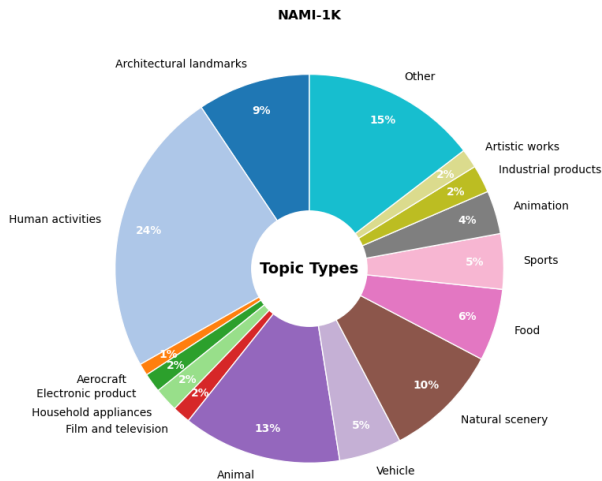


Figure 2. The distribution of topic types of NAMI-1K.

NAMI-1K, consisting of 1,000 prompts with diverse topic categories and varying length distributions. Specifically, 360 short prompts were selected from the open benchmarks GenEval and LumiereSet to characterize the alignment capability of text to image. While 320 human-created prompts were collected from community contributions and user interactions to reflect the model’s performance in real-world user scenarios. Additionally, 320 long prompts generated by CogVlm2, were used to assess performance in complex semantic alignment and instruction-following capabilities.

As shown in Figure 1 and Figure 2, We selected the top

14 topics and analyzed the distribution of sample counts across these categories. Compared to GenEval and DPG-Benchmark, the NAMI-1K dataset exhibits a significantly richer variety of prompt topics and is more closely aligned with real-world application scenarios.

We select five prompts from each of the three major categories for demonstration, as outlined below.

#### Open-benchmarks Prompts

1. "A photo of four apples."
2. "A photo of a toothbrush and a carrot."
3. "A photo of a green bus and a purple microwave."
4. "A beautiful sunrise on mars, Curiosity rover. High definition, timelapse, dramatic colors."
5. "View of a castle with fantastically high towers reaching into the clouds in a hilly forest at dawn."

#### Human-created Prompts

1. "The Little Match Girl."
2. "A pair of sisters happily folding paper airplanes."
3. "A man hugging a tiger in a lucid nightmare."
4. "How to raise a healthy and happy dog."
5. "Chester Zoo staff member taking a picture of cheetah footprint."

#### AI-generated Prompts

1. "A man wearing a red hat, with a beard and mustache, and a red scarf around his neck. He is looking directly at the camera, giving the impression of a portrait. The man appears to be well-dressed, possibly in a suit, and his attire is complemented by the red hat and scarf. The overall atmosphere of the painting is one of sophistication and elegance."

2. "A blue gate with a white fence, adorned with pink flowers. The gate is open, allowing a view of the garden beyond. The garden is filled with various potted plants, some of which are placed on the ground, while others are positioned on the fence. The plants are of different sizes and colors, creating a vibrant and lively atmosphere. The combination of the blue gate, white fence, and the abundance of flowers and plants make the scene visually appealing and inviting."

3. "A colorful illustration of a rocket blasting off into space, with a bright moon in the background. The rocket is positioned towards the top of the scene, while the moon is located towards the left side. The illustration is set against a dark background, which contrasts with the vibrant colors of the rocket and the moon. The overall scene is visually appealing and captures the essence of space exploration."

4. "The afterglow of the setting sun casts a golden hue

on the winding Great Wall, presenting a realistic scene. The bricks of the Great Wall appear ancient and sturdy under the golden sunlight, with each brick clearly visible. The distant mountains, illuminated by the sunset, show varying shades of orange and red, with distinct contours. Occasionally, a few birds fly over the Great Wall, adding a touch of vitality to the scene. In the background, the sky is dyed with brilliant shades of orange, red, and purple, with a few clouds scattered, making it exceptionally magnificent.”

5. ”A man in a black suit is standing on the street, holding a golden saxophone. His head is slightly bowed, and his eyes are closed, seemingly immersed in the music. The metallic sheen of the saxophone glistens in the sunlight, with details such as rings and keys clearly visible. The man’s shoes are black leather, polished to a shine. The background features a busy city street, with a few cafes on the roadside, their tables and chairs neatly arranged outdoors, and several customers leisurely sipping coffee. In the distance, there are skyscrapers, with blue skies and white clouds highlighting the city’s prosperity.”

## B. More Results of NAMI



Figure 3. ”Pug dog listening to music with big headphones.”, ”Low angle of pouring beer into a glass cup.”



Figure 4. ”Aerial view of a hiker man standing on a mountain peak.”, ”Close up of grapes on a rotating table. High definition.”



Figure 5. ”Time and Space Tunnel”, ”Wedding dress”



Figure 6. ”a photo of four bowls”, ”a photo of a bird”



Figure 7. ”a photo of a cat”, ”A fat rabbit wearing a purple robe walking through a fantasy landscape.”



Figure 8. ”A panda standing on a surfboard in the ocean in sunset, 4k, high resolution.”, ”A squirrel eating a burger.”



Figure 9. "Watch the fire from the shore", "red fox running in the snow"



Figure 13. "Teddy bear walking down 5th Avenue, front view, beautiful sunset, close up, high definition, 4k.", "Time lapse at a fantasy landscape, 4k, high resolution."



Figure 10. "Cesar, french sculptor, in his studio", "Grace Gummer at the party"



Figure 14. "a rustic lantern of wood, candles around and vintage vases with baby's breath", "cotton picking season. blooming cotton field. close up of the crop before the harvest, under a golden sunset light. - cotton stock videos & royalty-free footage."

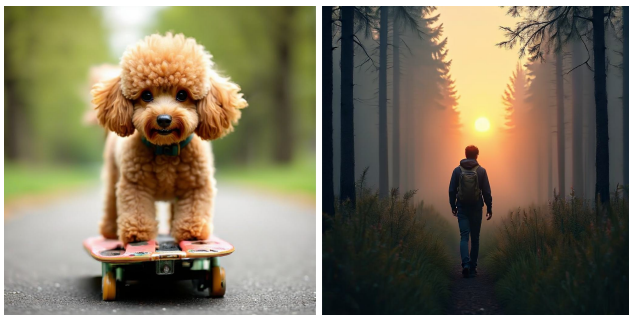


Figure 11. "Toy Poodle dog rides a penny board outdoors", "Traveler walking alone in the misty forest at sunset."



Figure 15. "a vast grassland appears particularly green under the sunlight, dotted with colorful wildflowers. A gentle breeze causes waves in the grass. In the distance, rolling mountains stand majestically against the backdrop of blue sky and white clouds. A few white sheep graze leisurely on the grass, with a shepherd dog guarding nearby. In the foreground, a clear stream meanders through, its water sparkling in the sunlight. Butterflies flutter among the flowers, and birds sing joyfully on the branches. The overall picture is rendered in a realistic style, with rich details and vibrant colors.", "Talk on paper"

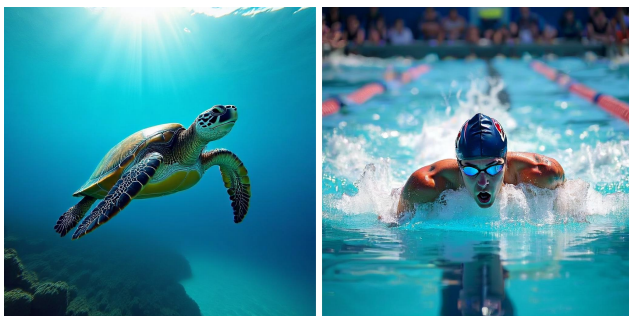


Figure 12. "Turtle swimming in ocean.", "PHOTOS: Life in the fast lane for our Olympic swimmers"



Figure 16. "Omar Elabdellaoui injured in the eye as a result of the accident at the New Year's Eve celebration", "A specimen of a variety of quartz showing conchoidal fracture"



Figure 17. "The image features a beautiful painting of a forest at night, with a full moon illuminating the scene. The moon is positioned in the upper left corner of the painting, and its light casts a glow on the trees and the surrounding area. The forest is filled with trees, some of which are taller and closer to the foreground, while others are smaller and further away. The painting captures the serene atmosphere of the nighttime forest, with the moon as the main focal point.", "a man wearing a cowboy hat and a suit, with a beard and mustache. He is looking directly at the camera, giving the impression of a confident and distinguished appearance. The man's attire and hat suggest that he might be a cowboy or a businessman with a unique sense of style. The overall atmosphere of the image is one of sophistication and confidence."

### C. More Discussion on the BridgeFlow Method

The proposed Linear BridgeFlow ensures cross-stage velocity consistency. In contrast, nonlinear mappings tend to induce trajectory curvature and destabilize stage-wise optimization. Specifically, the conditional probability path is defined as:

$$\hat{x}_{s_k} \sim \mathcal{N}(\alpha_k \odot \text{Up}(\text{Down}(x_1)), \Sigma_k), \quad (1)$$

$$\hat{x}_{e_{k-1}} \sim \mathcal{N}(\beta_{k-1} \odot \text{Down}(x_1), \Sigma_{k-1}), \quad (2)$$

we first apply a simple upsampling transformation to achieve resolution alignment

$$\text{Up}(\hat{x}_{e_{k-1}}) \sim \mathcal{N}(\beta'_{k-1} \odot \text{Up}(\text{Down}(x_1)), \Sigma'_{k-1}). \quad (3)$$

We can further complete the distribution alignment by learning a linear transformation

$$\hat{x}_{s_k} = W \cdot \text{Up}(\hat{x}_{e_{k-1}}) + B \quad (4)$$

to match the means and covariances of the two distributions. We also provide a comparison of different bridging strategies in Table 7 of the main text, showing that more complex modules do not yield further improvements.

### D. Example of NAMI for image editing

We provide a simple, training-free example of directly applying NAMI for image editing. As shown in Figure 18, we share the layouts and concept contours generated in the first stage, and modify the instructions in the subsequent stages to obtain edited images with different attributes. This demonstrates that NAMI is also convenient and promising for applications in other tasks.

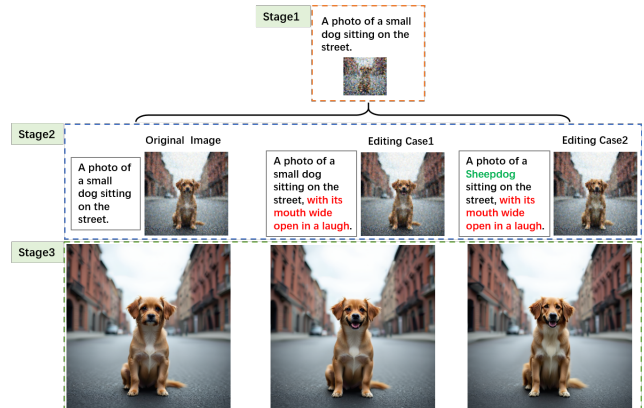


Figure 18. Image Editing of Directly Applying NAMI.

### E. Detailed explanation of the internal dataset

Our internal dataset contains approximately 100,000 high-quality image-text pairs, curated for text-to-image generation tasks and primarily intended for high-quality fine-tuning of generative models. The data was collected from diverse sources across the internet, covering a wide range of topics and visual styles to ensure rich diversity in both content and semantics. To improve caption quality and alignment, we used the CogVLM2 model to generate recaptions for each pair. The resulting image-text pairs then underwent multi-stage filtering as a whole, including evaluations of semantic coherence, aesthetic quality, and safety. Human assessments were also performed on the pairs to further ensure correctness, appropriateness, and overall quality. This rigorous process produced a meticulously curated dataset suitable for high-fidelity model fine-tuning.

### F. The visualization of different methods

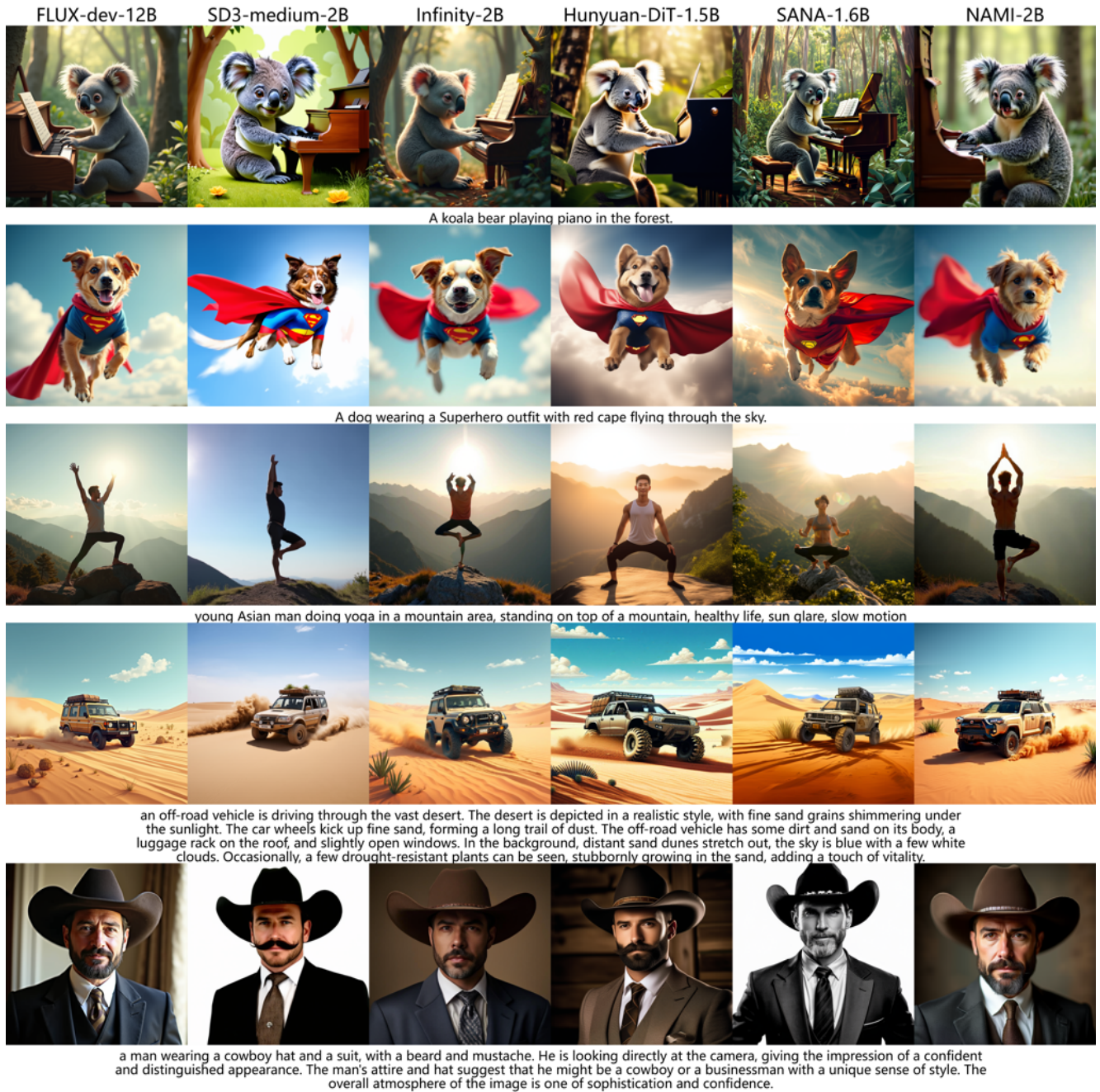


Figure 19. Visualization of image generation result comparison across different methods