Image Generation Models vs. My Character Designs: A Timeline

2025-11-13

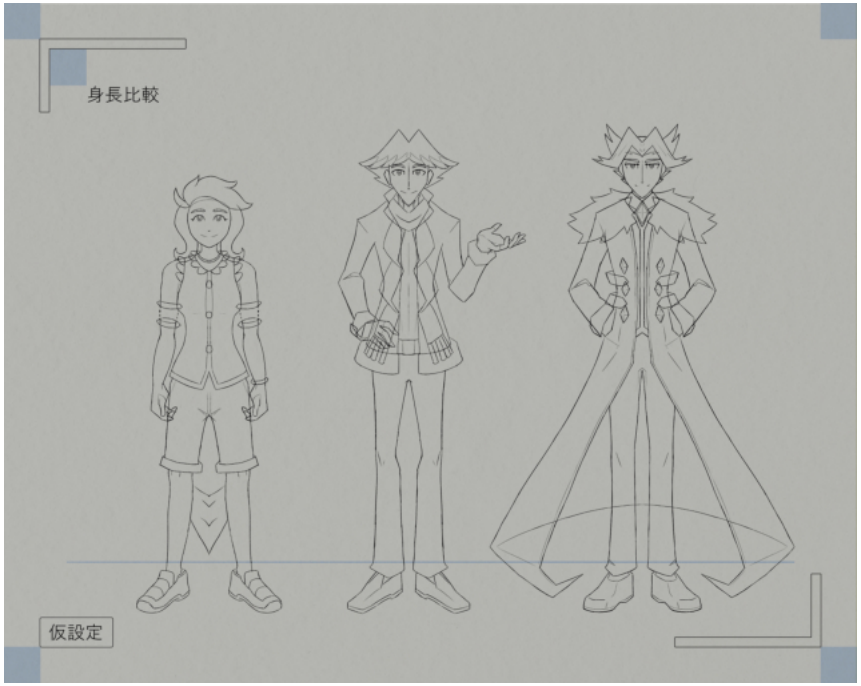
Agenda

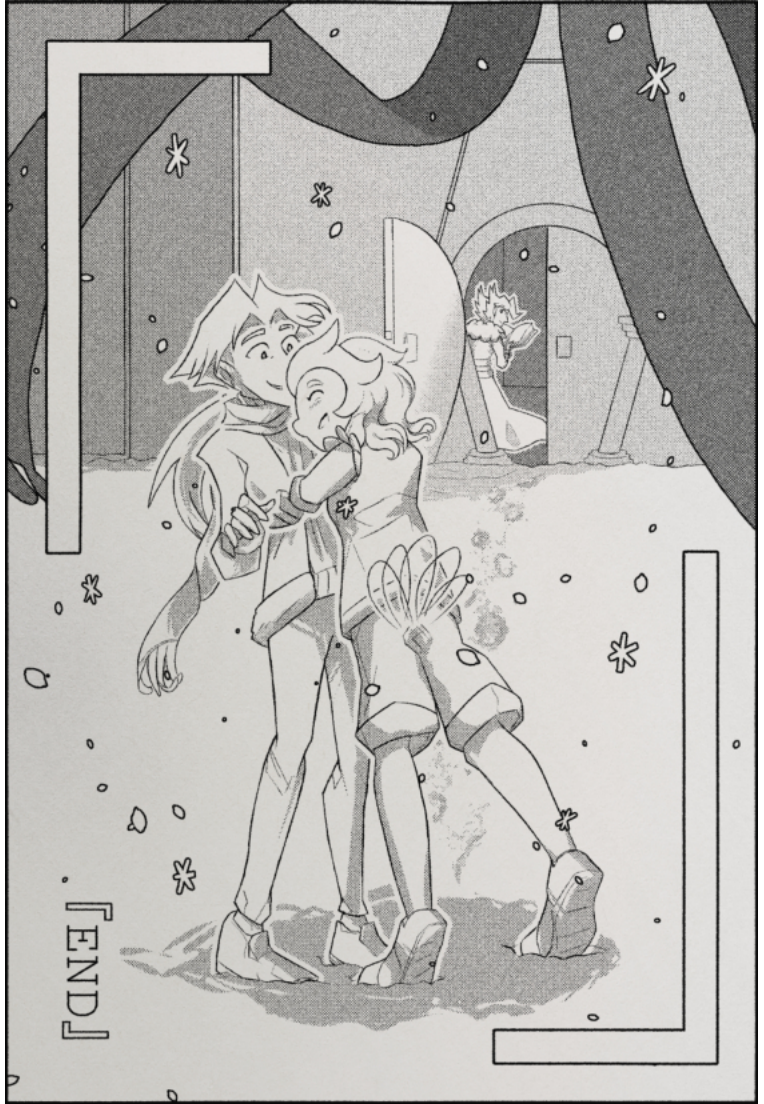
1. Character design TL;DR
2. Image generation model TL;DR
3. Why character designs?
4. My character designs
5. Share my experiences using the following image generation models:
    1. DALL·E 2
    2. Stable Diffusion 1.5
    3. Stable Diffusion 1.5 + LoRA
    4. Stable Diffusion XL + LoRA
    5. Gemini 2.5 Flash + Nano Banana
    6. + Bonus!
6. Closing thoughts

> **Note:** All dates in the presentation represent when I tried out the model, not when it first released.

Character Design TL;DR

- The structured process of creating a visual design such that the design is:
  - consistently drawn within your creative work (intrinsically identifiable)
  - distinctive outside your creative work (extrinsically identifiable)
  - …applied to fictional characters





Intelligence is knowing Solana is not a tomato.

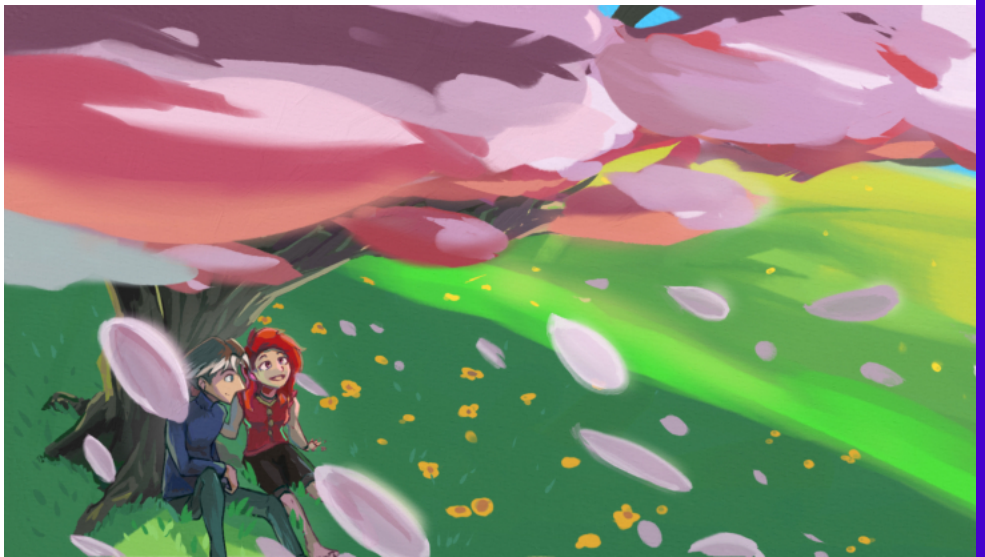
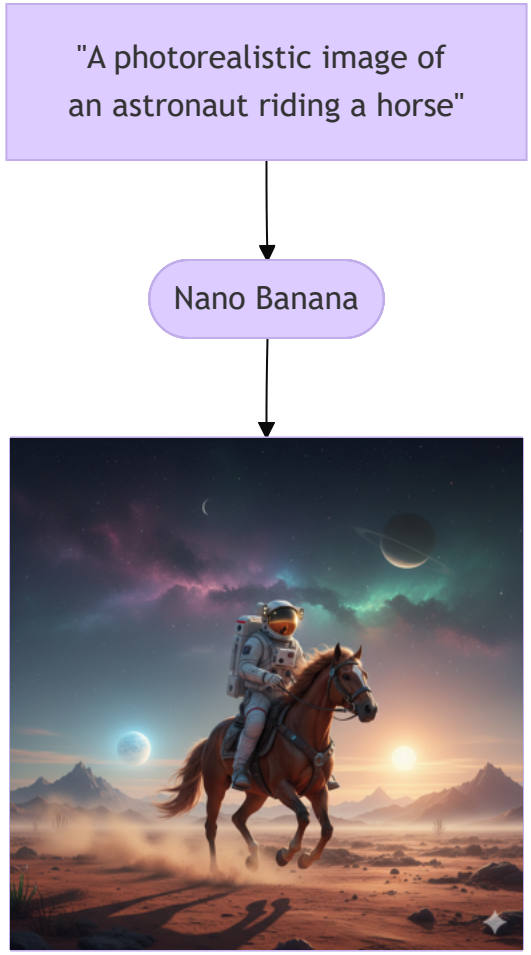


Image Generation Models TL;DR

- Computer software that can generate digital images from a text + optional image prompt



Why Character Designs?

Why character designs?

Benchmarking

- Character designs are distinctive without being too obscure for image models to handle.

Why character designs?

Example: text-to-image benchmarks

"A photorealistic image of an astronaut riding a horse"

DALL·E 2 [2022-04-06]

"a photograph of an astronaut riding a horse"

Stable Diffusion XL 1.0 [2023-08-17]

"a photograph of an astronaut riding a horse on moon"

Flux Pro [2024-12-13]

"A photorealistic image of an astronaut riding a horse"

Nano Banana [2025-10-16]

Why character designs?

Example: "concept-to-image" benchmarks

(no prompt)

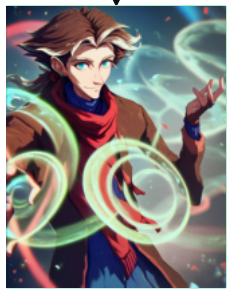
DALL·E 2
[2022-08-28]

"masterpiece,
armin_vicerre, portrait,
3/4 view, 1boy, solo"

Stable Diffusion 1.5
(Anything V3)
[2023-12-06]

"score_9, score_8_up,
score_7_up,
source_anime,
armin_vicerre, key art,
science sci-fi laboratory,
glowing portal gate in
background, biology,
detailed background, lean,
thin"

Stable Diffusion XL
(Pony Diffusion V6 XL)
[2024-08-08]



???

Nano Banana
[2025-08-29]

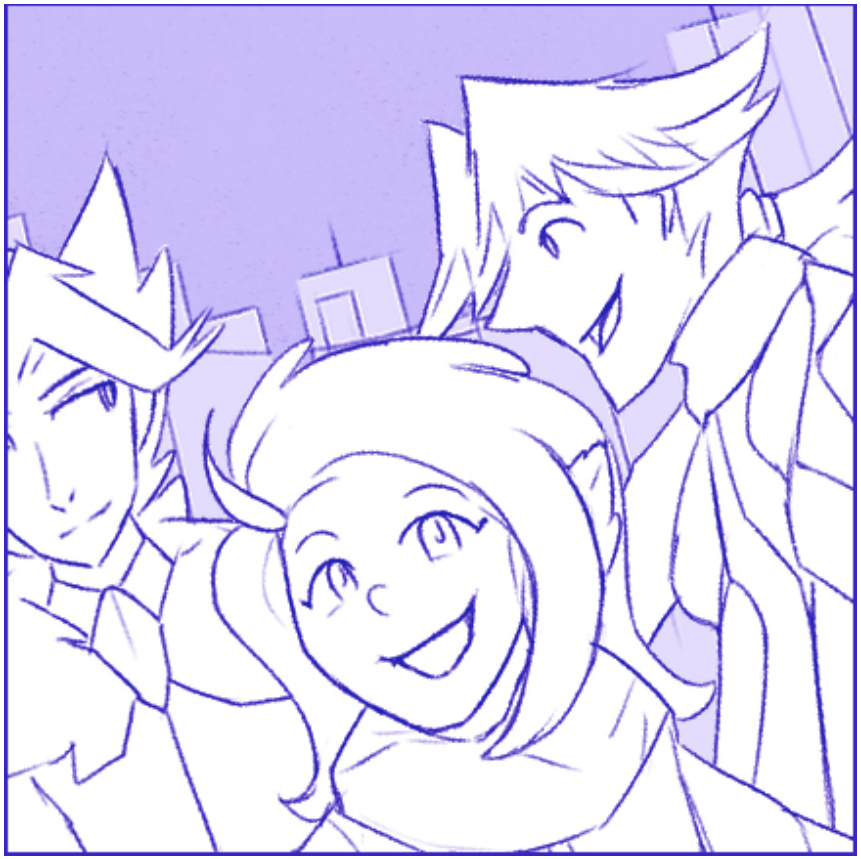
???

Why character designs?

Applications

- Project-specific inspiration
- Project-specific art/design learning material
- Personal interests
- This will be explained in more detail later!

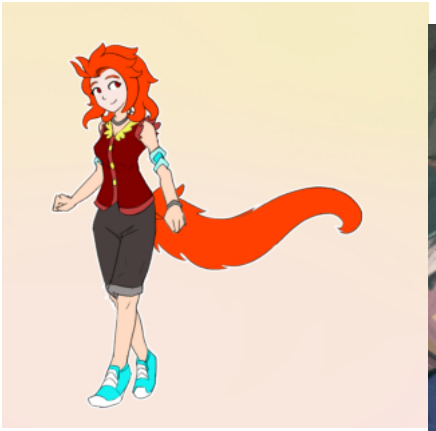
My Character Designs



My Character Designs – Solana

- "Starter" character design
- Has "easy" and "hard" design elements
- Has a visual theme

* Includes older designs.

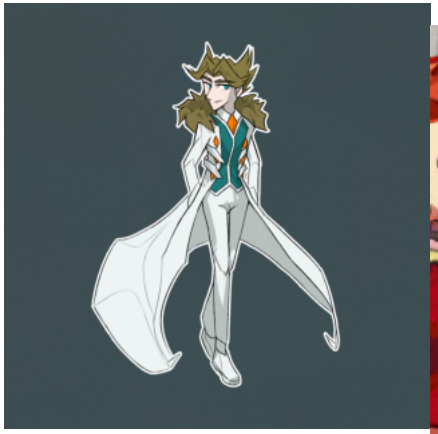


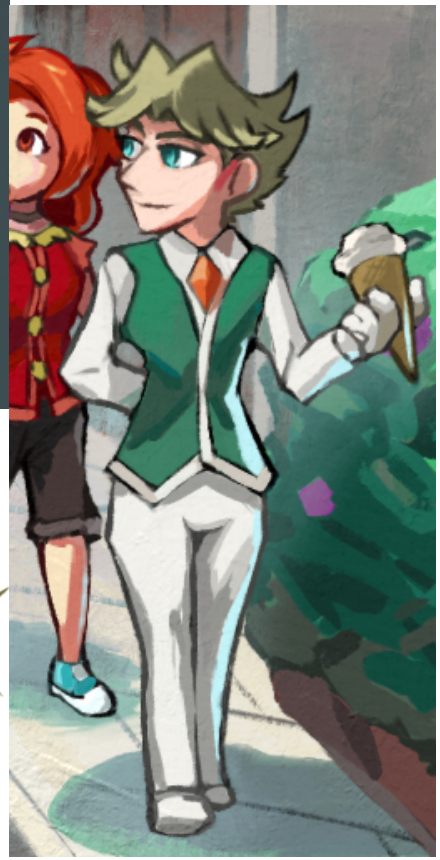
My Character Designs – Vic

- Default "test subject"
- Has tricky design features
- Useful for testing design fidelity across styles

My Character Designs – Alis

- Used in low-shot applications + failure cases





Summary

Solana

- Easy + hard design elements

Vic

- Uncommon design elements

Alis

- Edge cases

Checkpoint

Any questions so far?

DALL·E 2

2022-08-28

DALL·E 2

- Cloud-based
- First publicly available model to support image input → image output

Results

- Output images have qualities of input images 👍

Results

- Doesn’t copy subject directly

Summary

- Proof of concept
- Acts as a baseline for expectations

Stable Diffusion 1.5

2022-12-29

4 MONTHS, 1 DAY LATER

Stable Diffusion 1.5

- Allows for image-to-image generations
- Finer-grain control over generated output
- Runs on consumer hardware

Technical details:

- GPU: Nvidia GeForce RTX 3080 Ti (12 GB VRAM)
- Browser UI: Automatic1111
- Base model: Anything V3 (SD 1.5 fine-tune for anime-style art)

Results

- Can’t prompt for much if tokens are dedicated to visual description
- Prone to hallucinations

Summary

- Images start being usable!
- Very fun to poke at the model and see what it knows
- So much trial and error
- Hallucinations are fun

Stable Diffusion 1.5 + LoRA

2023-12-15

11 MONTHS, 16 DAYS LATER

Stable Diffusion 1.5 + LoRA

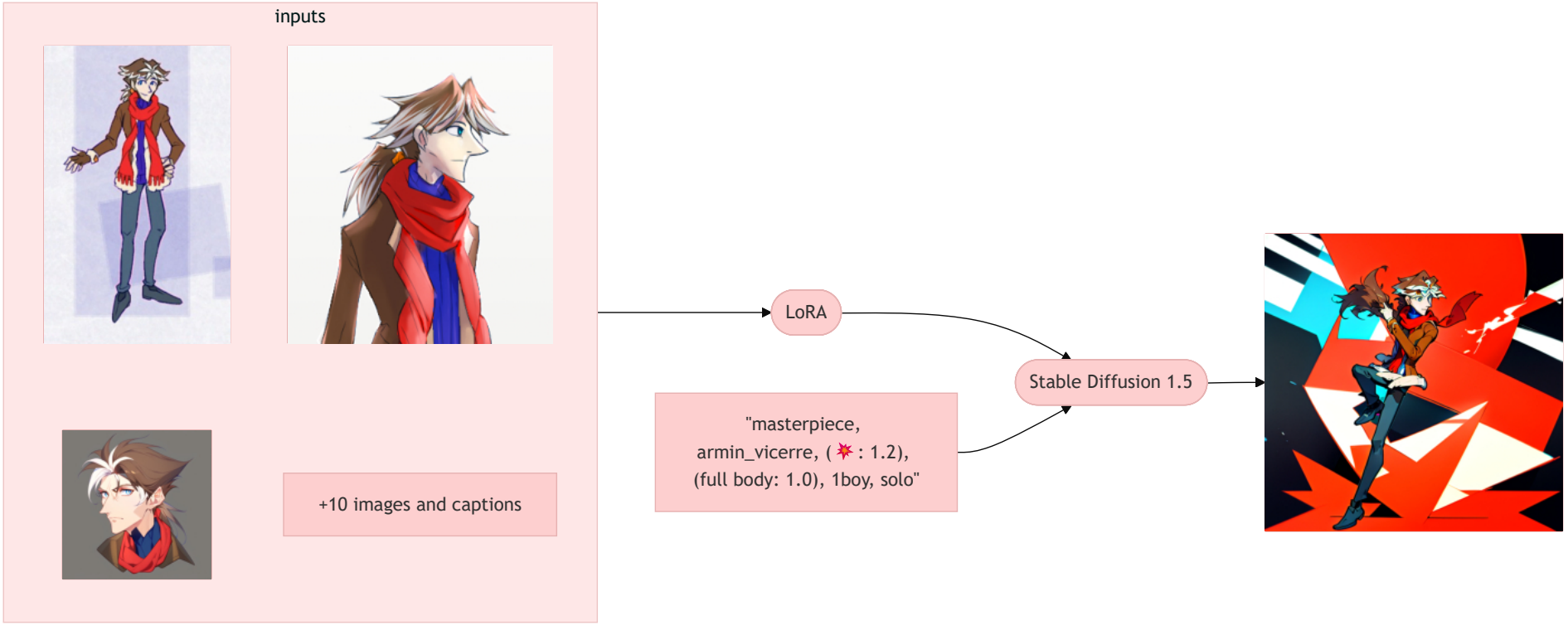
- Low-Rank Adaptation = "LoRA"
- Method of teaching image model new concepts, e.g., subjects and styles
- Vastly more efficient than training the entire model

Technical details:

- LoRA training browser UI: Kohya
- Image generation browser UI: ComfyUI
- Base model: Counterfeit-V3.0 (SD 1.5 fine-tune for anime-style art)

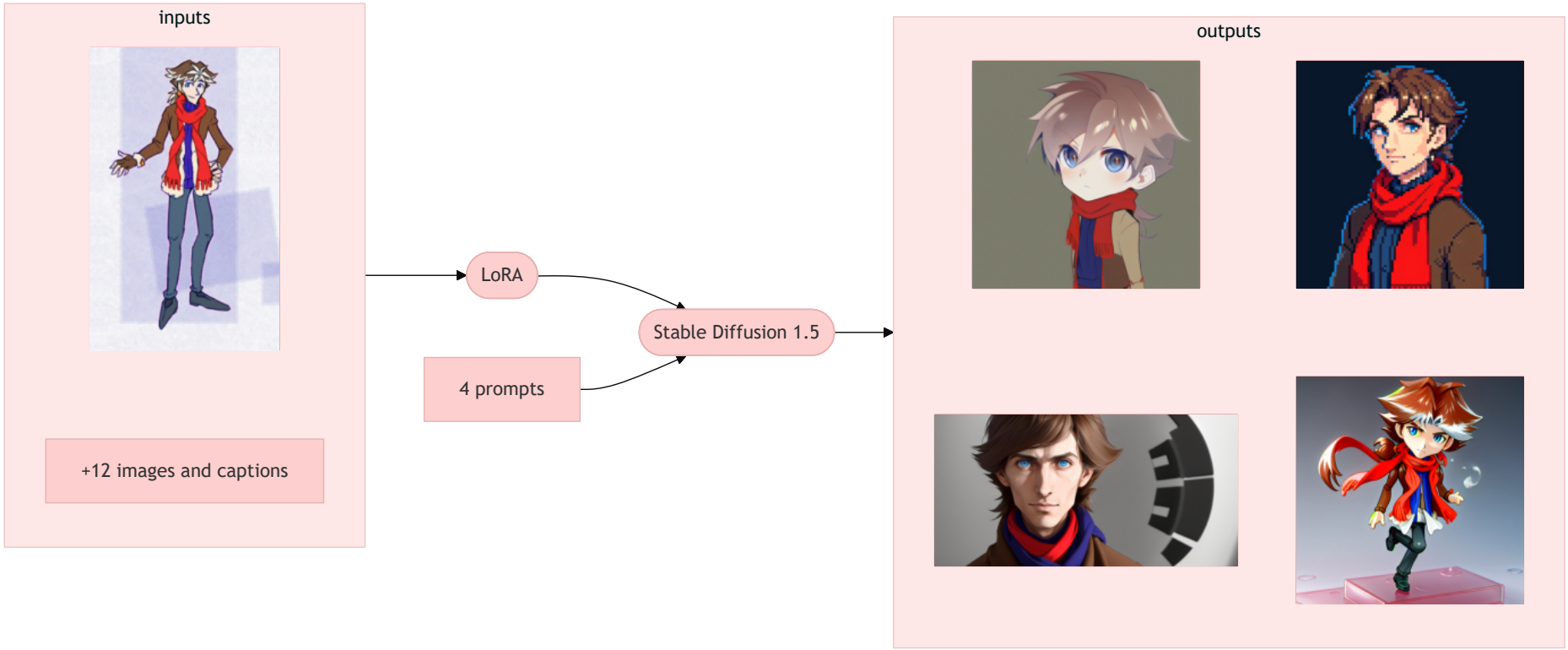
Results

- Actually a game-changer
- Good at inferences



Results

- Frees up tokens to be used on other image aspects

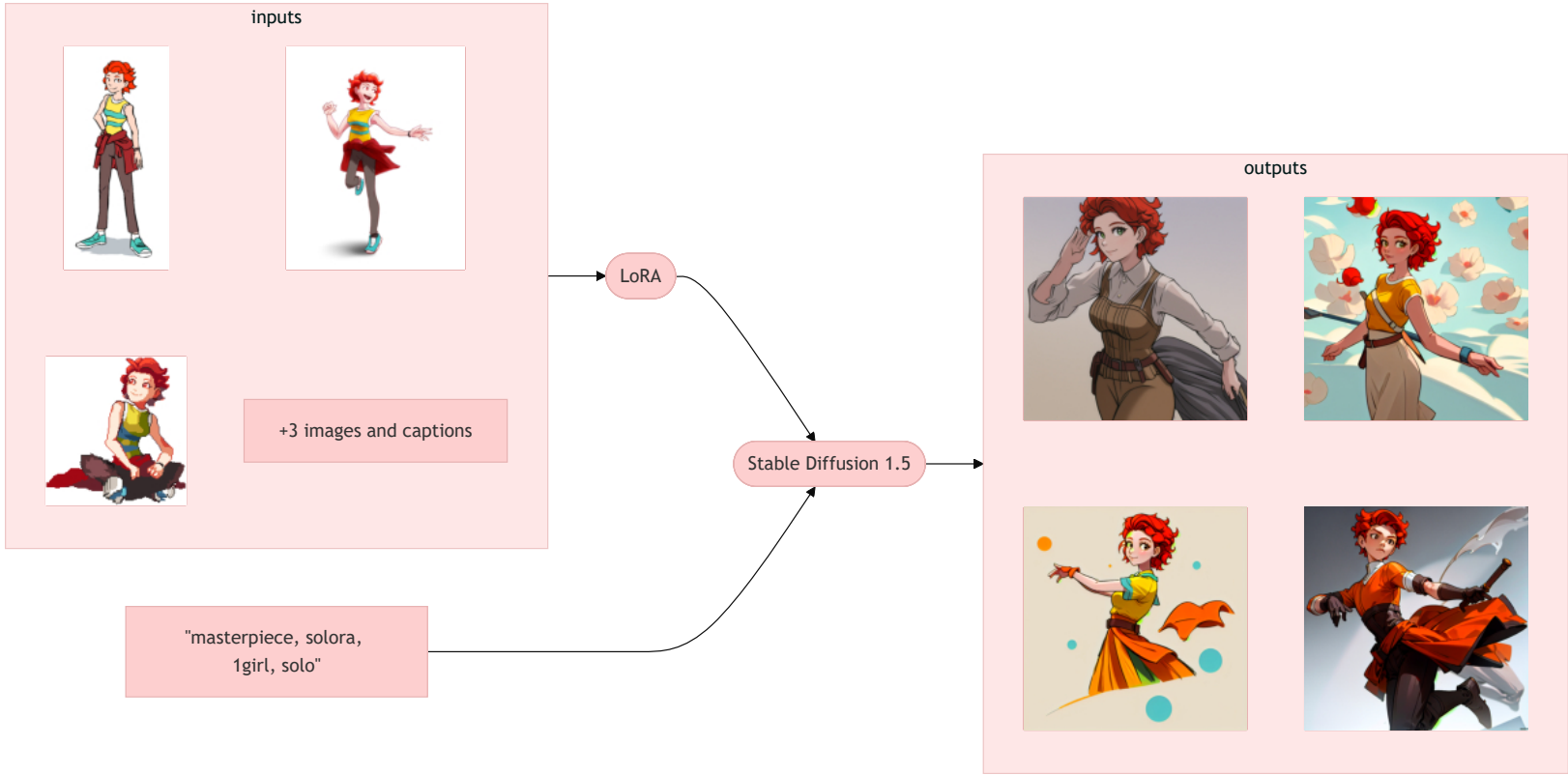


Results

- Not great at figuring out cohesion
- I don’t mind since image generations are inspirational at best
- Other people in the open-source community have developed tools to limit the extent of "nonsense" in outputs

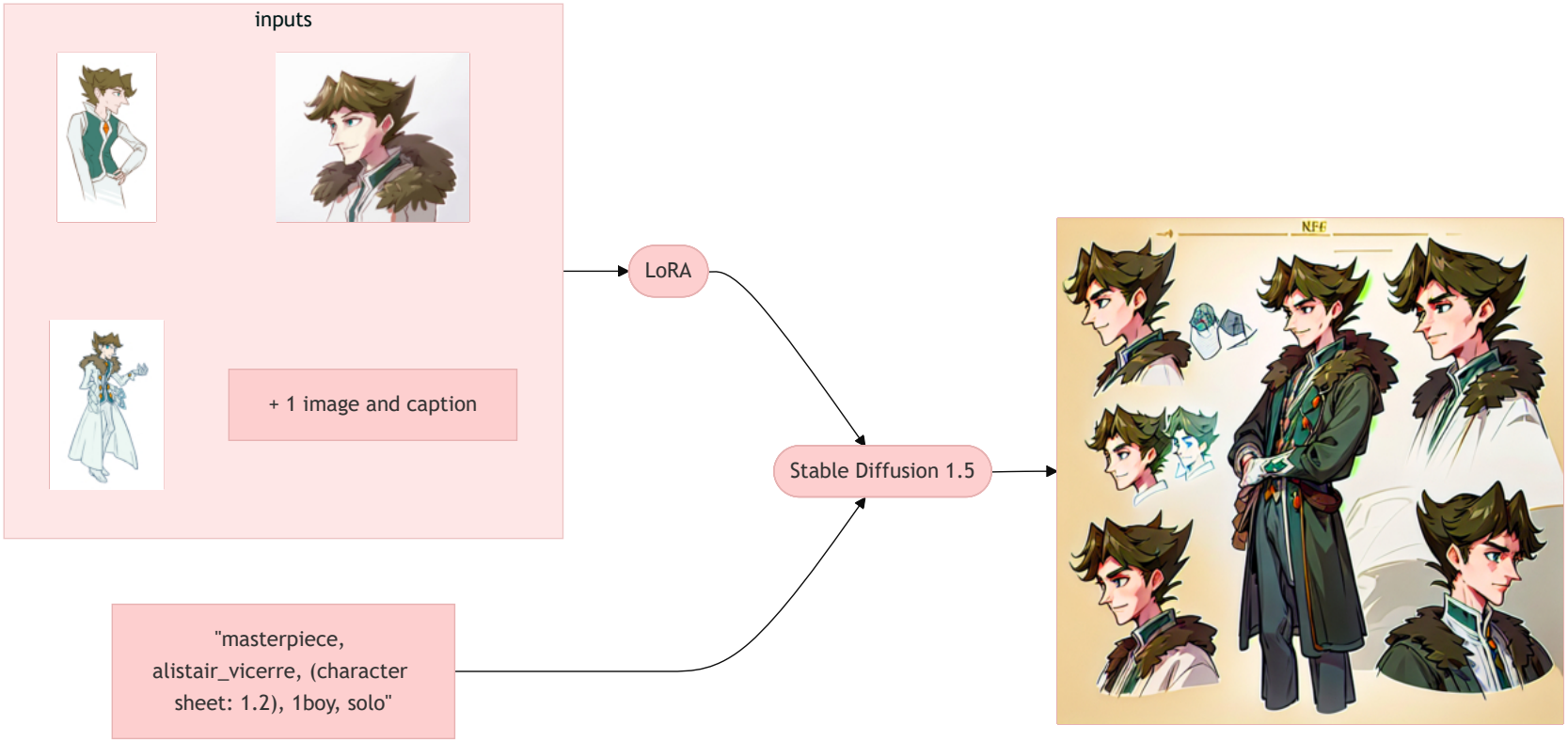
Results

- Underfitting



Results

- Overfitting on low-shot training data



Summary

- Amazeballs
- Addictive
- Unusual generations can be inspiring
- Or just entertainingly wrong

Stable Diffusion XL + LoRA

2024-02-20

2 MONTHS, 5 DAYS LATER

Stable Diffusion XL + LoRA

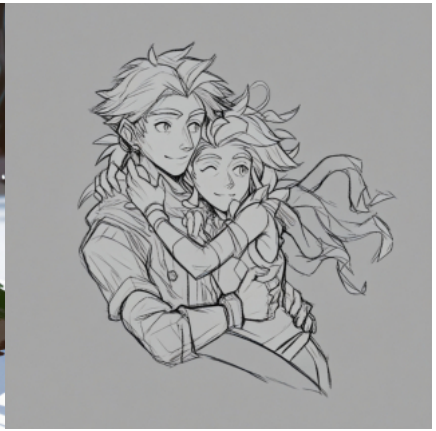
- Like Stable Diffusion, but XL
  - U-Net parameters: 860 million → 2.6 billion
  - Text encoder parameters: 123 million → 817 million
- Hard to give an apples-to-apples comparison due to training methods evolving in parallel

Technical details:

- Base model: Pony Diffusion V6 XL: (SDXL "retrain" for digital art)
- Workflow: Same as Stable Diffusion 1.5 + LoRA

Results

- Less "smeary", more cohesive
- Can handle more complex scenes

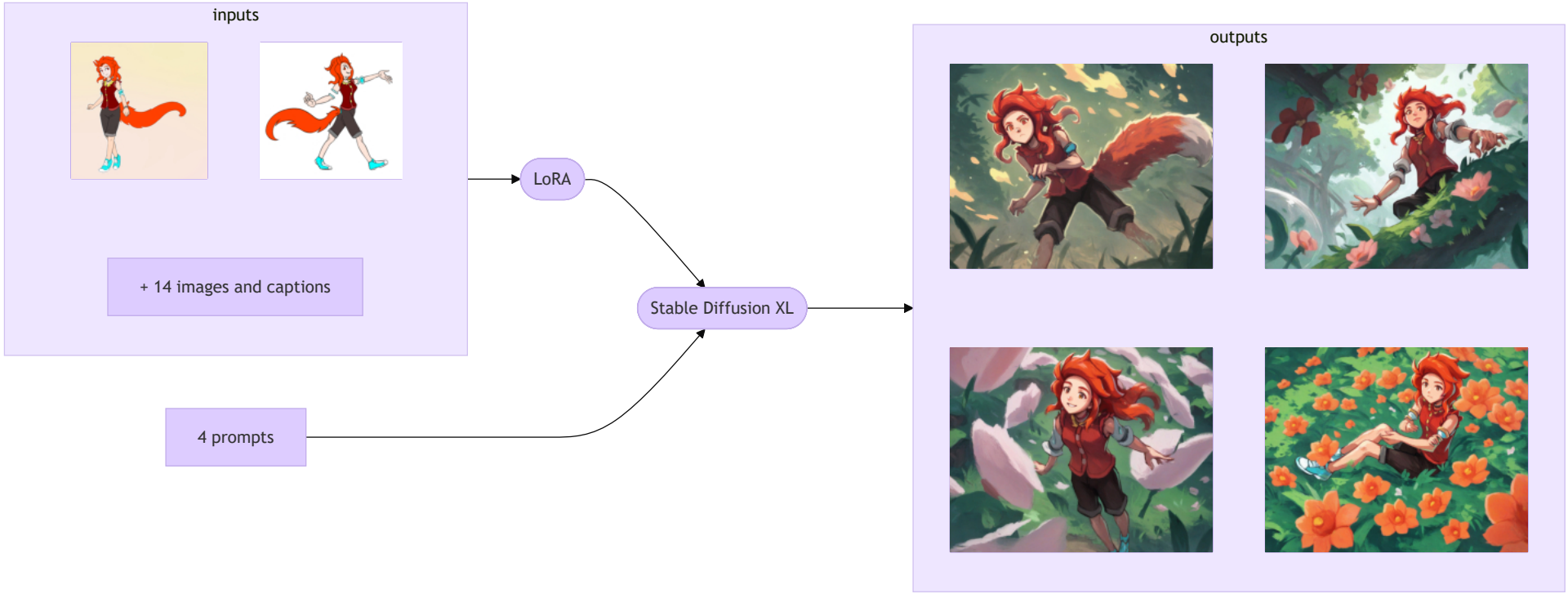


Results

- Just as good at adapting characters to different styles. 👍

Results

- Underfitting is much less problematic

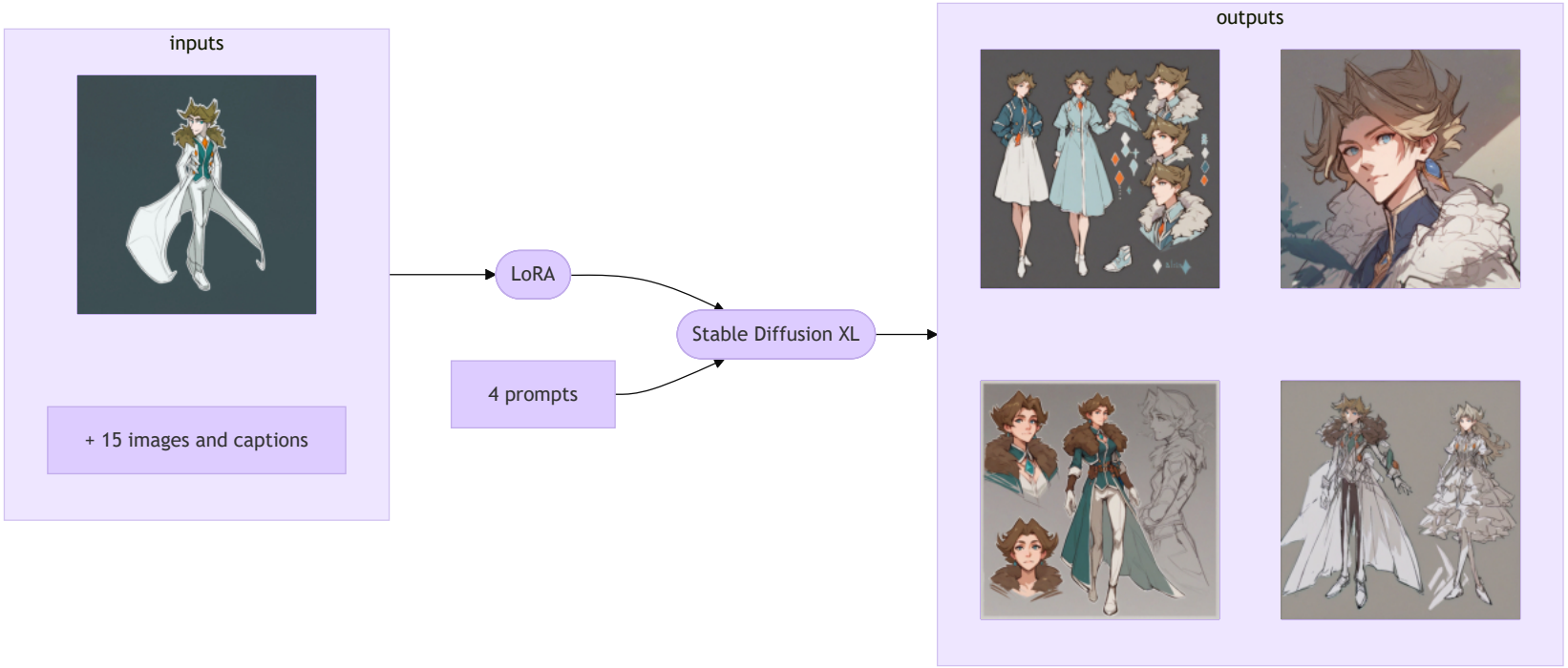


Results

- Overfitting is still an issue

Results

- Open-source, customizable model means biased training data



Summary

- Great improvements toward character fidelity
- Better at "understanding" a scene
- Failure cases are less entertaining wrong

Gemini 2.5 Flash + Nano Banana

2025-08-29

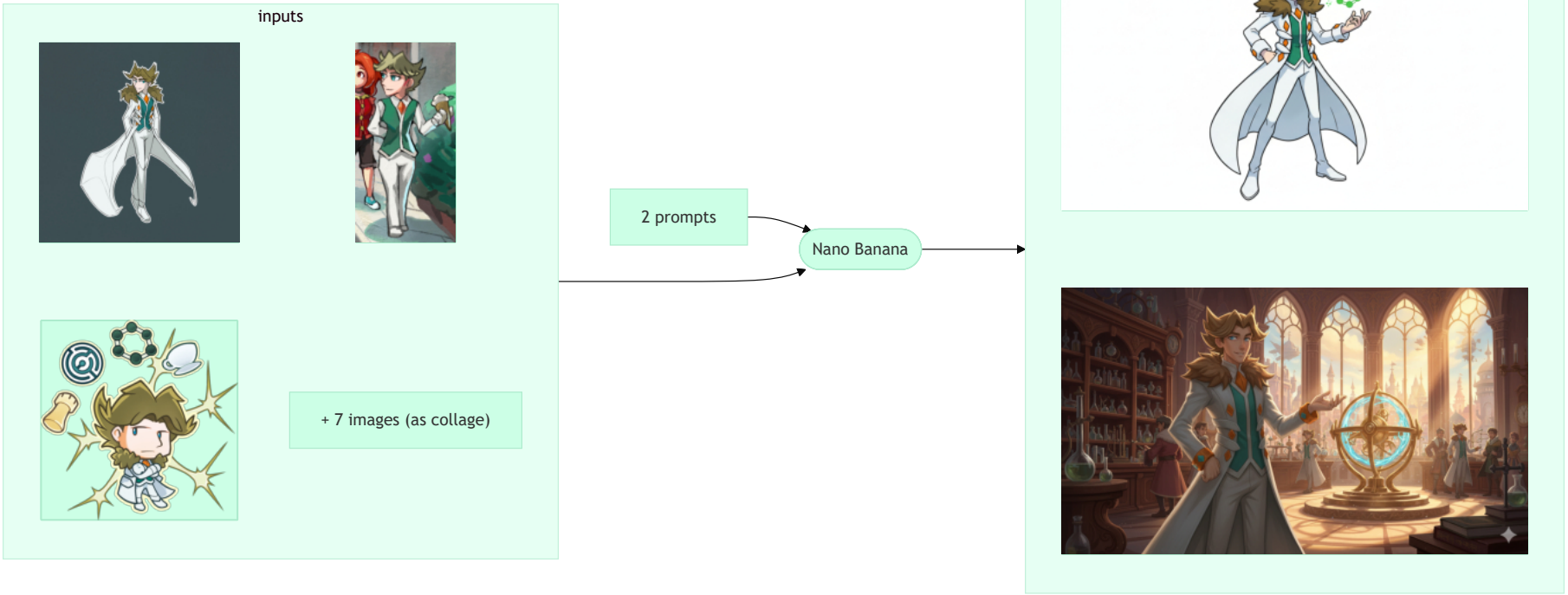
1 YEAR, 6 MONTHS, AND 9 DAYS LATER

Gemini 2.5 Flash + Nano Banana

- Cloud-based
- Multimodal model

Results

- Can properly draw Alis 👏

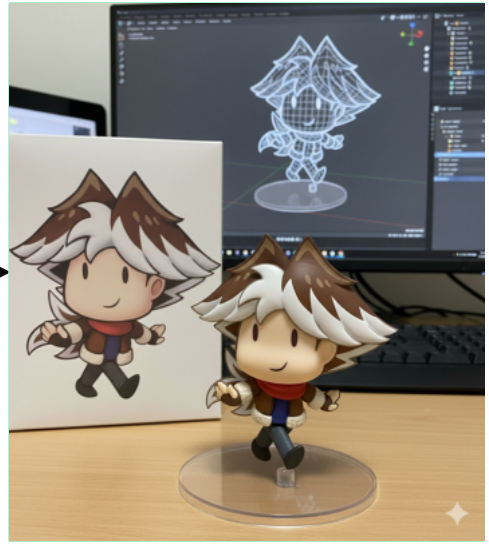


Results

- Can inference input images

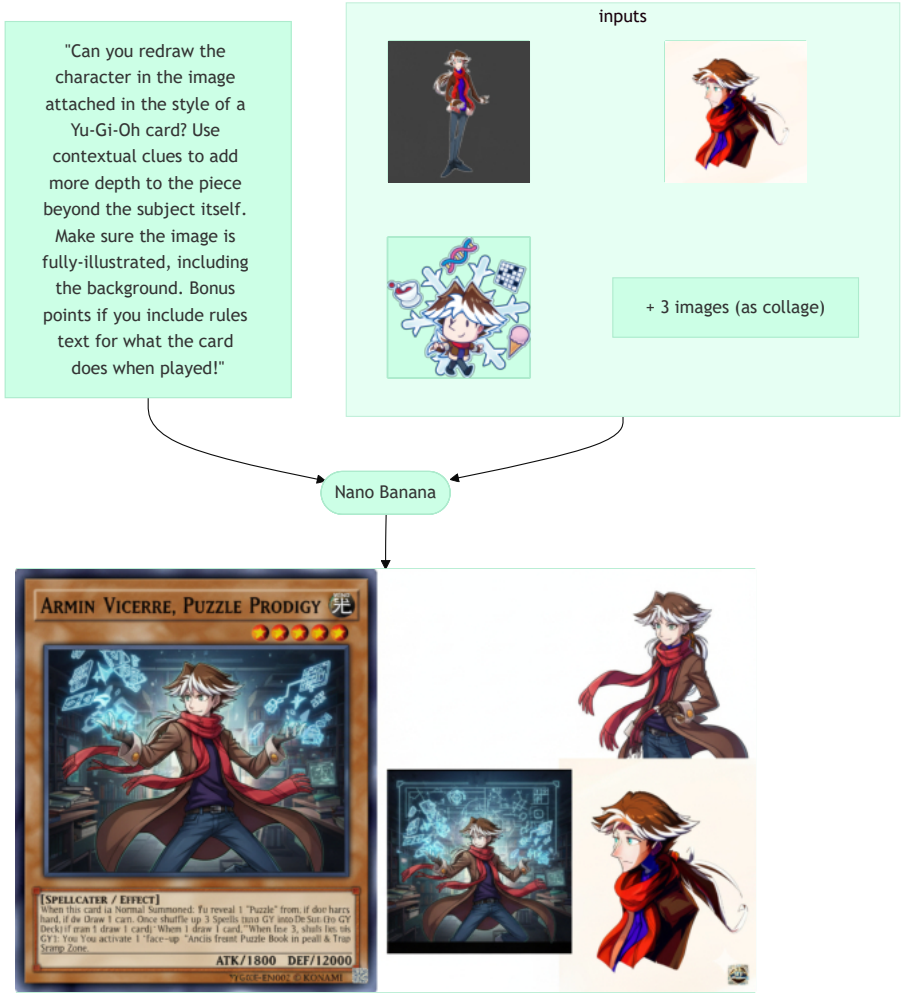
"turn this character into a character 1/6 figurine, behind it place a box of the figurine with the characters image printed on it, and a computer showing the blender modeling process on the screen in the background, add a round plastic base with the figure standing on it, the scene is indoors"

Nano Banana



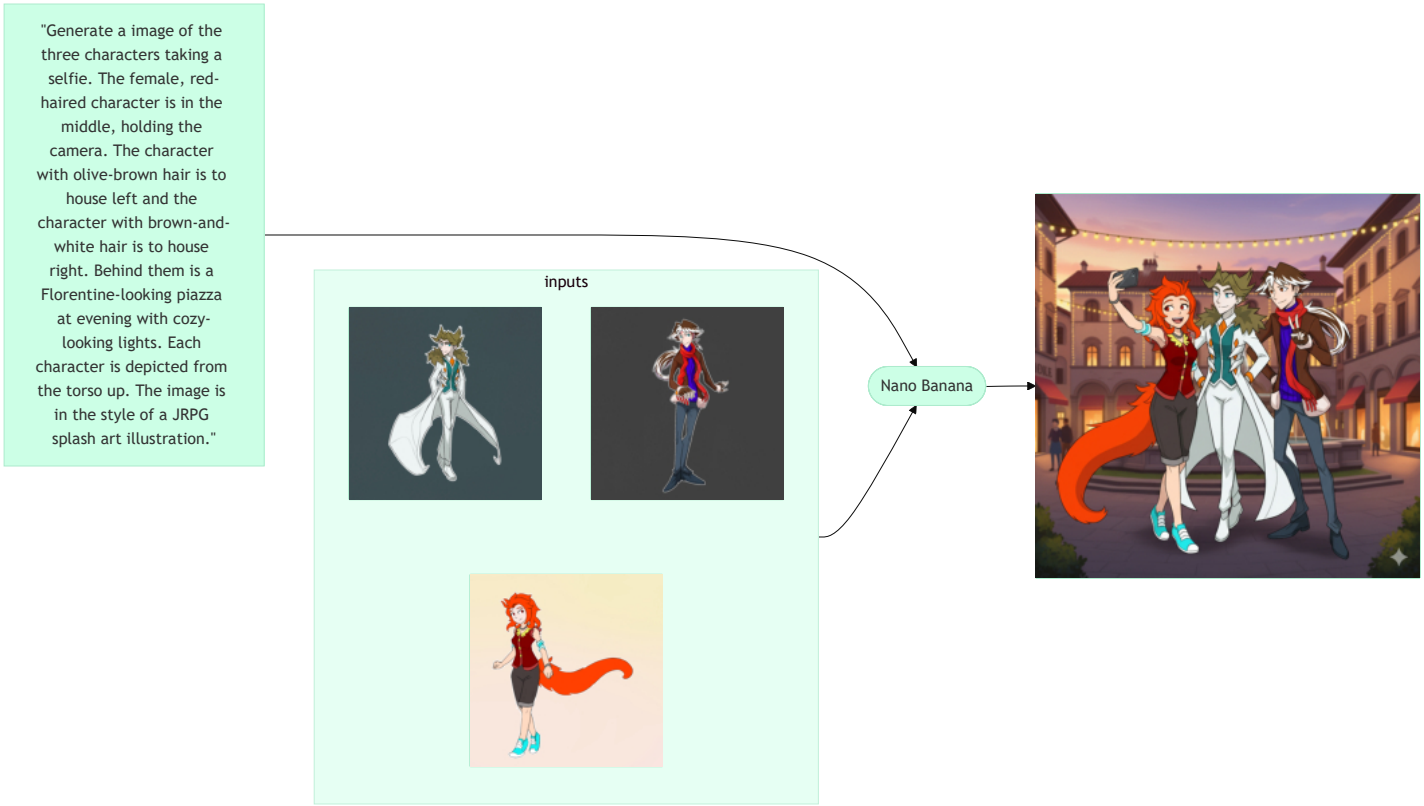
Results

- Can inference input images
- (and it has OCR, which is nifty)



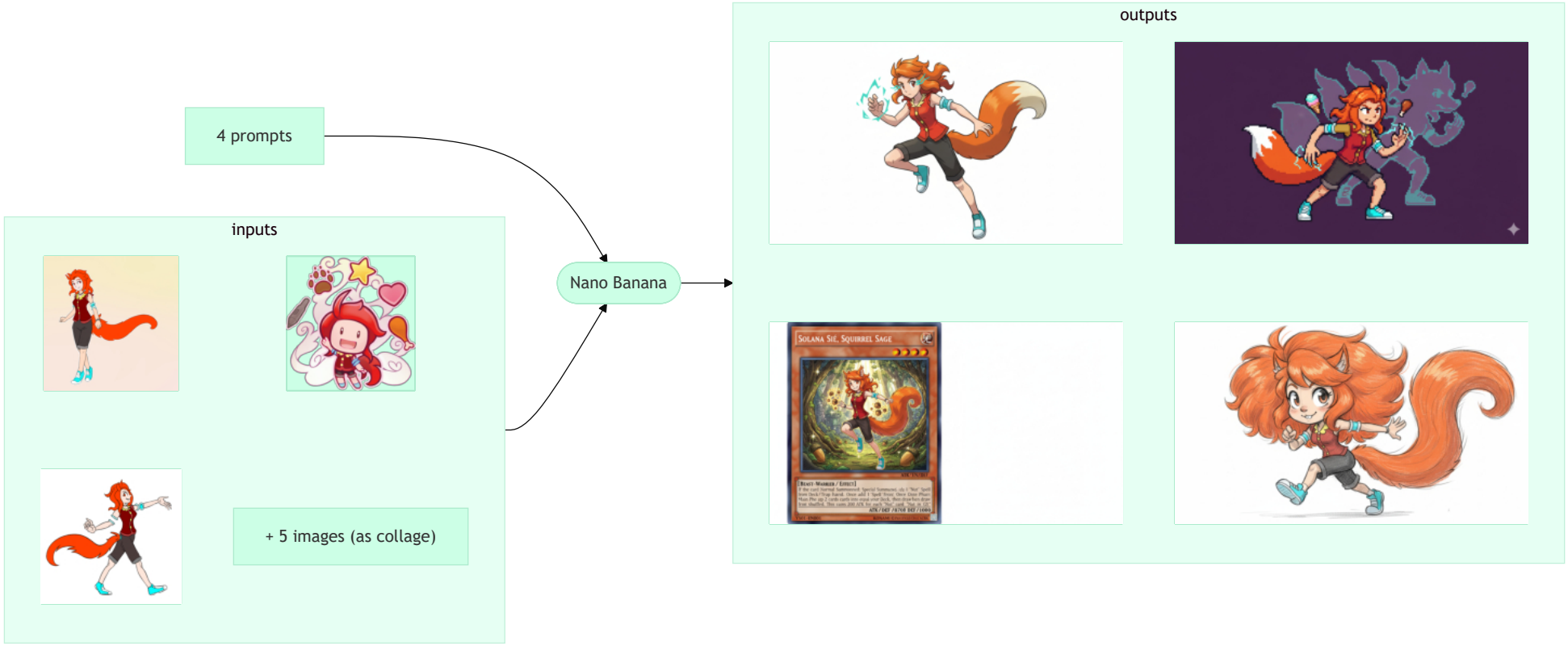
Results

- Excellent at image modifications + composition without copy-pasting or overfitting



Results

- Inferences can be incorrect, though not ungrounded



Summary

- Excellent understanding of image elements
- Excellent inferences
- Fun and surprising in its own way

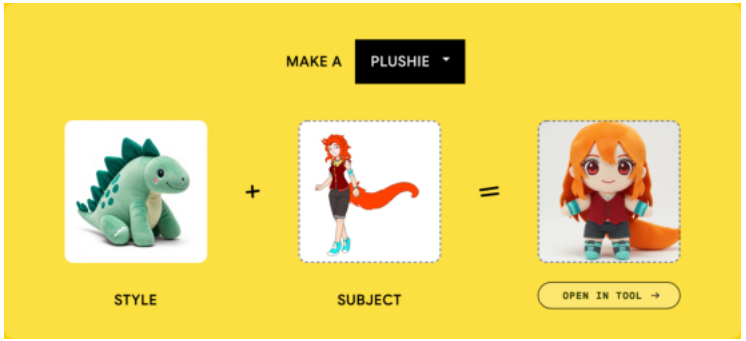
Bonus

Bonus

- Previous slides were not a comprehensive journey

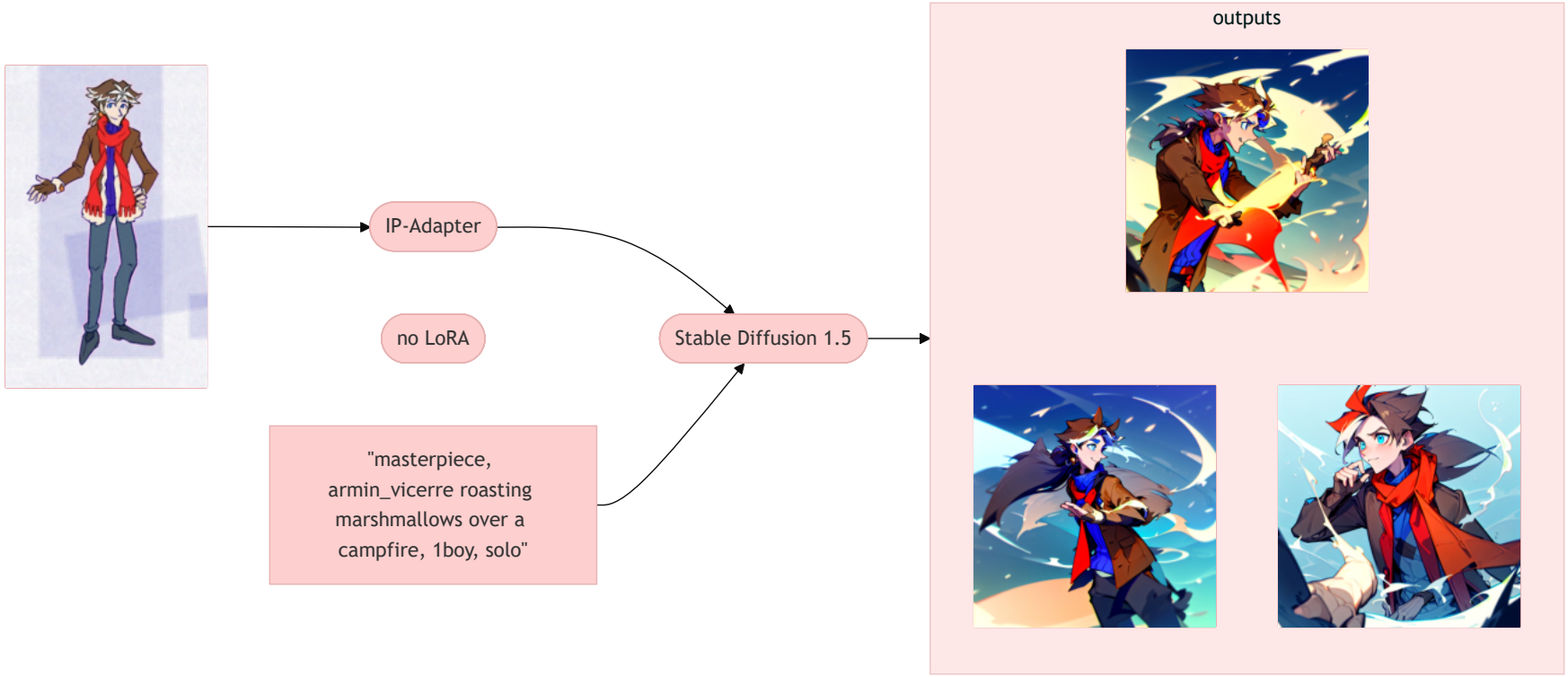
Bonus image generation models

- [2022-12-02] Different Dimension Me
- [2023-02-06] Midjourney
- [2023-03-21] Adobe Firefly
- [2024-02-12] Imagen
- [2024-08-01] Flux
- [2024-12-17] Whisk
- [2025-05-15] DALL·E 3



Bonus image generation technologies

- [2023-12-16] ControlNet: guide output with doodles, depth maps, or poses
- [2024-01-06] IP-Adapter: one-shot image-to-image

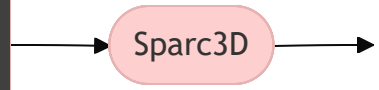


Video generation models?

- A video is just a moving image, right? 🧠

3D modeling models?

- A 3D model is just some math that produces an infinite # of 2D images, right? 🌌



Summary

- Lots of experimentation
- Very fun (and sometimes scary) to see what comes out :)

Where we are today

2025-11-13

2 MONTHS, 15 DAYS LATER

What do image generation models do well?

- Very high character fidelity
- Compositional understanding
- Good odds of "rolling" a good generation

What do image generation models do for me?

- Positive feedback loop: better human art → better training data for image generation models → better human art

What do image generation models do for me?

- Really fun to see your designs come to life

* Illustrations are generated, not the entire card.

What should come next?

- Integration with art programs (e.g. layers)
- Composition/storytelling
- Creativity and delight

That’s it!

- Q&A time
- Link to slides will be shared

Also check out:

- Citations in the source code
- Technical notes in the comments
- Metadata on Stable Diffusion generations