

Guide to Pony Images for creating AI

About

This is a guide for how to label images. Following this guide should result in better AI for pony image generation.

This guide is intended to be used with this website [[LINK](#)].

Thank you in advance for reading.

Section 1

Everything after a ">" means an example caption.



> Pinkie Pie is hugging Pinkamina (sad pinkie pie) and Surprise. White background.

1. Describe the character(s) and what they're doing.

It's essential that you describe each character in the scene and anything that one of them is doing.

This example shows why it's so important.

The characters tags are "pinkie pie, surprise". The tags tell the model there are 2 characters. The image has 3.

What does the model learn? The model learns that any image with “pinkie pie” should sometimes have 2 pinkies.

The interaction tags are “hug”. The tags do not tell the model who is hugging who.



> Princess Cadance, in bed, touching horns with Shining Armor. Princess Cadance is casting a spell. Shining Armor has hearts in his eyes. White background.

2. Describe the scene if it's not normal. If it is normal or can be guessed from the existing prompt then don't worry.

Example:

“Fluttershy helping bandage a bear.”

You do not need to mention grass or forest. The model knows that bears normally appear in forests.

You should mention if the bear is in Fluttershy's hut or Canterlot castle or anything interesting.

The model can infer things like Princess Luna images normally using the nighttime and Princess Celestia images normally using the daytime.

There is no negative to describing scenes in more detail, however writing a list of random objects can get tiresome.



> Chibby Twilight and Sunset Shimmer in bed together. Sunset is concerned and leaning up. Twilight Sparkle is sleeping. Flat shading. Flat background.

3. Increase your artwork standards. Always mention backgrounds that wouldn't exist in real life.

If you do not mention this artwork is flat, then the model will produce artwork that looks like this by default. If there is a flat background or an abstract background, always mention it. If the shading or quality is bad or even just average, mention it. That way the model will always produce high quality images when it isn't told anything.

Things you should be using a lot.

"Flat Background", "Abstract background", "Flat shading", "Sketch", "Monochrome"

(The original stable-diffusion model was trained on real life images. “purple background” doesn’t exist in real life so a purple background should be considered **very** abnormal.)



> Derpy with her front half in a bin. Flat background. Picture of a real life bin in the top left.

4. Picture in Picture, Multiple Scenes, Comics.

If there are multiple camera angles or things that would be impossible for a camera to capture in real life, they need to be mentioned.

In the top left of the above image is another image. That needs to be mentioned and

described.

If you don't mention this then the model will produce random faces and pictures inside the pictures it generates. The model is not smart enough to know when it should and shouldn't generate these pictures inside of pictures.



Hearts like this are normally worth describing. Depends if you think these should appear when the caption you've written is used in the model.



Dialogue does not need to be mentioned, Derpibooru has a tag for talking between characters so it can be handled in post.



Things like TV Screens, Mirrors and Window Reflections are perfectly fine. You do not need to mention things like this.



This image must be manually tagged with “floating text” or something similar. Despite having floating text in the image, derpibooru does not have any text tags for this image or others like it. When in doubt, mention the floating text.

Tips

- The model cannot see more than 70 tokens. A word or punctuation mark uses up a token. Any words or punctuation over the limit are trimmed. The website has a word counter in the bottom left, keep the number under 70.

(More details let you control the model more and slightly improve output quality. Less details make the model better at guessing what should be in the background. Feel free to use as much or as little detail as you want. There is no one correct answer.)

- Use full names to make programmers happy. “Rainbow Dash” is better than “Rainbow” because “rainbow” and “Rainbow” are actually completely different things.
- Capitalize the first letters in names. It helps with name identification.
- If you see a character as a species they’re not normally, label them. Sunset Shimmer should be “**pony** Sunset Shimmer” when in Equestria. Meanwhile Pinkie Pie should be “**human** Pinkie Pie” when in human form. This is not a strict rule, however it helps with transforming characters to different species.



>**human** Dinky Hooves riding **pony** Derpy Hooves while **human** Derpy Hooves happily watches while leaning on a fence. flat shading.

- If you see an image that is tiny like



Do not try to zoom in. The AI cannot zoom in.

Describe it as “extremely wide comic” or “detailed random colors” if you can’t make out any details.

- If you see a character that doesn’t have a name or you don’t know their name, you can describe the character instead.



> Surprise (**white pegasus pinkie pie with yellow mane**) sitting up in bed, barely resisting the urge to hug you with joy. bland shading. Party balloons in the background.

- Labeling signatures is not required.

Contact

Use cookietriplep@gmail.com to send feedback about this document or contact me.

Random

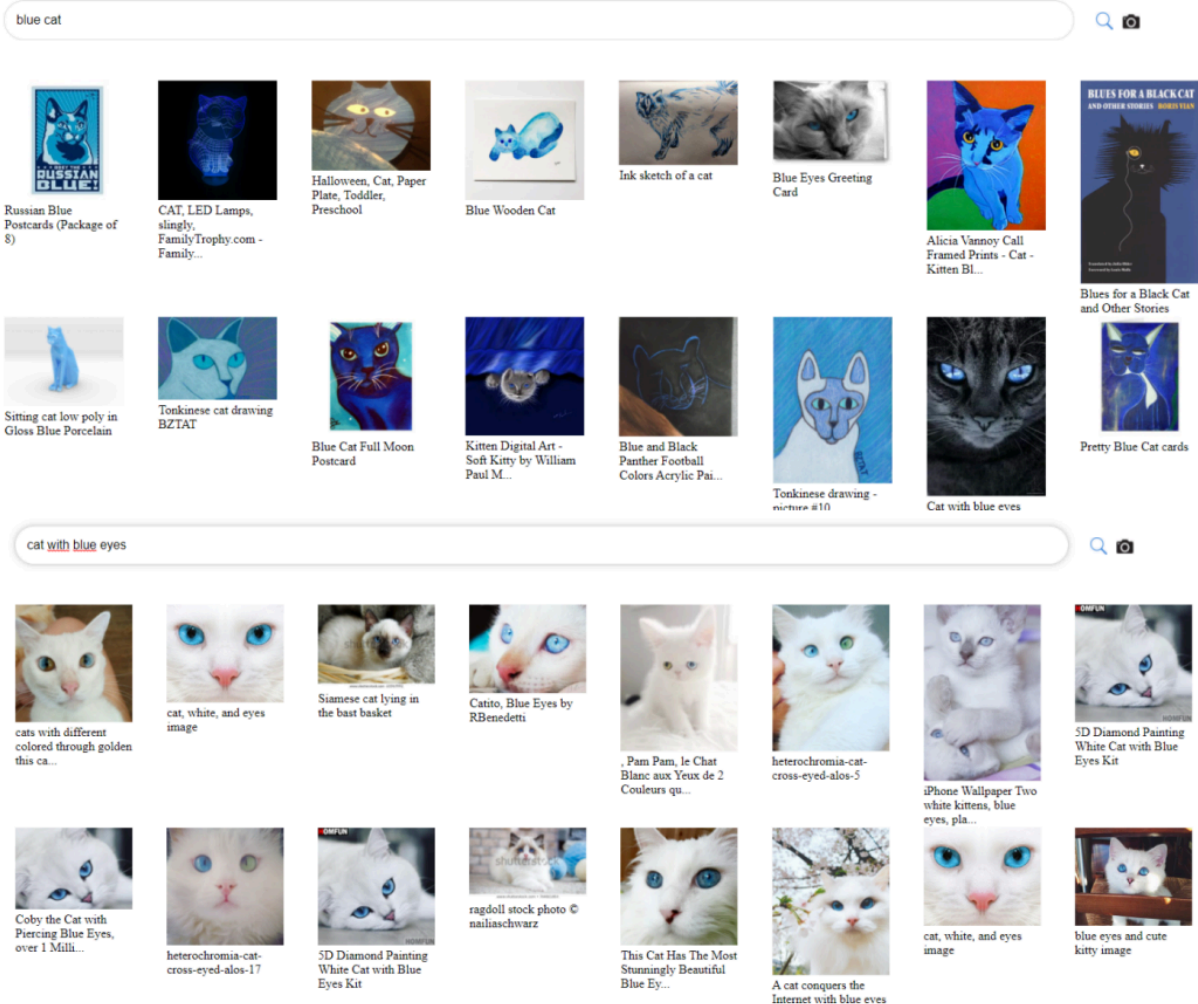


Figure 1: Sample images retrieved from the queries "blue cat" or "cat with blue eyes" in the web demo

OpenCLIP was trained within 0.2% accuracy of the original clip model using terrible captions. I wonder if we could create our own CLIP model that surpasses even the original CLIP?



Describe this image without going over 80 words. You may have to reduce your detail and focus on the most striking/important bits (and that's okay).

Instead of saying "Ninjas with swords, eyepatches and peglegs", you could say "pirate ninja's".