



BlueFaux's

DreamBooth

Guide + Study +promo

Guide: Visions of Chaos - DreamBooth Training

How to train a Stable Diffusion model to use you or an original character using the DreamBooth method locally with Softology's Visions of Chaos.

Study: Visions of Chads - DreamBooth Testing

What is the optimal amount of training steps?	(tl;dr ~ 1500 steps)
How many training images do I need?	(tl;dr ~ 5 minimum)

The Guide and Study are free for you and were funded by me - Promo for my paid services at the end

Table of Contents

1 - Title Page
2 - Table of Contents
3 - Introduction
4 - Prerequisites
5 - Softology and Visions of Chaos
6 - DreamBooth Guide
10 - DreamBooth Study
12 - Study Phase 1
18 - Study Phase 2
24 - Study Phase 3
31 - BlueFaux Studios Promo



The Pandemic pushed many people to distance socially and professionally. The lack of commute times mixed with the lack of things to do led to millions of tech minded individuals to start taking on more hobbies.

This led to the fabled transcendence in the Computer Science sector...



AI models absolutely exploded thanks to becoming an open-source, community bolstered superproject.

Stable Diffusion is a major fruit of this achievement.

But how do you as an individual run Stable Diffusion? How do you train an SD Model to make pictures of yourself instead of generic people?

Thanks to Softology's program, Visions of Chaos, we can set that up. Aaaaaand Good thing you have me as a teacher. I will link/show you how to:

- 1.) Get Standard Diffusion running locally.
- 2.) Train a model using your own dataset.
- 3.) Set the optimal parameters for training.



Guide: Visions of Chaos DreamBooth Training

What you need BEFORE you start:

1.) Visions of Chaos with Machine Learning Mode

- a.) Visions of Chaos is a program dedicated to bringing open source simulation and AI software to the general public.
- b.) Enabling Machine Learning Modes is a long process that involves downloading several programs (like Python and VS) so your computer can run the AIs.

2.) A CUDA enabled GPU with **24GB+** of VRAM

- a.) This training method generally takes just under 21GB of VRAM.

3.) A folder with 512x512 images of your subject

- a.) Visions of Chaos offers several different methods to resize your images, but stretching a non square image will give you a distorted picture so be careful

4.) A <token> and a <class> for the training.

- a.) The <token> should be something unique. For the study, I am using Chad, so to make that unique and easily remembered, I am going to use TheChad
- b.) The <class> is the general category that the subject belongs to. For TheChad, I could use man, guy, person, dude... to be honest, you could use anything because it will still train towards your dataset, but the <class> is what generates your Regularization images.

5.) Enough space to actually save the models.

- a.) A full, unpruned model is over 11GB.
- b.) A pruned model is over 2GB.
- c.) If you save checkpoints along the way, the autopruned checkpoints are over 4GB

Softology

The science of software

Information and Links

Website:

[Softology](#)

Visions of Chaos Download:

[Softology - Visions of Chaos](#)

Step-by-Step for Enabling Machine Learning Mode:

[Softology - Visions of Chaos - Machine Learning Support](#)

“Softology is my way to entertain, share knowledge and inspire others. This site and Visions of Chaos are produced as a hobby by one person and not as a company as some people have assumed.”

-Jason Rampe

Contact Infos

[Discord](#) - [Twitter](#) - [Email](#) - [Reddit](#)

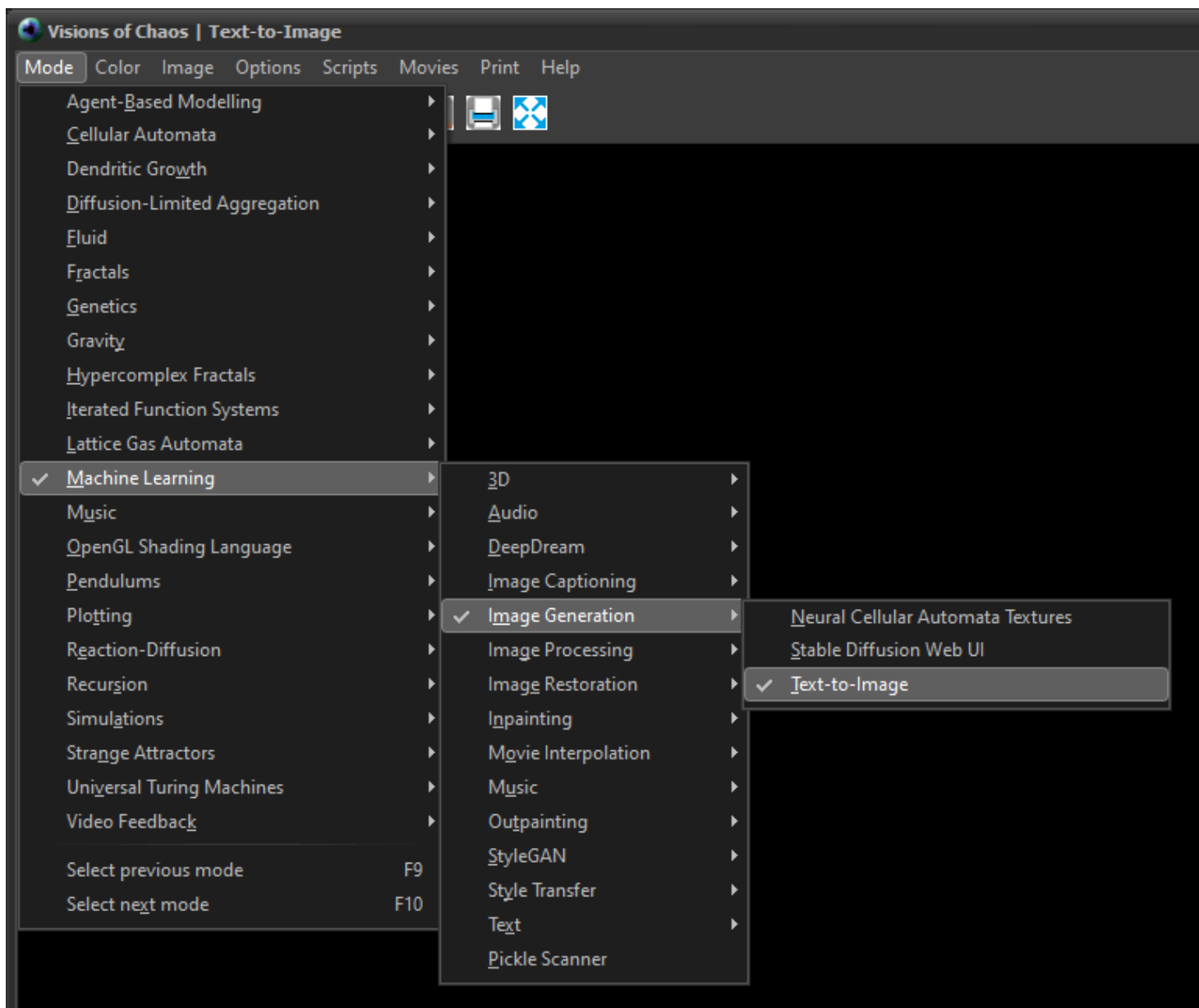
BlueFaux is not affiliated with Softology, but we did contact Jason for permission to plug his amazing works and ways to contribute for those who want to show appreciation for an altruistic tech philanthropist instead of a Capitalist Entity which exists only to make its shareholders an easy buck.

[Patreon](#) - [PayPal](#)

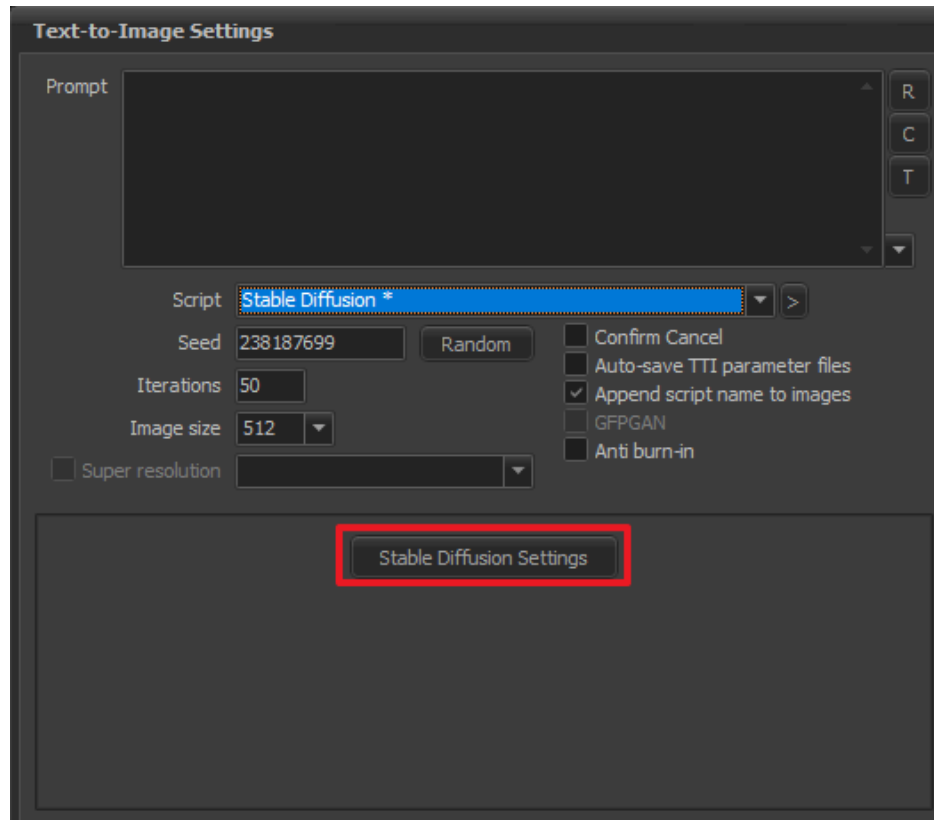
1.) Open Visions of Chaos



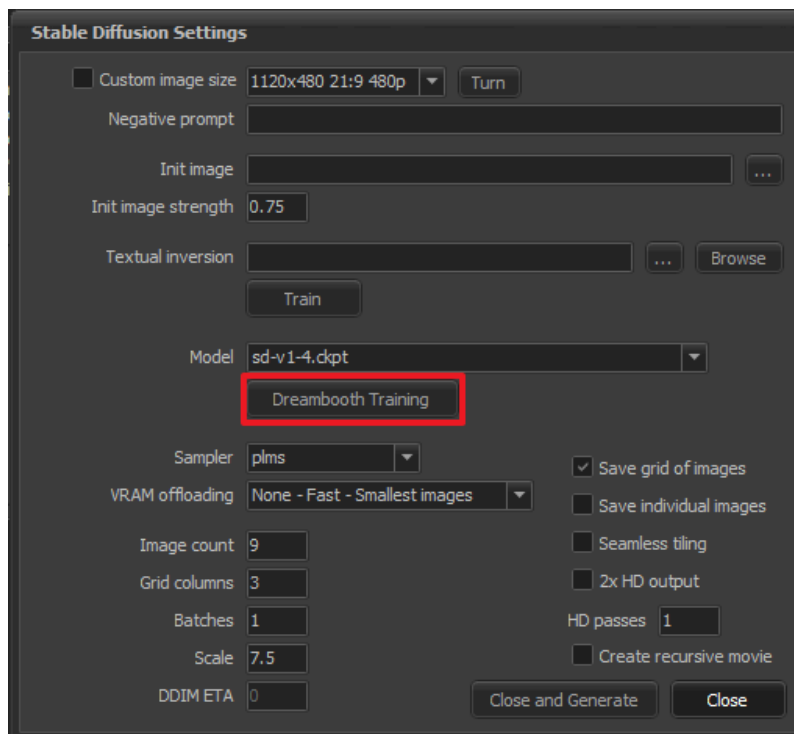
2.) Using the top menu, select [Mode] -> [Machine Learning] -> [Image Generation] -> [Text-to-Image]



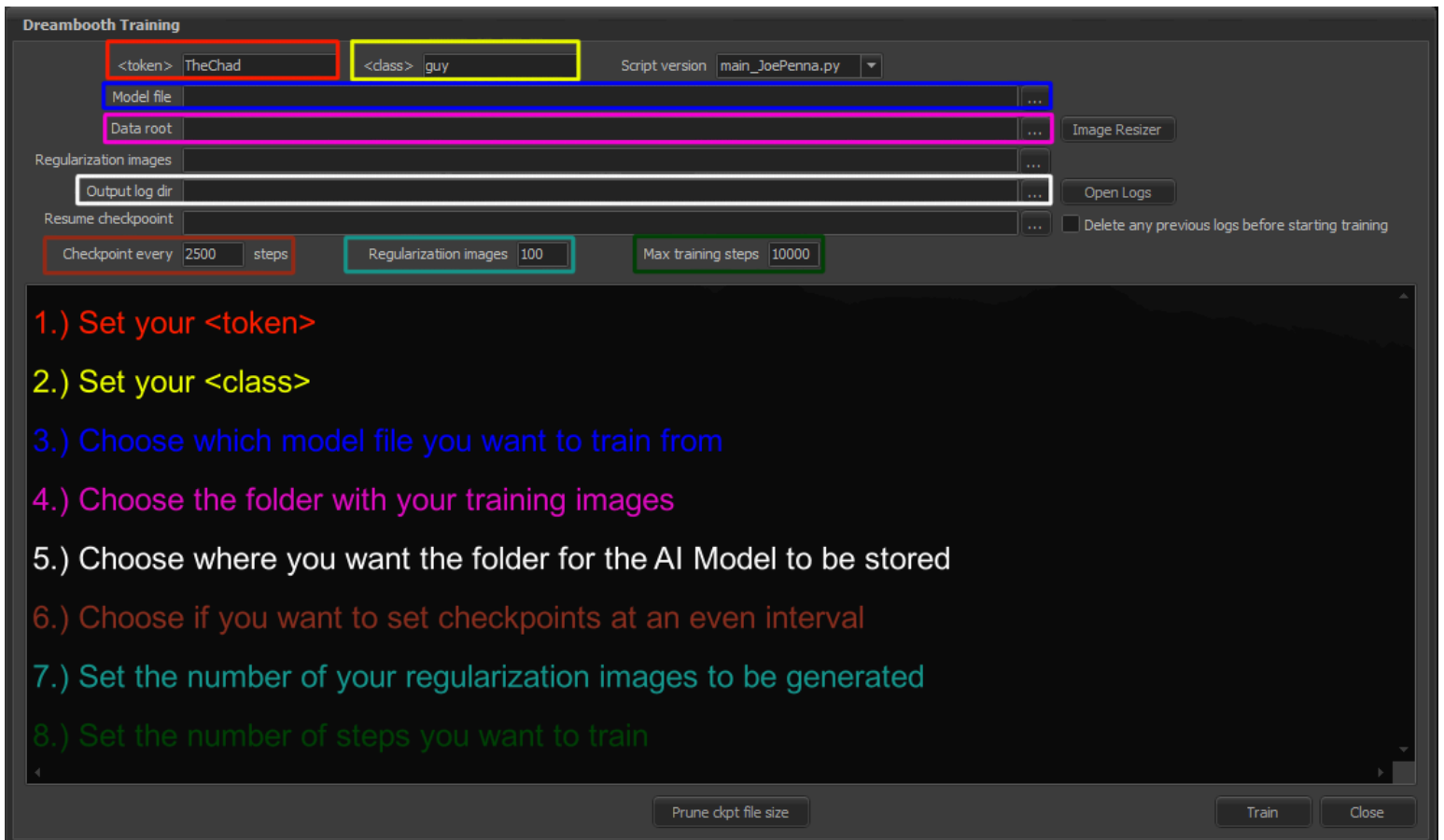
- 3.) Select “Stable Diffusion” from the [Script] drop menu
- 4.) Go to the [Stable Diffusion Settings]



- 5.) Select [Dreambooth Training]



6.) Follow the color coded directions below



1.) Set your <token>

2.) Set your <class>

3.) Choose which model file you want to train from

4.) Choose the folder with your training images

5.) Choose where you want the folder for the AI Model to be stored

6.) Choose if you want to set checkpoints at an even interval

7.) Set the number of your regularization images to be generated

8.) Set the number of steps you want to train

1.) Set your <token>

a.) This is the unique keyword you want to call on your training

2.) Set your <class>

a.) This is the category of your training subject

3.) Choose which model file you want to train from

a.) This will be a .ckpt file

4.) Choose the folder with your training images

a.) They need to be 512 x 512 (5 images should be your lower bound, but try to get as many quality images as you can for the dataset. Over 50 is excessive/redundant for a character)

5.) Choose where you want the folder for the AI Model to be stored

a.) Ensure this drive will not run out of space

b.) A fully trained, unpruned model is over 11GB

6.) Choose if you want to set checkpoints at an even interval

a.) This will save a partially pruned .ckpt file at each of those intervals

b.) A partially pruned .ckpt model is over 4GB

7.) Set the number of your regularization images to be generated

a.) Regularization images just basically generate images using your selected <class> as a prompt to help with training, guy uses "guy" as a prompt to generate the images

8.) Set the number of steps you want to train

a.) Higher doesn't mean better. Stick with 1500 steps for characters unless you are experienced

7.) Train

- Press [Train].
- This should take less than 1.5 hours for an RTX 3090 for 1500 steps and 100 regularization images.
- Do not use your GPU for other tasks during this training as it may crash.

8.) Prune

- Press [Prune] and select the last.ckpt file from the folder that was created in your output directory.
- This only takes a few seconds, but it lowers the model size by almost 10GB.

9.) Rename

- Name your model something that you will recognize.
- I usually follow the format “[token]_[class]_[#ofsteps].ckpt ” or “TheChad_guy_1500.ckpt”

10.) Move the model to your AI's Model Directory

- The following will show the base Path for three separate AIs that VoC has

Stable Diffusion - (C:\Users\[username]\AppData\Roaming\Visions of Chaos\Examples\MachineLearning\Text To Image\Stable Diffusion\models\ldm\stable-diffusion-v1)

Deforum Stable Diffusion - (C:\Users\[username]\AppData\Roaming\Visions of Chaos\Examples\MachineLearning\Text To Image\Deforum Stable Diffusion)

Automatic1111's Stable Diffusion WebUI - (C:\Users\[username]\AppData\Roaming\Visions of Chaos\Examples\MachineLearning\Text To Image\stable-diffusion-webui\models\Stable-diffusion)

11.) Use the Model to generate your images

- The the format for the prompt should follow **<token>** **<class>** format
- This means something like “a photo of **TheChad** **guy** eating a burrito”

~FIN~



How many steps do I use?

How many images do I need?

*We are no
longer
cave
mans*



*We is
modern
mans...*



With
SCIENCE



*We can command
The Universe to tell us its
secrets if we know the right
questions to ask...*

So let's ask...

Study: Visions of Chads - DreamBooth Training

Overview:

I have been using DreamBooth for my own projects since it came out, but I haven't seen an answer to several questions I have about it. To answer these myself, I am doing this study to find out the best ranges for some of the parameters to train a Stable Diffusion .ckpt model file for unique characters.

Method:

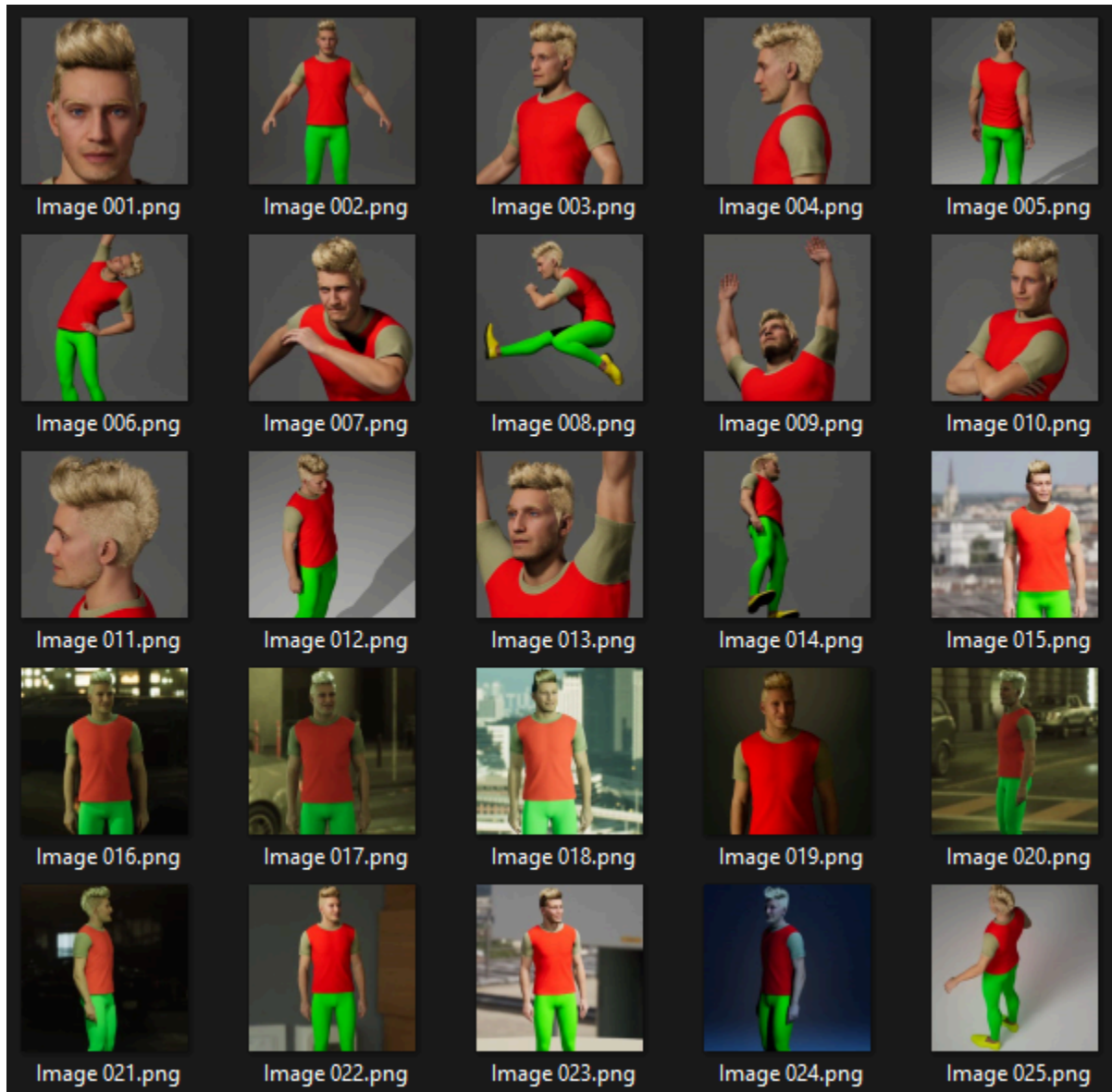
- ★ Phase 1 - What is the best range for Training Steps and Diffusion Steps?
 - 5 separate models with a different number of Training Steps will be trained on the same dataset
 - The dataset has 25 Images of Chad wearing the same outfit in different environments and poses.
 - These models will be used to generate images with a different number of Diffusion Steps
 - There will be 5 tables showing the results according to a specific prompt
- ★ Phase 2 - How many images do I need for the Training Dataset?
 - 5 separate models will be trained for 1500 Training steps
 - These models will use a different number of images in their Training Dataset
 - There will be 5 tables showing the results according to a specific prompt
- ★ Phase 3 - Trained model showcase on 3 Prompts
 - This uses 1500 steps on the full 50 image dataset
 - Each Prompt will have a table of fixed seeds for One-Shot images.

TL;DR Results:

- ★ Phase 1
 - Between 1,000 and 2,500 training steps seem to give the best results.
- ★ Phase 2
 - While you can get away with just one image in your Training Dataset, more images generally seem to make the model better, but every model is usable, surprisingly.
- ★ Phase 3 - Showcase
 - One-Shot images are generally never good enough for a final product, but do display the fitness of the trained model.
 - Final image examples are also included to showcase the trained models to their full potential.



















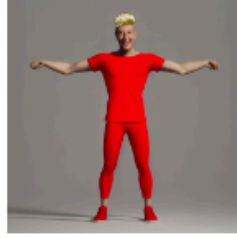






Phase One

- ★ This phase will explore the effect of Training Steps on Output quality
- ★ The training will be done on the same set of (25) 512x512 images of my Chad [MetaHuman](#)
 - MetaHuman is a tool from Epic to make “realistic” 3D models for free.
- ★ There will be 5 models trained from this image set with the Training Steps of [100], [500], [1,000], [2,500], [5,000]



- ★ There will be no control group because 0 steps is just the base model that is not even trained on the keyword. How the base model interprets a novel keyword is pretty chaotic and not very useful for this study.
- ★ The results are shown in the tables on the following 5 pages

Phase 1, Table 1

Prompt 1: "TheChad guy flexing"		Sampling Method: Euler a	512 x 512 image	CFG Scale: 7	SEED: 42069
	100 Training Steps	500 Training Steps	1000 Training Steps	2500 Training Steps	5000 Training Steps
20 Diffusion Steps					
50 Diffusion Steps					
75 Diffusion Steps					
100 Diffusion Steps					
150 Diffusion Steps					

Notes:

- 1.) 100 steps is undertrained
- 2.) 1000 - 2500 Training Steps seems best for model fitness
- 3.) 75 Diffusion Steps seems best for these one-shot generations
- 4.) 5000 steps is very close to just outputting a screenshot of the 3d model in the virtual studio environment.

Phase 1, Table 2


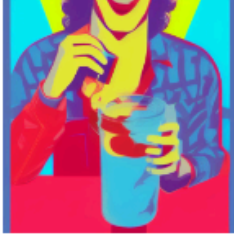



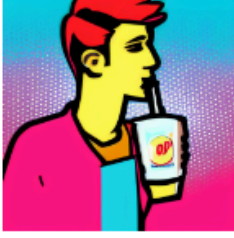
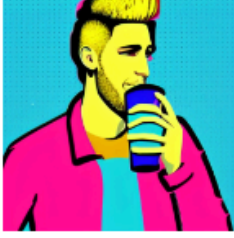

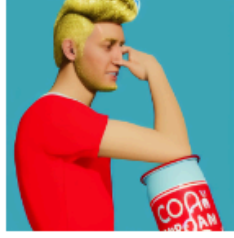

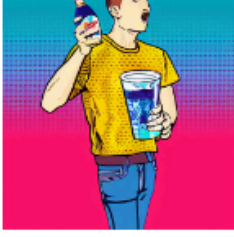



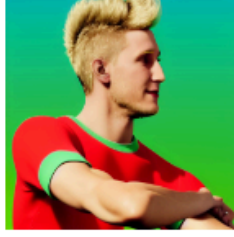



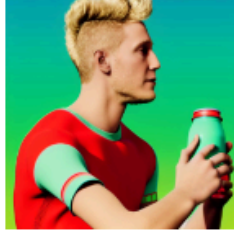
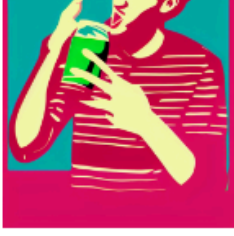
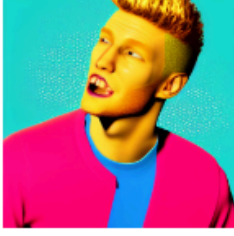



Prompt 2: "TheChad guy wearing sunglasses on top of a skyscraper in the style of GTA V splash art"		Sampling Method: Euler a	512 x 512 image	CFG Scale: 7	SEED: 42069
	100 Training Steps	500 Training Steps	1000 Training Steps	2500 Training Steps	5000 Training Steps
20 Diffusion Steps					
50 Diffusion Steps					
75 Diffusion Steps					
100 Diffusion Steps					
150 Diffusion Steps					

Notes:

- 1.) Notice how it loses the fidelity of the non trained part of the prompt. At 5000 training steps, it looks just like the MetaHuman style 3D model instead of the 2D splash art style. This is an example of an overfitted model. If you add (emphasis) or (weight:1.5) to the nontrained parts of the prompt, you can still use an over trained model.

This means "TheChad guy wearing sunglasses on top of a skyscraper (in the style of GTA V slash art)"

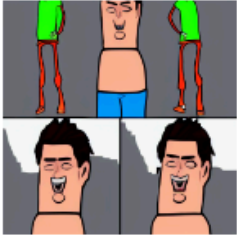







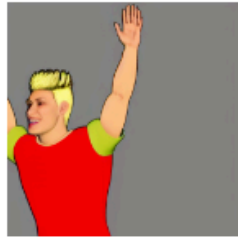

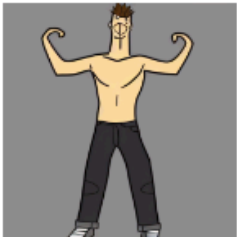

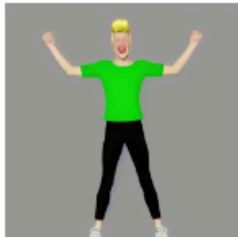
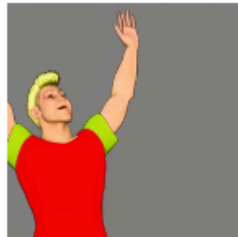



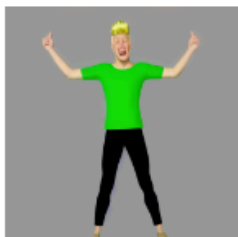
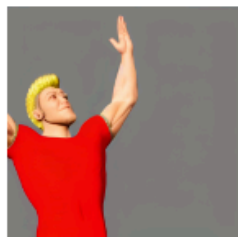






Phase 1, Table 3

Prompt 3: "TheChad guy drinking a soda, digital pop art poster, vaporwave aesthetic"		Sampling Method: Euler a	512 x 512 image	CFG Scale: 7	SEED: 42069
	100 Training Steps	500 Training Steps	1000 Training Steps	2500 Training Steps	5000 Training Steps
20 Diffusion Steps					
50 Diffusion Steps					
75 Diffusion Steps					
100 Diffusion Steps					
150 Diffusion Steps					

Notes:

- 1.) 1000 Training Steps is the clear winner for One-Shots.
- 2.) The 500 step model is looking pretty good as well.
- 3.) The Vaporwave aesthetic is strong enough to influence the overtrained 5000 step model.














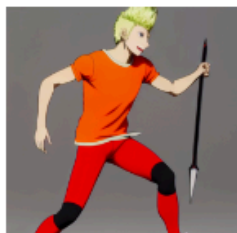

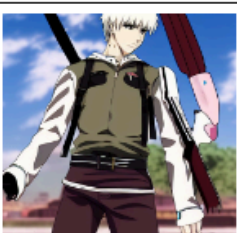
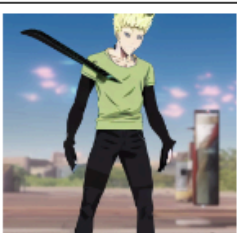

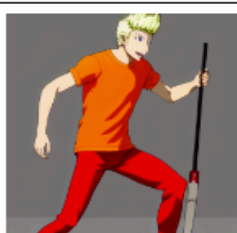

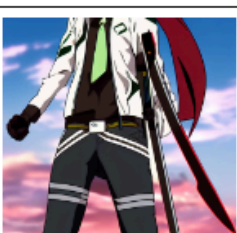


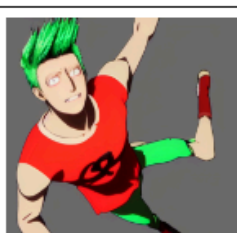
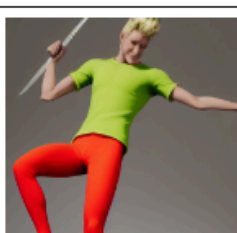
Phase 1, Table 4

Prompt 4: "TheChad guy flexing ironically as a new meme, drawn in ms paint"		Sampling Method: Euler a	512 x 512 image	CFG Scale: 7	SEED: 42069
	100 Training Steps	500 Training Steps	1000 Training Steps	2500 Training Steps	5000 Training Steps
20 Diffusion Steps					
50 Diffusion Steps					
75 Diffusion Steps					
100 Diffusion Steps					
150 Diffusion Steps					

Notes:

- 1.) This one is interesting because it seems to wrestle between making it MS paint or MetaHuman and not in between. There are probably some underlying token interactions that can be avoided with changing the prompt

Phase 1, Table 5

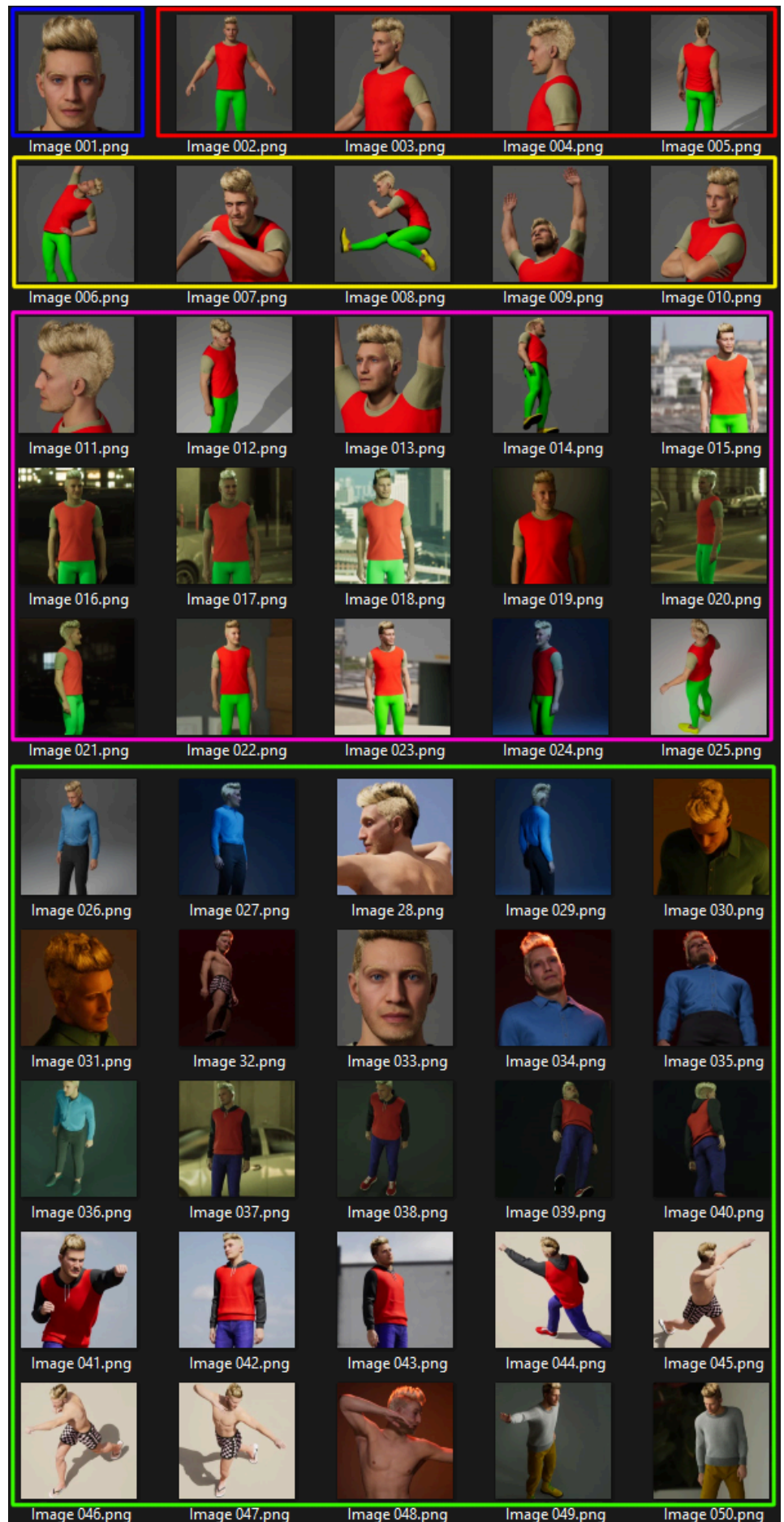
Prompt 5: "TheChad guy as an anime character using ODM gear to slay a titan (by Hajime Isayama)"		Sampling Method: Euler a	512 x 512 image	CFG Scale: 7	SEED: 42069
	100 Training Steps	500 Training Steps	1000 Training Steps	2500 Training Steps	5000 Training Steps
20 Diffusion Steps					
50 Diffusion Steps					
75 Diffusion Steps					
100 Diffusion Steps					
150 Diffusion Steps					

Notes:



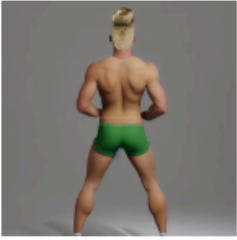

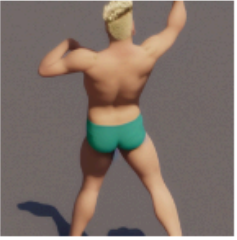



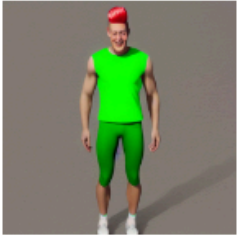
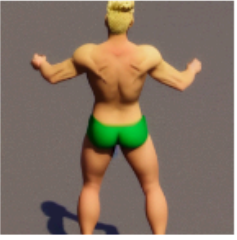



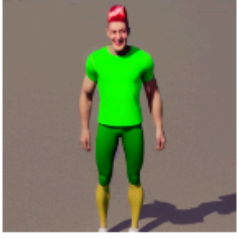











- 1.) 500 - 2500 training steps are all very good
- 2.) Notice how the background loses fidelity as it becomes overtrained. This can be fixed with an inpainting mask in post or by changing prompt and weights.

Phase Two

- ★ This phase will explore the effect of [# of training images] on output quality.
 - ★ The training will be done at 1500 steps with a cumulatively increasing dataset of 512x512 images of my Chad MetaHuman.
 - ★ The Models and their datasets are color coded, cumulative, and shown here —————>
- Model 1
 - 1 Image
 - Frontal face
 - Model 2
 - 5 Images
 - Neutral spread
 - Model 3
 - 10 Images
 - Dynamic Poses
 - Model 4
 - 25 Images
 - Different Environments
 - Model 5
 - 50 Images
 - Different Outfits
- ★ The results are shown in the tables on the following 5 pages


















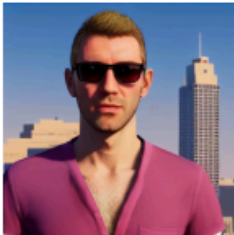


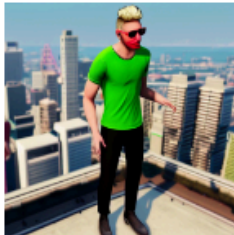




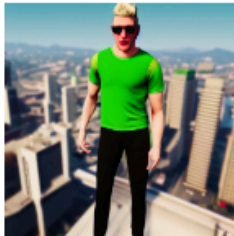
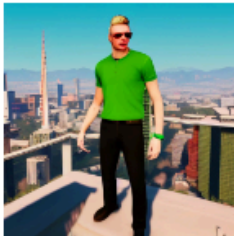
Phase 2, Table 1

Prompt 1: "TheChad guy flexing"		Sampling Method: Euler a	512 x 512 image	CFG Scale: 7	SEED: 42069
	1 Image Data Set	5 Image Data Set	10 Image Data Set	25 Image Data Set	50 Image Data Set
20 Diffusion Steps					
50 Diffusion Steps					
75 Diffusion Steps					
100 Diffusion Steps					
150 Diffusion Steps					

Notes:

- 1.) Model 1 is not able to produce OneShots of the prompt.
- 2.) All other models can produce decent OneShots.
- 3.) No other background is coming out in Model 4.
- 4.) Model 5 with different outfits is producing prompt appropriate clothing.

Phase 2, Table 2

Prompt 2: "TheChad guy wearing sunglasses on top of a skyscraper in the style of GTA V splash art"		Sampling Method: Euler a	512 x 512 image	CFG Scale: 7	SEED: 42069
	1 Image Data Set	5 Image Data Set	10 Image Data Set	25 Image Data Set	50 Image Data Set
20 Diffusion Steps					
50 Diffusion Steps					
75 Diffusion Steps					
100 Diffusion Steps					
150 Diffusion Steps					

Notes:

- 1.) Who is this man in Model 1? This result shows me that the previous prompt could produce more than just the face, so technically it would be usable, just would take more generations to hone into CHAD
- 2.) All other models are good enough to use.


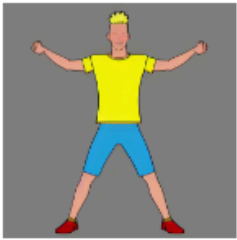























Phase 2, Table 3

Prompt 3: "TheChad guy drinking a soda, digital pop art poster, vaporwave aesthetic"		Sampling Method: Euler a	512 x 512 image	CFG Scale: 7	SEED: 42069
	1 Image Data Set	5 Image Data Set	10 Image Data Set	25 Image Data Set	50 Image Data Set
20 Diffusion Steps					
50 Diffusion Steps					
75 Diffusion Steps					
100 Diffusion Steps					
150 Diffusion Steps					

Notes:

- 1.) Lol all models are good enough to use, even Model 1 in a limited capacity.

Phase 2, Table 4

Prompt 4: "TheChad guy flexing ironically as a new meme, drawn in ms paint"		Sampling Method: Euler a	512 x 512 image	CFG Scale: 7	SEED: 42069
	1 Image Data Set	5 Image Data Set	10 Image Data Set	25 Image Data Set	50 Image Data Set
20 Diffusion Steps					
50 Diffusion Steps					
75 Diffusion Steps					
100 Diffusion Steps					
150 Diffusion Steps					

- Notes:
- 1.) Model 1 face again.
 - 2.) Model 5 is actually able to produce the MS Paint look more consistently.
 - 3.) Model 5 is the best.

Phase 2, Table 5

Prompt 5: "TheChad guy as an anime character using ODM gear to slay a titan (by Hajime Isayama)"		Sampling Method: Euler a	512 x 512 image	CFG Scale: 7	SEED: 42069
	1 Image Data Set	5 Image Data Set	10 Image Data Set	25 Image Data Set	50 Image Data Set
20 Diffusion Steps					
50 Diffusion Steps					
75 Diffusion Steps					
100 Diffusion Steps					
150 Diffusion Steps					

- Notes:
- 1.) All models are usable.
 - 2.) Model 5 is the best.

Phase Three

While these AI Models can produce great generalized OneShots, the meat of any art medium is how to get good outputs for your work consistently. OneShots are not good at producing very specific images consistently yet. You could generate 1500 images and not have the right facial expression, body position, etc. HOWEVER, there are many ways to guide Stable Diffusion to

produce highly specific results and using any of them will push your digital art results to a higher tier of quality.



Using an Initial Image - In this method, you use one of your generated images as a base image for another generation. There will be an option that controls the amount of noise it adds to increase or decrease the variance of the next image generated.










Inpainting - In this method, you paint a mask to tell it where to actually change the image. It still has a noise schedule to control the variance. This is one of the most useful tools in the Stable Diffusion arsenal right now due to its ease of use for someone who is not experienced in digital art programs.

Manual Image Editing - Krita, GIMP, or any other image editing or drawing program used in conjunction with Stable Diffusion by a traditionally trained/experienced digital artist is absolutely insane. With Inpainting, the AI still has trouble drawing hands. Too many fingers, not enough fingers, impossibly positioned fingers..... AI sucks at hands, man.... But if you quickly bring an output over to Krita, draw a crude, but anatomically accurate hand, then run it through Stable Diffusion again as an Initial Image, the hand will take a moment to fix instead of playing the RNG game with the other methods. Obviously this improves every aspect of controlling your output quality.

This Phase just showcases some OneShots of Three more refined prompts, but also generates some higher quality final images to showcase the “Optimally” trained Chad model.

I used the best training parameters from Phase 1 (1500 Training Steps) and Phase 2 (50 Image Dataset) on 15 different seeds for 50 Diffusion Steps.

They are displayed on the next six pages. . . .

Prompt 1: "TheChad guy as a medieval fantasy knight looking into the camera, he is photogenic, (character concept art), dnd, d&d, dungeons and dragons"		Sampling Method: Euler a	512 x 512 image	CFG Scale: 7	1500 Training Steps 50 Diffusion Steps
1		227		1337	
Seed	Image				
10		2227		808813	
101		22227		314159265	
10101		222227		161803398	
1010101		2222227		271828182	


As you can see, Stable Diffusion is not amazing at doing weapons, but it can do armor very well. Employing external drawing programs to draw the weapon you want and then iterating through SD again is probably the best way to get the kind of weapon you want.

When farming for a base composition, generally AI Artists will make a batch of random seeds and then work from the best image for that.



For these images, I only used Initial Images and Inpainting. As you can see, the weapons are still pretty bunk. The spear is probably the best, but the generic Sting next to it is laughably awkward.

I really recommend using an external drawing program to help with this issue.
















Prompt 2: "TheChad guy as an anime protagonist from Attack on Titan, by Hajime Isayama, (TheChad guy is zooming by while doing a spin attack through the air, Attack on Titan style)"		Sampling Method: Euler a	512 x 512 image	CFG Scale: 7	1500 Training Steps 50 Diffusion Steps
1		227		1337	
Seed	Image				
10		2227		808813	
101		22227		314159265	
10101		222227		161803398	
1010101		2222227		271828182	

To you, 2227 years from now... The same AI that was trained on a 3D model of a photorealistic human can capture the anime style, even in OneShots, very well. BUUUUT yet again, these frikken weapons are BUNK. barely any ODM gear as well.

I do like these results though.



For the left images (top and bottom), I brought the images to Krita to draw some crude swords on the top one and some ODM gear on the bottom. All four of these images were made in less than 30 minutes total of work. They aren't perfect, but I am just giving you a proof of concept.

Prompt 3: "a photo of TheChad guy wearing a business suit, professional studio photography (half body shot), ultra realistic skin detail"		Sampling Method: Euler a	512 x 512 image	CFG Scale: 7	1500 Training Steps 50 Diffusion Steps
1		227		1337	
Seed	Image				
10		2227		808813	
101		22227		314159265	
10101		222227		161803398	
1010101		2222227		271828182	

Even though the Chad MetaHuman is a “photorealistic” model, these OneShots are expressing aspects of photorealism that the 3D Rendering can’t achieve easily.



I bet that playing with the prompts longer could have made a better photorealism look, but these are already very good after such a short time of work. The most glaring thing is the saturation of his eyes. You can literally add “desaturated eyes” to your prompt to tackle this specific issue.

Remember to add literal adjustments like that to your prompt if you keep noticing the same aspect of your outputs are sticking out in a bad way.

BlueFaux Studios

BlueFaux Studios is dedicated to advancing visual mediums in the digital realm. Our services will expand over time, but as of Nov 26, 2022, here are our current services:

Services	Description	Uses	Deliverables
\$50 DreamBooth Training	We will train a Stable Diffusion .ckpt model to your specs with the data set you provide.	- This is for AI Artists who do not have a GPU that can train .ckpt files, but want to make a personal model.	1 pruned .ckpt file
\$50 Personal Image	We will produce a High Quality Trained Digital Image based on a dataset you provide. (\$30 per extra Image)	- Personal Social Media photos - Personalized Character Concept arts - Wall Arts	1 Personal Use Digital Image
\$200 Commercial Package	We will produce 5 High Quality Trained Digital Images based on a dataset you provide. (\$50 per extra Image)	- Business Social Media - Illustration - Art for products - Game Arts - Promotional Materials	5 Commercial Use Digital Images
(5x modifier) Any NSFW Content	Any NSFW content will have a 5x price modifier and require additional vetting.	- DreamBooth Training on a NSFW Dataset - Personal Use Illustrations with nudity or violence - Commercial Use Illustrations with nudity or violence	Legal NSFW content

Commission Inquiries - bluefaux.surrealism@gmail.com

The Stable Diffusion License can be read here: <https://huggingface.co/spaces/CompVis/stable-diffusion-license>

BlueFaux Studios LLC will deny any request of service for content breaking the Stable Diffusion License.

* Prices are for my time. All of this stuff is free and open source, so you can do it for free.

** This training method is basically obsolete now already, I will be working on EveryDream when I can

Meet BlueFaux

BlueFaux is a recently disabled veteran who turned into a tech enthusiast. She is now devoting her life to the beautiful field of Computer Vision. The Pandemic caused a massive leap in this field and allowed people like BlueFaux to enter the field and contribute meaningfully as a layman. She intends to do this through sharing resources, information, and mentorship to people looking to get into the field.

While only part of a small network of other disabled veterans and AI Enthusiasts, she intends to expand her network to more meaningfully impact these communities in a positive way in the future.

If you want to support someone who is passionate about demystifying AI Art and democratizing it to the masses, please consider becoming a Patron to support her and gain access to her services.



<https://www.patreon.com/BlueFaux>