

Sep 30th Research Update: First attempt at fine-tuning

Process ([tutorial](#)):

- 1) Install NVIDIA CUDA compiler (nvcc)
- 2) Clone [git repository](#) of scripts for preparing data and training model.
- 3) Create virtual environment and install pytorch & related packages
- 4) I skipped the section on generating image captions. I tried it but they ended up being poor captions. ChatGPT with context could do a far better job.
- 5) Prepare the latents from input.
- 6) Fine-tune (1 hour)

Fine Tuning Hyperparameters & Settings

Input: 28 screenshots of grid UI layouts, light and dark themes

Learning rate: $5e-6$

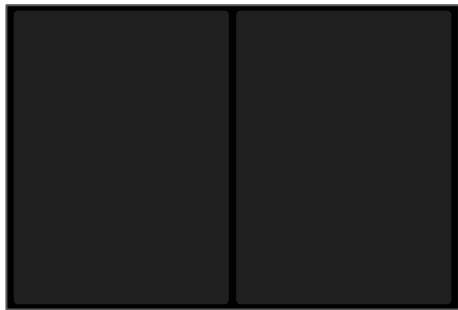
Batch size: 1

36 epochs, around 1000 “training steps”, took an hour

- I saved a checkpoint at 500 steps, and at the end.
- Training accuracy plateaued around halfway through training

I'm unsure if it used multiple GPUs in training or just 1.

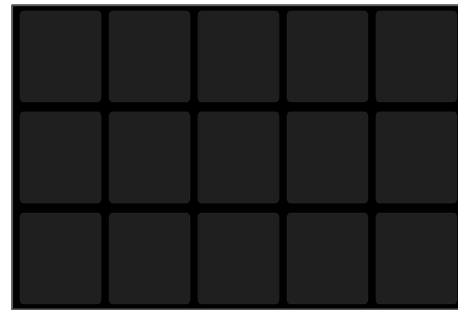
Training Data Examples



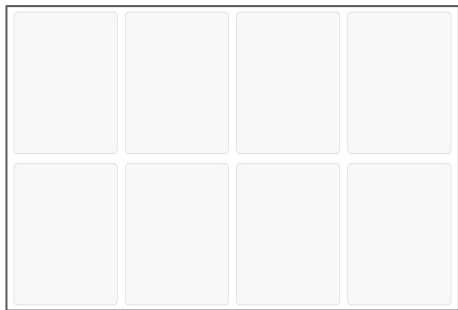
1 row of 2 columns with a dark theme



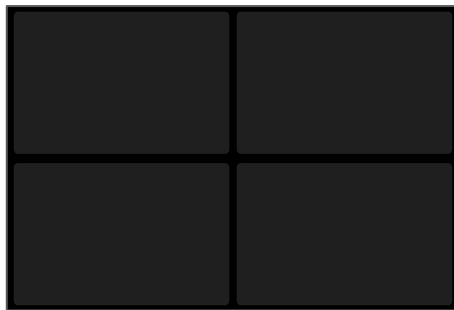
3 rows of 1 column with a light theme



3 rows of 5 columns with a dark theme



2 rows of 4 columns with a light theme



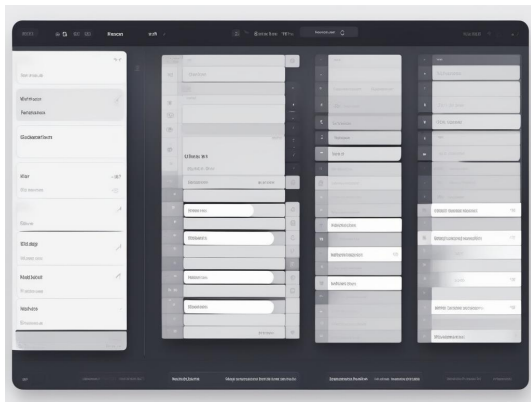
2 rows of 2 columns with a dark theme



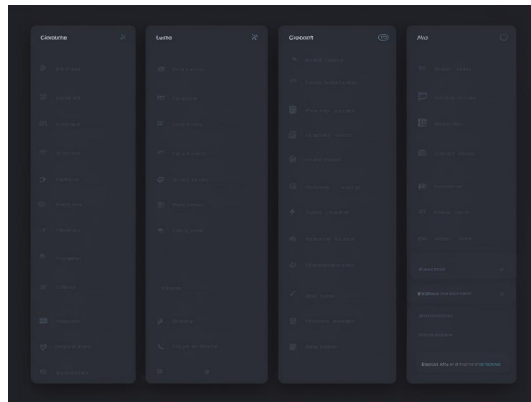
2 rows of 3 columns with a light theme

Example Prompt: “A user interface with 2 rows and 2 columns with a dark theme”

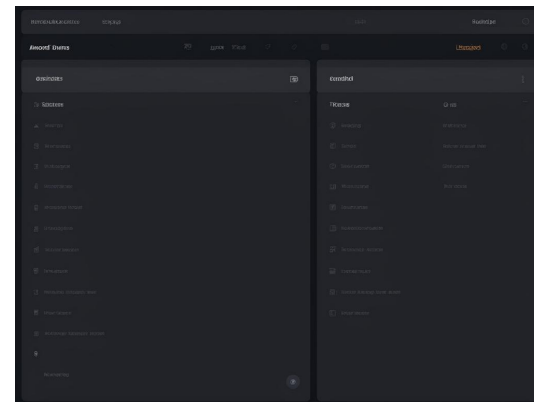
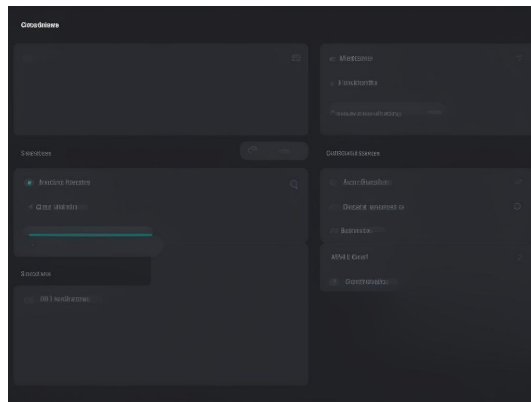
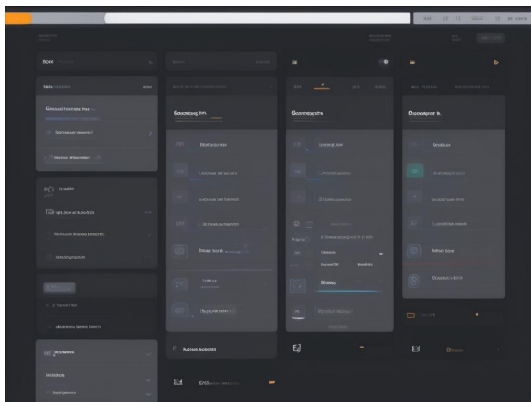
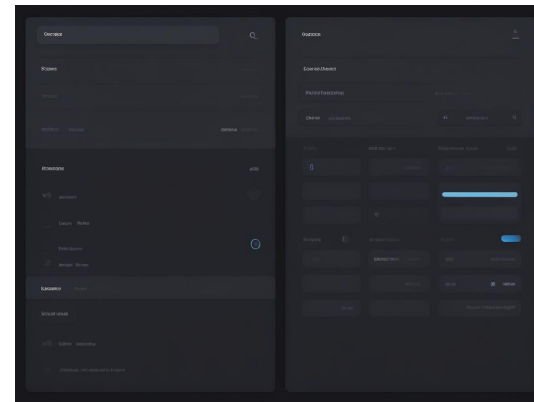
Base Model



After 18 Epochs

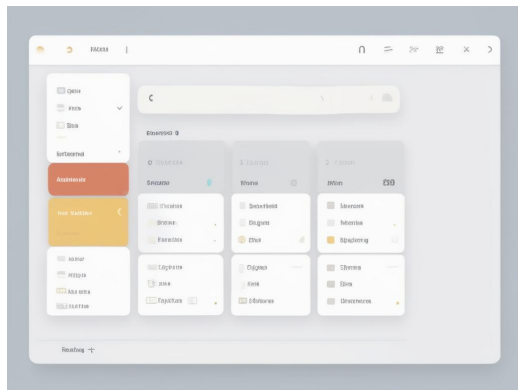


After 36 Epochs

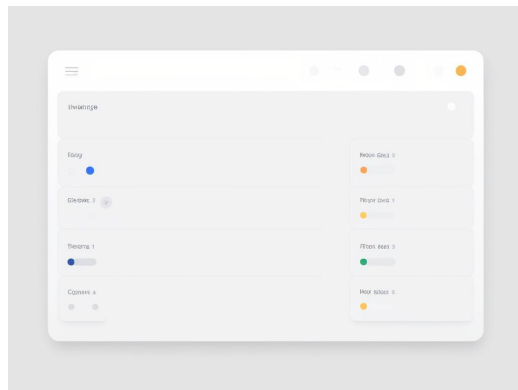


Example Prompt: “A user interface with 2 rows and 2 columns with a light theme”

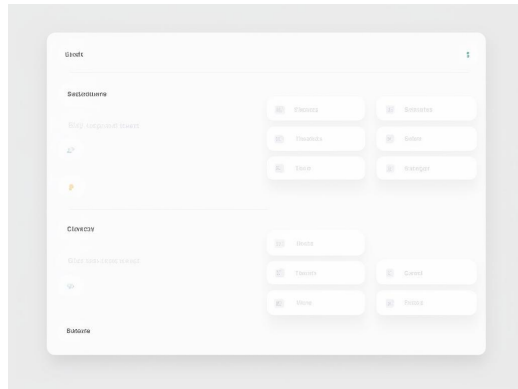
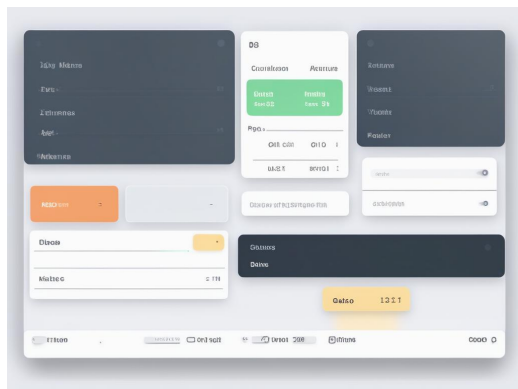
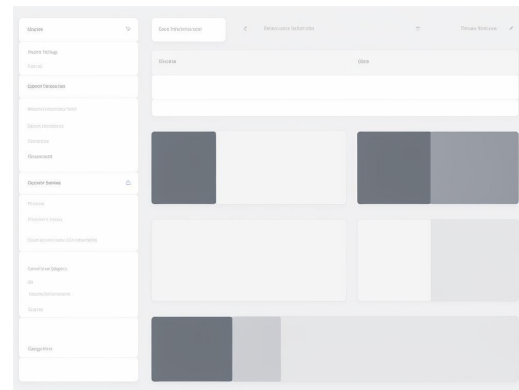
Base Model



After 18 Epochs

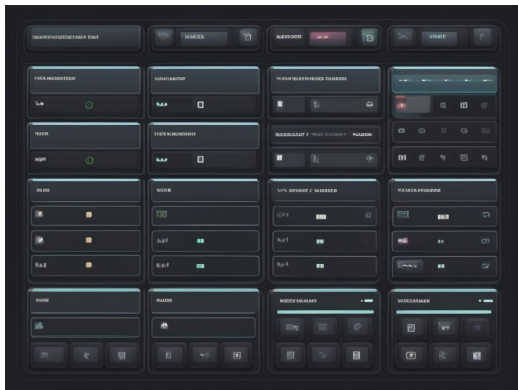


After 36 Epochs

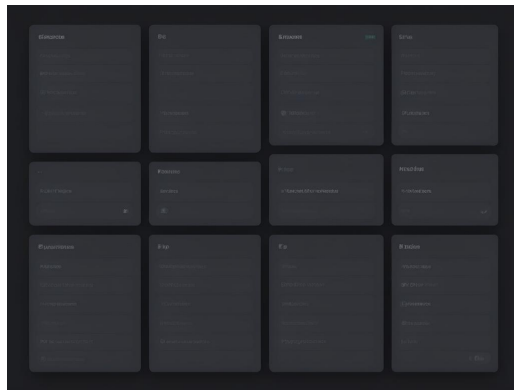


Example Prompt: “A user interface with 4 rows and 4 columns with a dark theme”

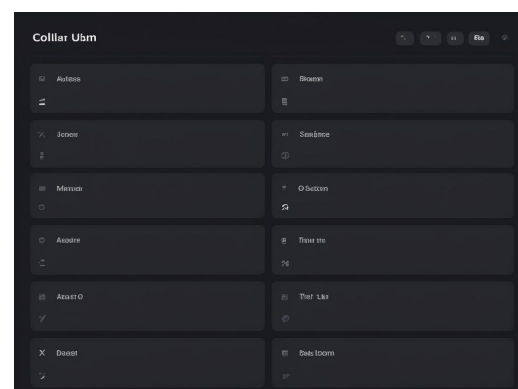
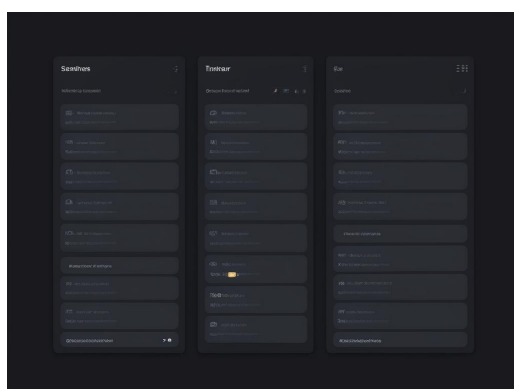
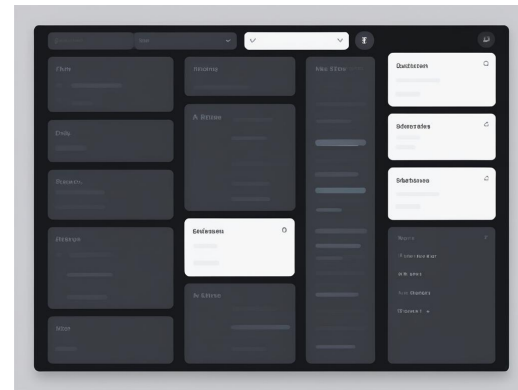
Base Model



After 18 Epochs

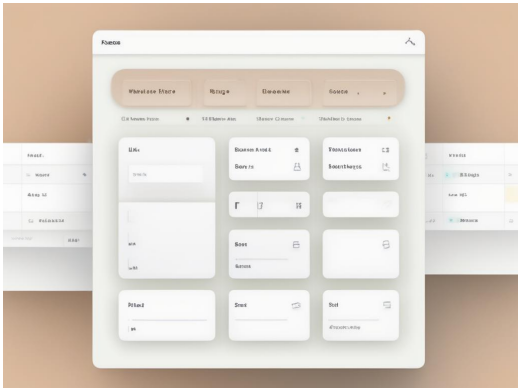


After 36 Epochs

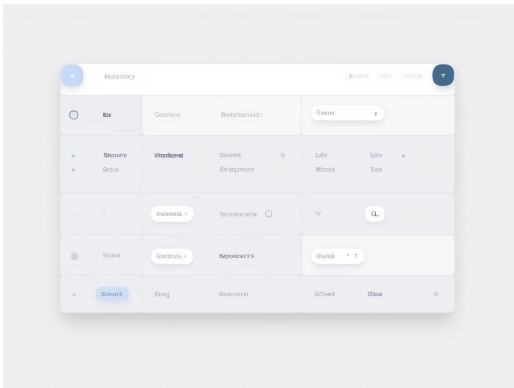


Example Prompt: “A user interface with 4 rows and 4 columns with a light theme”

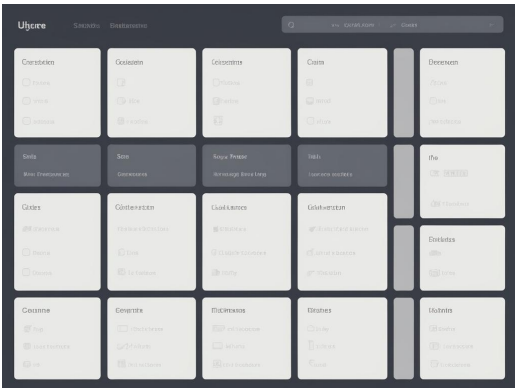
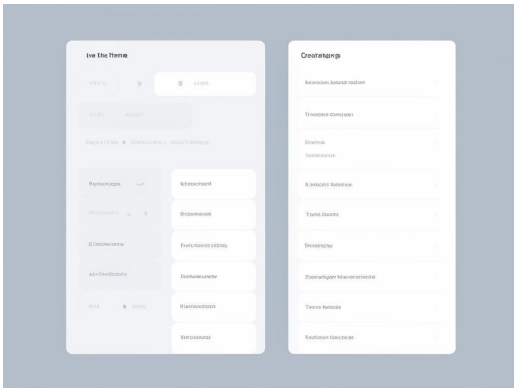
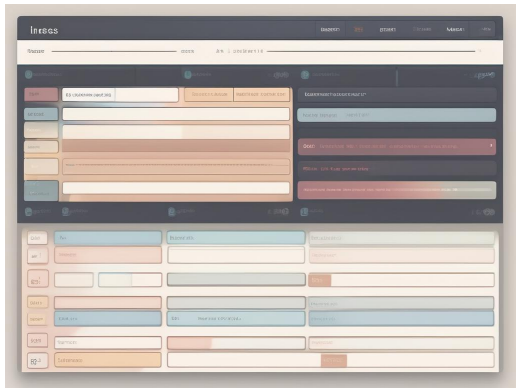
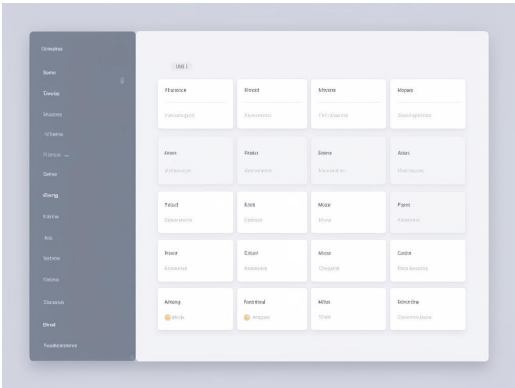
Base Model



After 18 Epochs

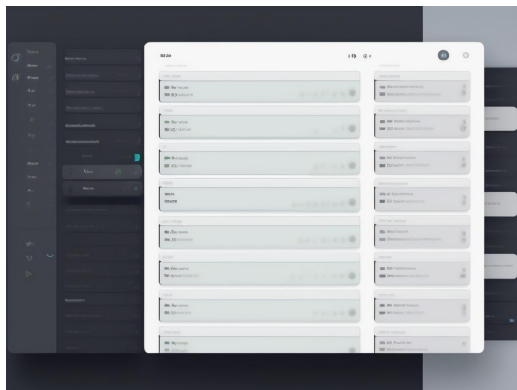
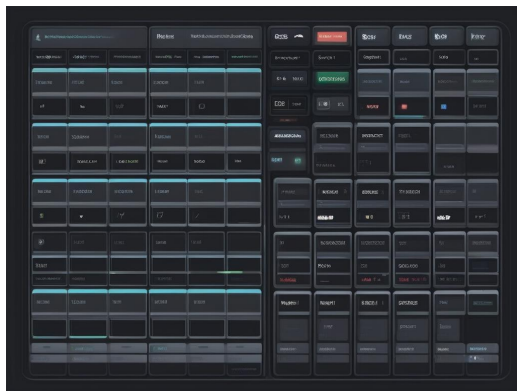


After 36 Epochs

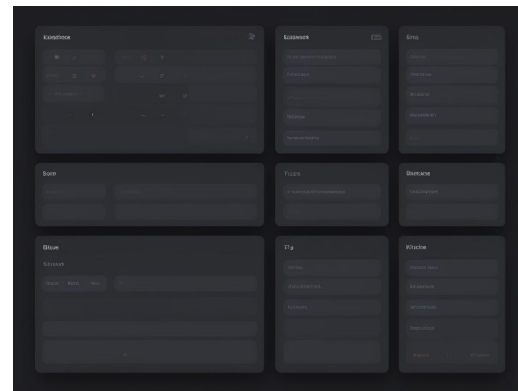


Example Prompt: “A user interface with 5 rows and 5 columns with a dark theme”

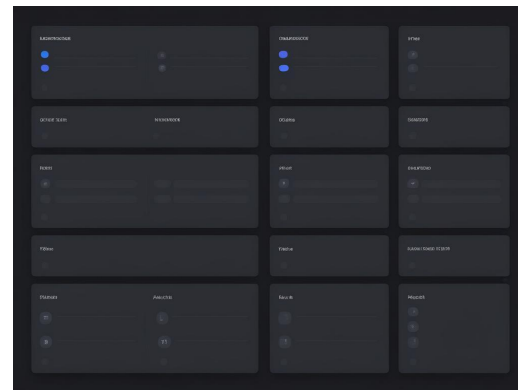
Base Model



After Fine-Tuning



checkpoint



final

Observations

- Model did not learn:
 - Exact grid width and height (the meaning of numbers)
 - Perfect themes
- Model learned:
 - General grid layouts
 - Much better looking, especially at larger grid/table size
 - Approximate grid size (small vs large)
 - The shape and style of the UIs I gave it
 - Rounded corners, color scheme

Where I think I should go next

More training images

- HTML elements (in progress)
- Sidebars, Navbars, Menus

Train smarter

- Adjust and train multiple times to figure out the best learning rate, batch size, length of training, etc.