Investigating the usefulness of a generative AI when designing user interfaces

Andreas Bleichner
Nils Hermansson
Investigating the usefulness of a generative AI when designing user interfaces

Andreas Bleichner
Nils Hermansson

Abstract

Generative AI is a hot topic as of 2023, with huge financial investments in the industry. The areas of use for it are rapidly expanding. One potential benefit of generative AI could be in the field of UX design.

For this master's thesis, a Stable Diffusion model has been fine-tuned to create pictures of login screens based on text prompts written by a user. A set of these pictures have been used in a prototype and the concept has been evaluated through user tests. The prototype and the concept of using generative AI in UX design received a positive reception from testers. It was established that further work on the fine-tuned model and how to use it as a tool is required for it to be effectively integrated into the design process of user interfaces.
Teknisk-naturvetenskapliga fakulteten, Uppsala universitet.

Utgivningsort Uppsala/Västerås.

Handledare: Förnamn Efternamn,

Ämnesgranskare: Förnamn Efternamn,

Examinator: Förnamn Efternamn
Acknowledgments

We want to thank Katie Winkle for her great ideas and guidance throughout the project. We also want to thank Mikael Axelsson for his technical expertise and support. Thanks to Consid for providing us with hardware to train the model, and thanks to all of the testers for contributing with great feedback and insights.
# Contents

1 Introduction 1  
   1.1 Fine-tuning the Model for Specialized Login User Interface Generation 1  

2 Background 5  
   2.1 Text-to-image Generative Models 5  
      2.1.1 Generative Adversarial Networks 6  
      2.1.2 Diffusion Models 7  
   2.2 Stable Diffusion 8  
      2.2.1 The Autoencoder/Decoder 9  
      2.2.2 A Conditioned U-net Neural Network 9  
      2.2.3 Conditioning 10  

3 Purpose 11  
   3.1 Research questions 11  

4 Related Work 12  
   4.1 Galileo-AI 12  
   4.2 Uizard 12  
   4.3 Midjourney 12  

5 Method 14  
   5.1 Creating a New Model 14  
   5.2 Fine-tuning 14  
   5.3 Using Low-Rank Adaptation to fine-tune Stable Diffusion 15  
   5.4 Finding a Suitable Dataset 17
5.5 Training ......................................................... 18
  5.5.1 First training session ................................. 20
  5.5.2 Second training session .............................. 22
  5.5.3 Final training session ................................ 23

6 Evaluation Method ........................................... 25
  6.1 Evaluation Requirements ............................... 25
  6.2 Delimitations ............................................ 26
  6.3 User Tests ................................................. 26

7 Evaluation Results .......................................... 31
  7.1 General impression of the prototype .................. 32
  7.2 Likelihood of using the images in the prototype as inspiration when designing login screens .................. 33
  7.3 User impact on generated images ....................... 34
  7.4 Positive impressions .................................... 34
  7.5 Negative impressions ................................... 35
  7.6 Ideas for improvement of the prototype .............. 35
  7.7 The use of a further developed prototype in a work environment ............................................. 36
  7.8 The effects of a further developed tool similar to the prototype on the UX design profession ............ 36

8 Discussion ...................................................... 37
  8.1 Fulfilment of the Evaluation Requirements ........... 37
  8.2 Qualitative Feedback on the Prototype ................ 37
  8.3 Potential uses & effect on the UX profession .......... 38

9 Conclusion ...................................................... 39
1 Introduction

As of early 2023, generative AI:s is a hot topic in popular culture and the tech industry. Multiple websites offer the user to write a sentence and return an image portraying what the user has written. Some websites allow this for free while others ask the user to pay for the service. On January 23, 2023, Microsoft announced a multi-billion dollar investment in OpenAI, the AI research laboratory that invented ChatGPT and DALL-E[12][13] which implies the importance and potential of the AI field. AI has already seen use in many industries such as health care, manufacturing, retail, and banking. The relatively new field of text-to-image generative AI:s (see section 2.1) further expands the use field of AI:s. A potential area of application that we investigate in this paper is that of user experience design (from now on referred to as UX). For example, a UX designer or a student who is working on a web application project could potentially find usefulness in creating a mock-up user interface that can be used as a first draft and further improved.

1.1 Fine-tuning the Model for Specialized Login User Interface Generation

For this paper, an already existing generalized text-to-image generative AI model was adapted and specialized in generating images of login user interfaces. This process is called fine-tuning the model. Significant improvements in the generated images can be seen when comparing the basic model compared to the fine-tuned model when given similar text prompts related to user interfaces depicting login screens for web applications.
1 Introduction

(a) Image generated before the model was fine-tuned, the result of the prompt: "Login screen with a nature background"

(b) Image generated after the model was fine-tuned, the result of the prompt: "Lgn-scrn with a nature background"

Figure 1 Comparison of the results from a similar text prompt.

Note that in figure 1a the model does not seem to recognize the general attributes that are usually connected to a login screen, such as a text box for the user name/email and a text box for the password. Figure 1b displays a login screen more similar to login screens that you commonly see across the internet.

The fine-tuned model was trained on a dataset consisting of 60 captioned images of various login page user interfaces from websites. A GitHub repository containing the model weights as well as a script that can be used to generate images has been created and is accessible here: https://github.com/nilshermansson/login-screen-lora. There is a readme file in the repository with instructions on how to get started.

In order to evaluate the fine-tuned model an interactive prototype was created. The prototype was created using the web application Figma. The prototype lets the user make a series of choices regarding the properties of the login user interface that is to be created and displays an image based on the chosen properties such as theme, color, and style. Evaluation requirements were set to clarify the expectations of the fine-tuned model, its results, and what data was to be collected from the user tests. The evaluation requirements were largely fulfilled. Testers appeared to have recognized that the images depicted login screen user interfaces and felt like they had an impact on the resulting images based on their choices in the prototype. Testers generally also enjoyed using the prototype, finding it fun and easy to use. Multiple testers experienced that the prototype did not offer enough options regarding the customization of the pre-
sented images that the model had produced. The reason for the limited customization options was due to the model’s ability to handle additional options beyond what it was trained on. The options had to be restricted because of time constraints and limitations in the training dataset. Further expanding the customization options of the prototype would have required more time, training, and data if the quality of the generated images were to satisfy the desired level of standard. Testers were positive toward the further development of a tool inspired by the prototype and thought that it could prove valuable for people in the UX area of work, as well as for hobby developers.

Some concern was raised regarding such a tool competing with designers and doubts if such a tool could really develop to the level that it can be used efficiently in the design process.

Figure 2 shows an image of the work process. It is recommended that the reader go back to this image during the different sections of the report.
1 Introduction

Figure 2 The work with the model and how it was used from start to finish

Declaration of Division of Labor In this project, both of us contributed to doing research and writing the project report as well as all other areas of the project. Nils contributed more to the fine-tuning of the model and Andreas worked more on the evaluation.
This project focuses on the usefulness of generative AI when designing user interfaces. Terms like text-to-image models and diffusion models are relevant to our work. In this section, we describe some central terms.

2.1 Text-to-image Generative Models

The term text-to-image has become increasingly popular since the emergence of the deep learning model DALL-E, developed by OpenAI[1]. The concept of a text-to-image model is that the user enters text into a prompt. An AI trained to produce images based on words and phrases then returns one or more images that should depict what the user has written in the prompt (see figure 3).

One of the early text-to-image models is alignDRAW[28] from 2015, created by researchers from the University of Toronto. The model managed to generate images that were, although blurry, generalized beyond the training set which was demonstrated by using sentences describing scenarios that are unlikely to occur in real life and were not depicted in the training set such as "a herd of elephants flying in the blue skies"[28].

As of the year 2023, there are multiple high-quality text-to-image models available online.
2 Background

Figure 3 An example of how Stable diffusion, a text-to-image model released in 2022, depicts ”a lonely caterpillar wandering the streets of Paris at night in the moonlight.”

2.1.1 Generative Adversarial Networks

Generative adversarial networks (GANs) are a type of neural network used for unsupervised machine learning which can be utilized for generating new data that is similar to some provided input data.
GANs consist of a generator and a discriminator (figure 4). If for example, a GAN was to generate pictures of flowers, the job of the generator is to generate so-called "fake samples", pictures that are not part of an existing dataset depicting flowers. The discriminator is then trained on pictures from the dataset as well as pictures generated by the generator and tries to distinguish them from one another.

If the discriminator makes a correct assessment and judges the picture not to be a part of the real dataset, the discriminator remains unchanged whereas the generator adapts. If the discriminator assesses the generators picture to be from the original dataset, it will learn that the given picture was not from the real set of samples but rather from the generator, and will have to adapt so that it does not get tricked by the generator the next time[24]. The goal is to use the generator after training is complete in order to generate images.

**Figure 4** Image showing a Generative Adversarial Network, taken from https://developers.google.com/machine-learning/gan/gan_structure

### 2.1.2 Diffusion Models

A diffusion model gradually adds noise to an image until it becomes pure noise and then trains a neural network to reverse this process. After training, the diffusion model is to generate an image by giving it randomly sampled noise that can, as a result of the training, be removed so that an image is obtained.

Diffusion models have been established to be preferred for image synthesis over GANs [30]). One reason for this is the risk of mode collapse among GANs. Mode collapse is when a generator produces an output so plausible that it starts to only generate extremely
similar outputs. If the output is “too good” the discriminator will keep getting tricked by the generator that only does subtle changes to the output. The resulting model will use only a small part of the total data set which will give bad results when used with a text prompt. Another issue is that even if mode collapse does not occur there is a risk that the generator reaches good results in tricking the discriminator with a relatively small set of the total data sets, which will lead to bad results for the model even if a mode collapse never occurred.

Because of such issues diffusion models tend to use a wider spread of the training set to generate output than GANs[26].

**Figure 5** A simple diagram showing how inference is performed using Stable Diffusion. In this case the number of inference steps is $N$ and the current iteration of the denoising process is $t$.

### 2.2 Stable Diffusion

Stable diffusion is a latent diffusion model. What this means is that the model works with latent representations of images, meaning the images are compressed. Images are compressed using an autoencoding model that can convert to and from a space perceptually equivalent to the pixel space but with reduced computational complexity[31]. The diffusion process is made on images transformed into a latent space meaning each image is compressed while still retaining its core characteristics. This is useful since it makes the model quicker to train and produce samples. To process data and train the model in the pixel space would require much more time and hardware than when you
utilize a latent space[15]. In total stable diffusion consists of three main parts:

### 2.2.1 The Autoencoder/Decoder

An autoencoder and decoder are used to convert images between latent and pixel space. The term *latent space* is used to describe the part of the model where images and other inputs have been processed from pixels or words to lower-dimensional representational data. *Pixel space* is used to describe the part of the model where images are represented by pixels.

The autoencoder performs a perceptual compression that removes high-frequency details but learns semantic variation. It encodes the data from pixels into representational data that is faster for the model to process.

As previously mentioned an advantage of using an autoencoder as opposed to training the model on actual images consisting of pixels is that the model becomes quicker to train. Instead of forcing the diffusion model to evaluate gradients on all pixels, it can be done on more compressed data. Furthermore, this technique allows for re-usability. An autoencoder can be trained once and then shared for reuse in multiple diffusion models. This enables the possibility for developers to experiment with many different diffusion models while not having to worry about processing the data from pixel space to latent space, given that the autoencoder at hand is satisfactory in its task to process the data[23][31].

Encoded data can be restored to pixel space with a single pass through the decoder for the given iteration of the data $z_t$ where $z$ is the processed data and $t$ is the time step.

### 2.2.2 A Conditioned U-net Neural Network

A U-Net is a convolutional neural network (CNN) originally designed for biomedical image segmentation[32]. When trained, the U-Net will be able to iteratively denoise an image in the latent space until it is ready to be converted to the pixel space. This means the U-Net needs to be time-conditioned by a scheduler to know how far into the denoising process it currently is, or in other words, how much of the image it should denoise in the current iteration.

There are multiple schedulers to choose from and your choice will affect the output. A cosine noise scheduler gradually decreases the amount of noise added during training, while a linear noise scheduler adds a fixed amount of noise at each training iteration. It was discovered that a cosine noise scheduler yielded better results than a linear one [29].
2.2.3 Conditioning

Conditioning in the context of a stable diffusion model is the process where the model learns for example what text input correlates to what type of images. By augmenting the U-net with cross-attention mechanisms the diffusion model is turned into a conditional image generator. It can, for example, be conditioned with semantic maps, images, text, or other inputs [31]. In the context of the U-net cross-attention helps the model preserve particular characteristics and attributes when generating the output.

Call the input data, for example, a text $y$. To pre-process $y$ a domain-specific encoder $\tau_\theta$ is introduced which can be seen to the right in figure 5. In the case of text input, the text is converted to an intermediate representation, in this case, a set of vectors resulting in the input $\tau_\theta(y)$.

The encoded data $\tau_\theta(y)$ then goes through a switch that controls how the data is to be integrated with the U-net depending on the data type of the input. $\tau_\theta(y)$ is then mapped to the intermediate layers of the U-net through the earlier mentioned cross-attention layer which is then used in the training of the model.
3 Purpose

The purpose of this project is to investigate the potential of using generative AI to create user interfaces based on text input, so-called text-to-image synthesis. The aim is to explore the usefulness of such a tool. UX designers, computer science students, people who create web applications in their spare time and many more could potentially find such a tool useful in their respective fields. Additionally, the aim of the project is to understand people’s attitudes toward AI-generated user interfaces and assess their perceived utility and potential future use.

3.1 Research questions

To summarize what we wanted to investigate at the start of the project we posed the following research questions:

- **RQ1: Can generative AI be used as a tool in the process of UX design?** With AI being used more and more in all parts of society, we want to investigate if there is an area of use for it in the process of UX design.

- **RQ2: What do people think about the widespread use of a generative AI will affect the UX profession?** Will automation of some parts of the design process benefit people working with UX design or will it reduce the demand for people with that skill set?
4 Related Work

This section describes some products that are to a certain extent similar to our project and in what ways they differ from our project.

4.1 Galileo AI

Galileo AI[10] is a project that was announced with a tweet by Galileo AI:s Twitter account on February 9 in 2023[5]. Galileo AI appears to be a tool used to design user interfaces based on user input through a text prompt. On Galileo AI:s website it is said that the design is to be editable in Figma[4] which is a popular designer tool that is used to create interactive mock-ups of user interfaces. It is also stated that Galileo AI can be used to create images and illustrations that can be used in the user interface design. No further information about the product is currently available, but the idea appears to be very similar to the idea behind the project that this paper is about. The difference appears to be that this project is focused on web applications, while Galileo AI appears to be focused on mobile app design based on the images on the website.

4.2 Uizard

Uizard is a service that utilizes AI as a tool to assist in designing user interfaces[17]. Among their services, they have the function to scan wireframes that are then generated into a UI mockup with customizable components. Another function is that you can upload a screenshot, which is then generated into a customizable UI mockup as well as the function to generate heat maps that predict where the user of a given user interface is most likely to focus its attention. Uizard has shared some theoretical work behind the models that they utilize for their AI tools, but they have not shared information about the specifics of their models. Uizard differs from our project since it does not have the function to enter text to generate user interfaces, but needs to be provided a picture of an already existing wireframe picture.

4.3 Midjourney

Midjourney[14] is an independent research lab that has developed a text-to-image AI with the same name. Midjourney can be accessed through Discord[2] and through a web app. After entering a prompt the user is presented with four images that were generated
using the prompt. The user can then choose to produce more variations of one of the images or, if satisfied, upscale the resolution of a generated image. Midjourney is a general-purpose model but produces high-quality user interfaces when given appropriate text prompts. Examples can be seen in figure 6.

Figure 6 Results of feeding Midjourney v4 the following prompt: "Flower-themed login screen with blue color for a food delivery app that has username and password fields and a change language button".
5 Method

This project was conducted in multiple steps. As the project proceeded, some major changes were made in terms of the approach to the project. This section describes what methods were used to reach the result and how the project developed over time.

5.1 Creating a New Model

An attempt to create a new model that could be trained was first made since that would have the benefit of full control of how the model works, with the advantage of being able to niche it for the purpose of this paper. This proved to be time-consuming and hard for a few reasons, primarily:

- Training a model from scratch is generally extremely time-consuming. According to Emad Mostaque, the CEO of Stability AI that created Stable diffusion, the model was trained for 150,000 GPU-hours[18].

- Training a model from scratch requires hefty hardware. On top of the high prices of GPUs, the price of electricity to keep the GPUs running during training often adds up to large amounts as well. E.Mostaque has stated that the total cost to train Stable diffusion was around $600,000[18].

- Large datasets are required to acquire a well-trained generalized model that can generate a wide set of different images from a wide variety of words and sentences. It is said that a training set consisting of somewhere between 2.3-5.8 billion images were used for Stable Diffusion (the exact numbers vary in different sources)[3][16]. To acquire such a dataset is not necessarily hard since there are public datasets of this size, but for a small group of people implementing a model, the problem might be to store and work with such large datasets.

Because of those reasons, this idea was abandoned for the benefit of fine-tuning.

5.2 Fine-tuning

Instead of creating a new model, the concept of fine-tuning was used. Fine-tuning relieves the user from training a model from scratch by using an already pre-trained general-purpose model which has been trained on a wide set of classes or images. By not having to train a model from scratch the issues listed in section 5.1 can be
avoided. When using fine-tuning for training deep learning models a well-suited pre-trained model is required as the base model. Alternatively, you can also train an existing base model yourself, but earlier mentioned problems will then have to be handled. The result of fine-tuning a model will be unsatisfactory if the pre-trained base model is unsatisfactory since the results from the fine-tuned model are based on the pre-trained model.

5.3 Using Low-Rank Adaptation to fine-tune Stable Diffusion

Fine-tuning deep learning models with a large number of parameters, like Stable Diffusion, is time-consuming and hardware-intensive. Attempts were made to fine-tune an existing pre-trained model to become specialized in outputting user interfaces. This approach involved updating the weights of Stable Diffusion using a much lower learning rate compared to when you are training the model from scratch. These attempts were unsuccessful due to hardware limitations, specifically, an insufficient amount of VRAM on the available hardware.

Figure 7 A diagram of how LoRA can be applied[27]. Only the LoRA weight matrices are trained while the pre-trained model weights are frozen.

Instead, LoRA, which is a technique to fine-tune models was used. LoRA stands for Low-Rank Adaptation which differs from earlier fine-tuning techniques by utilizing
rank-decomposition matrices (also known as low-rank matrices) in order to reduce the number of parameters that are trained when fine-tuning a model[27]. Rank decomposition matrices are a family of matrix factorization techniques that can be used to approximate a matrix using two or more lower-rank matrices. The pre-trained model weights are frozen and pairs of rank-decomposition weight matrices are added and trained instead (see figure 7). It was initially intended for large-language models like ChatGPT but it has recently been applied to Stable Diffusion. According to the paper which coined the term LoRA[27], there are many benefits of using it in contrast to traditional fine-tuning:

- When fine-tuning a model weights are adjusted at a low learning rate meaning the fine-tuned model is of the same size as the pre-trained model. When using LoRA however, the pre-trained model is frozen. All you need to save are the additional LoRA weights which are significantly smaller (3.9 MB in contrast to 7.7 GB in the case of Stable Diffusion 1.5). This allows users to easily switch tasks and share their LoRA weights.

- LoRA makes training more efficient since it only optimizes the injected low-rank matrices which are much smaller.

- No added latency during inference, which can happen when using other techniques which extend the model’s depth.

- Models fine-tuned using LoRA are less prone to forgetting previously known concepts (known as catastrophic forgetting), since the base model weights are frozen.

The first person to use LoRA when fine-tuning Stable Diffusion was Simon Ryu[25]. They accomplished this by adding pairs of low-rank matrices that are trained instead of training the whole model (see figure 7) to the attention layers of the model. On their GitHub[33] they have shared their code which has since been integrated into the Diffusers[8] python library. Diffusers was created and is maintained by Hugging Face[6]. It enables users to perform inference as well as the ability to train their own diffusion models and it is the library that was used for fine-tuning in this project.

In order to make sure the training would work on the available hardware as well as to ensure the quality of the resulting fine-tuned model we ran a training session on a dataset containing pictures of Pokemon that was available on Hugging Face. The results, while a bit uncanny, were promising and showed that the training was effective (see figures 8a and 8b).
5 Method

(a) Before fine-tuning.  (b) After fine-tuning.

Figure 8 The output of our model before and after fine-tuning using a dataset containing pictures of pokemon. The prompt used was "A blue pokemon with red eyes".

5.4 Finding a Suitable Dataset

Initially, attempts were made to find an already existing dataset that could be suitable for LoRA training but they were unsuccessful. Some datasets containing user interfaces were investigated, this was however abandoned once the decision was made to limit the scope of the output of our fine-tuned model to login screens from applications running on computers. No datasets containing only login screens were found and the process of creating a dataset was initiated.

Initially, 60 pictures of login screens in different styles, colors, and layouts were gathered. The next step was to caption each image in the newly created dataset. A caption is a text which describes the contents of the image and is used during training. The caption allows visual features and visual patterns to be associated with textual information which is crucial when training a text-to-image model. This was more difficult than anticipated since information on how to optimally caption images for LoRA training was scarce. First of all, when fine-tuning using LoRA on a new subject, in this case, login screens, it is important to give that subject a trigger word. This trigger word needs to be included in all captions of your dataset, and using it in a prompt after the training has finished will make the fine-tuned model try to draw that subject. It is important when choosing a trigger word to not use a word that already exists. Because of this, we decided that our trigger word for creating a login screen would be “lgnscrn”.

17
Other than the trigger word, more information in the caption to capture the characteristics of the image was needed. When deciding what other information to include we initially decided on including the following things:

- The color of the background.
- If there are any additional sign-in options like Google, Facebook, etc.
- If there is a forgot password link or button.
- If there are any pictures and where they are located.

![Sign In Screen](image)

**Figure 9** This image initially had this caption: "Sign in screen with blue background and a forgot password link and google login button and Facebook login button and Twitter login button"

Having added captions to all images in the dataset the next task was to upload the dataset to the Hugging Face Hub[9], which is a space where users can create and share models and datasets, in preparation for training. This was done using the Hugging Face Datasets Python library[7].

### 5.5 Training

When looking for information on what version of Stable Diffusion would be suitable as a base model most users at the time seemed to prefer Stable Diffusion 1.5. It also
happened to be the model which was used by default in the guide to LoRA training on Hugging Face. It was decided to try using Stable Diffusion 1.5 and, if necessary, reassess this decision later.

As mentioned earlier the Diffusers[8] python library was used during training. It contains a training script for fine-tuning using LoRA which requires a pre-trained model and a dataset which both need to be hosted on the Hugging Face Hub. During training, it is possible to upload training data to Weights & Biases (from now on referred to as Wandb)[19]. Wandb is a platform that allows users to manage and visualize machine learning experiments. In this case, Wandb was used to monitor the progress of the output of the model during training.

The script, aptly named train_text_to_image_lora.py, takes in a few arguments which can alter how the training is performed:

- **max_train_steps** specifies how many training steps should be performed.
- **train_batch_size** specifies how many inputs (images in this case) are processed in each step. Having a higher value speeds up training but due to hardware limitations, we had to set this value to 1.
- **learning_rate** specifies at what rate the weights of the model should be adjusted after each training step. When fine-tuning the traditional way this value defaults to 1e-06 but when fine-tuning using LoRA, the learning rate needs to be higher. We used the default value which was 1e-04.
- **resolution** specifies resolution of input images. All images will be resized to this resolution during training. It was set to 512 since that was close to the resolution of the images in the dataset.
- **validation_epochs** specifies how often validation should occur by setting the number of epochs that need to pass. An epoch refers to processing the entire dataset once (in our case this means one epoch is 60 training steps). Validation was performed by generating 4 images with the current state of the model and uploading them to Wandb.
- **validation_prompt** specifies the prompt that should be used when generating the validation images.
5.5.1 First training session

The first training session performed above expectations considering no adjustments to training parameters or image captions had been made yet. The training ran for 12000 steps which took approximately 14 hours. It took a lot longer than estimated since validation was performed at every epoch which was deemed unnecessary and adjusted in later training sessions. The output was inconsistent, but some images were deemed to be of good enough quality to be used in the user tests which was promising. It was, however, not possible to add additional login option buttons as well as the forgot password button/text. Adding these to the prompt did not seem to affect the output at all. This might be because their appearance in the dataset differed a lot. It was decided to remove these features from the image captions in the dataset.

![Image](image.png)

**Figure 10** Output after the first training session using the prompt "Lgnscrn with a nature background"

When looking at the validation images on Wandb to see how the model performed at certain training steps it was determined that the model did not appear to show any noticeable increase in output quality after 6000 training steps had been performed.
5 Method

(a) Model validation at 732 training steps

(b) Model validation at 2501 training steps

(c) Model validation at 6100 training steps

(d) Model validation at 10919 training steps

Figure 11 Validation which was performed during training displayed on Wandb.
5.5.2 Second training session

While planning the second training session the user test was being created simultaneously. It was decided to adjust the captions of the images to hopefully make the model perform according to the specifications of the evaluation requirements (see section 6.1). At the time the user test would allow the user to select the style and the colors of the login screen. First, we categorized all images in the dataset as either vivid or minimalist by changing the captions of the images to include one of those properties for every image in the dataset. Minimalistic login screens had less vibrant colors and a simpler layout than vivid ones. Secondly, many login screens in the dataset have a box containing the login fields which often has a different color than the background. In an attempt to control the color and size of that box as well as the color of the background, the captions were adjusted. The smaller boxes were dubbed cards and the large boxes were dubbed boards (see figure 12).

![Image of examples](image.png)

**Figure 12** Examples of images and their captions from the dataset used for the second training session.

This time the training was completed in roughly 4 hours. It was significantly faster due to it only being run for 6000 steps as well as only performing validation every 10 epochs. The quality of the output was not significantly different from the first training session but the attempts to control the color of the previously mentioned boxes proved unsuccessful. Adding styles to each caption proved successful since the output was affected in the intended way.
5.5.3 Final training session

In preparation for this training session adjustments were made to the captions of all images in the dataset. The captions now only contained the color of the background, the style of the login screen as well as if any images existed. The training ran for 6000 training steps and was completed in just shy of 5 hours and the results seemed slightly better than the previous training sessions. When this training session was finished, the user tests were almost fully prepared. In order to see if this model would perform well enough, the output of the model was evaluated by trying to generate images that could be used in the user tests. The output was inconsistent but sometimes good enough to be used in user tests. Due to time constraints and satisfactory output, the model was considered good enough to be the version used in the user tests and further training was discontinued.
(a) vivid lgnscreen with green colors and a (b) minimalistic lgnscreen with green colors in the city in the background and nature background

**Figure 14** Examples of output after the final training session
6 Evaluation Method

The generative model trained in this project and the concept of using generative AI to generate user interfaces were evaluated with a prototype through user testing. Before the user tests, requirements were made in order to clarify what we were trying to achieve with the evaluations. By defining the evaluation requirements beforehand, clear goals were established which helped to assess the effectiveness of the fine-tuned model and attitude toward the use of generative AI for creating user interfaces. This also provides explanations and context of our evaluation methodology to readers, so that they can understand and interpret the results accurately.

6.1 Evaluation Requirements

Requirements were made to clarify our expectations of the model itself, as well as what data we needed to gather from the users through user tests.

The requirements that were set for the model were:

- The tester should recognize that the presented user interface is a login screen
- After the user test the tester should feel that the choices they made had an impact on the result (for example if the tester entered that they want a tropical theme on the user interface they should recognize that in the final image of the interface)

Furthermore, we wanted to investigate the tester’s attitude toward generative AI, specifically in the area of creating user interfaces. We deemed it of value to understand if the tester:

- Enjoyed having an AI generating images of user interfaces based on the tester’s specifications.
- To what extent did the tester feel in control when using the prototype?
- Based on the tester’s experience with the current version of the product, do they believe there is potential for further development? If so, what improvements or features would they suggest to enhance its potential in the future?
6 Evaluation Method

6.2 Delimitations

Because of time constraints, we delimited the scope of the tests to login screens. This allowed us to specialize the model by training it on a dataset consisting of login screens so that we could focus on ensuring the quality of that type of user interface.

Another choice of delimitation was that we did not let the tester access the text prompt. Instead, a prototype was used where the tester could choose between a few limited choices of interface properties, such as choosing between only two styles, "minimalistic" or "vivid".

This is because the quality of the output of the model could not be guaranteed if the tester gets access to the text prompt. The model has been trained on certain words and phrases. If the tester was to not use some particular phrases the output would become unpredictable. Furthermore, the consistency of the test experience among all the testers needed to be guaranteed. Any form of randomness that could affect the tester was undesired, for example, if one tester received a white image while another tester received an advanced login screen from the same text prompt. For these reasons, it was decided to generate the user interfaces in advance, and just present the user interface that matched the testers' choices during the tests. For more on how the tests were designed, see "User tests", section 6.3

One alternative way of testing that was considered was to create an application that would be connected to the model through an API and let the tester fill out certain words in a sentence to generate an image of an interface, for example, we could have the sentence:

"A login screen with a and b colors, c theme and d style"

where the tester could replace a, b, c and d with options from drop-down menus with a limited amount of alternatives, but ultimately we decided that even this presented a layer of unpredictability that was unwanted in the tests.

6.3 User Tests

User tests were made to evaluate the model. The evaluation method was inspired by the book Interaktionsdesign och UX: om att skapa en god användarupplevlse[20] by Linköping University researcher Mattias Arvola who teaches and research in UX, interaction design, and cognitive science[11].

A prototype in Figma and a survey created with Google Forms were used for the user tests.

In Interaktionsdesign och UX: om att skapa en god användarupplevlse Arvola mentions
Evaluation Method

that the testers should represent the target group for the product[21]. Testers were chosen with this in mind, more about the testers in section 7.

In order to help the tester understand their task during the test, three user stories were created. In each user story, the tester was given a background story that described what the tester wanted to achieve in the specific scenario and why. The user stories can be found in section A of the appendix.

To test the concept of using a generative AI for UX design and to get a general opinion about some of the images that the fine-tuned model produced, an interactive prototype was created in Figma. We created a vertical prototype, which Arvola describes as a prototype used to show some limited functionality and how this functionality is supposed to look and work[22].

The tester was first presented with instructions for how to do the user test, as well as the three user stories (found in the appendix, section A). In the instructions, a link to the Figma prototype was provided. The tester was instructed to read the user stories and perform the tasks provided in the user stories by using the Figma prototype. In the prototype, the tester first got to choose a general theme for the login screen (see figure 15).

![Figure 15 The tester is choosing the theme of the user interface that is to be created](image)

The tester then got to choose a color scheme for the user interface(figure 16).
The tester then picked a style for the user interface (figure 17). Style refers to a general feeling of the user interface, in this case, the tester got to choose between "vivid" and "minimalistic"
Lastly, the tester was presented with two images of user interfaces that implemented the
6 Evaluation Method

selected properties (figure 18). The tester was instructed to follow this process once for every user story. The tester started over the process in order to fulfill the next user story by clicking "restart".

When the tester was finished with the prototype, the tester answered the last part of the survey. When the user stories were completed and the tester was done with the prototype, the tester filled out the survey. In the first part of the survey, the tester responded to questions related to their previous experience with generative AI:s. The tester then answered the second part of the survey which provided some information about the tester, such as the tester’s occupation and earlier experiences with generative AI. The tester was not required to provide any personally identifiable information like name, age, or sex since we wanted the tester to be as anonymous as possible to protect the tester’s integrity. The third and last part of the survey concerned the evaluation requirements and provided us with data used to control whether the set requirements were met or not.
7 Evaluation Results

The survey and all of the answers can be found in the appendix, section B.

In total, there were nine people who completed the user tests. Five of the testers were IT engineering students. They represent the part of the target group that could use this type of application in their studies since IT engineering students work with UX and usually have to design user interfaces during their studies. Four of the testers worked with IT in some capacity which represent another target group who could use a tool similar to the prototype in their profession. All testers did have some experience designing user interfaces which was of most importance. Most testers said they had done it in the past occasionally, two testers said they do it regularly and one tester said it is their main occupation (see figure 19). Having testers with different levels of experience when it comes to user interface design provides different perspectives which is a good thing when evaluating. Almost all testers were familiar with generative AIs with only one tester saying they were somewhat familiar with the expression (see figure 20). Most testers had used ChatGPT and about half reported having used a text-to-image generative AI like Dall-e and Stable diffusion. One tester reported not having used any type of generative AI.

![Survey results when the tester was asked how familiar they are with designing user interfaces](image)

**Figure 19** Survey results when the tester was asked how familiar they are with designing user interfaces
7 Evaluation Results

7.1 General impression of the prototype

In the survey answered by the user testers some words and phrases were re-occurring in regard to the general impression of the prototype. The testers expressed that:

- The prototype was fun to use
- The prototype was easy to understand and use
- Some testers felt too restricted by the predefined options

One tester expressed that some of the color combinations in the produced images were displeasing. Another tester expressed that the test was so bare bones that it hurt the prototype.
7.2 Likelihood of using the images in the prototype as inspiration when designing login screens

Do you think that the resulting images in the prototype could be used as inspiration when designing a login screen?

9 responses

![Bar Chart](image)

**Figure 21** When asked if the resulting images in the prototype could be used as inspiration when designing a login screen, 44.4% of the testers answered with 5 on a scale from 1 to 5 where 1 meant "not at all" and 5 "fully agree". 33.3% answered 4 and 22.2% answered 3.
7.3 User impact on generated images

When asked if the testers felt like the choices they made when picking properties for the login screens had an impact on the final images, 77.8% of the testers answered with 5 and 22.2% answered 4 on a scale from 1 to 5 where 1 meant "not at all" and 5 "fully agree".

7.4 Positive impressions

When asked if there was anything in particular that gave them a positive impression, the responses were mixed but summarized that:

- Multiple testers were impressed by the generated login screen designs. The color schemes, the interplay of design elements, the layout, the use of real imagery, and the overall designs received positive feedback.

- Multiple testers responded that the ability to produce many user interfaces quickly gave a positive impression.

Some other responses were: "...produces recognizable login screens" and "easy to use and distinct correlation between choices and results".
7.5 **Negative impressions**

The response to if there was anything in particular that gave the testers a negative impression was mixed:

- One tester thought that the images were hard to see in the prototype
- Two testers responded that the text produced by the AI was annoying since was not consisting of real words, but rather just cluttered the design
- Another two testers responded that they wished for more options and to be able to influence the output more

The rest of the feedback primarily related to some colors looking bad, a bit too generic layouts, and some UI elements feeling out of place.

7.6 **Ideas for improvement of the prototype**

When asked if they had any ideas for improving the prototype the most popular responses were:

- To add more options during creation. For example more styles and the ability to control the layout.
- The ability to iterate on generated images. For example by creating variations of the image or by using a text prompt to adjust an image to your liking.

One tester requested to have a "re-do-based-on-the-same-prompt-button" which could be used to generate more images quicker. Another tester wanted to have direct access to the prompt used to generate images.
7 Evaluation Results

7.7 The use of a further developed prototype in a work environment

To the question if the testers could see themselves incorporating a further developed version of the prototype in their line of work or for other purposes 77.8% of the testers answered ”Yes” and 22.2% answered ”Maybe”

When asked how the testers that answered ”Yes” would use it the majority of answers were centered around speeding up the initial design process, brainstorming, first drafts, and getting inspiration.
One tester responded that besides speeding up the design process it would also be practical to use since it removes a layer of discomfort that comes with criticizing another person’s work, for example if you are to give feedback on another person’s initial user interface design. The tester means that it would feel easier to give feedback on an AI-generated user interface.
Another tester responded that if the model was developed so that it can produce more than login screens they would use it in their line of work. It would have to produce for example individual components, and it would be even better if it could generate the code from the generated user interface as well.

The tester that answered ”Maybe” responded that it would need to be further developed to the level where the model is integrated into a website builder where the whole framework for the user interface could be generated. In this case, the tester could potentially use it as inspiration for a personal blog or website.

7.8 The effects of a further developed tool similar to the prototype on the UX design profession

Lastly, the testers were asked to briefly elaborate on how they thought that widespread use of a tool similar to this prototype would affect the UX-designing profession as a whole. The response was overwhelmingly positive. All testers answered that it could speed up the work for UX designers and similar roles in the design field as well as give non-designers a helpful tool.
Only two testers expressed concern besides their positive feedback. One of the testers expressed that if such a tool start to produce too good results it could become a problem for the whole UX design area of work. The other tester expressed that it might be hard for such a tool to generate examples that adhere to existing design principles.
8 Discussion

The user tests were conducted without any issues reported by the testers. Based on the results of the survey we could evaluate and reflect on the tester’s attitude towards generative AI and the quality of the prototype.

8.1 Fulfilment of the Evaluation Requirements

The evaluation requirements were mostly satisfied. Based on the user tests the testers appeared to have recognized that the produced user interfaces depicted login screens and seemed to feel that they impacted the resulting images with their choices. The testers also provided relevant insights regarding the questions listed in the evaluation requirements.

As multiple testers expressed that they felt some lack of control due to the few options for customization in the prototype, it appears as if the users did not feel autonomous enough when handling the prototype. In the case of the prototype, as earlier mentioned, this was a conscious decision made by us since restrictions were required, both for the quality of the images produced by the fine-tuned model and for the quality of the tests (see section 6.2 for more delimitations). It was considered to inform the testers about this so that they could focus on other aspects of the tests, but ultimately it was decided that all kinds of feedback surrounding the subject would be interesting.

8.2 Qualitative Feedback on the Prototype

One purpose of the prototype was to show the testers how a finished product could be used and inspire them to imagine why and how it could be useful. All testers listed at least one area of potential use which suggests the prototype fulfilled its purpose. One of the questions in the user tests asked the testers how they could see themselves using a similar product currently.

- Most examples provided by testers involved using the output as inspiration when designing a user interface, which aligns with the intended use.

- Overall the testers enjoyed using our prototype and found it easy to use. Many testers did however feel like they did not have enough control over the process of generating the login screens. This becomes apparent when looking at the sug-
gested further improvements to the prototype which almost all included either adding more options or allowing a user to modify an already existing image

- Some feedback was directed toward the fine-tuned generative model rather than the prototype: regarding the testers’ choices of properties and the outcome some testers did not experience that the colors matched the chosen options well

The consensus among the testers seemed to be that in order for a product similar to the prototype to be useful to them there needed to be significant improvements made. The fine-tuned generative model needs to be further improved for it to effectively be used as a tool. Since testers expressed dissatisfaction with the color schemes and the colors in the output compared to the chosen colors by the user, it is apparent that the model needs further training on a wider spread of user interfaces, possibly with a larger variety of image captions in the dataset. We considered including a question in the evaluation questionnaire regarding how well the images adhered to design principles. We decided against this due to the images not having sufficient quality.

### 8.3 Potential uses & effect on the UX profession

We wanted to know how the testers thought that the usage of a further developed version of the prototype could affect the UX designer profession as a whole.

All testers expressed positive feedback saying it could speed up the design process for UX designers and might provide effective assistance to non-designers. This technology-friendly attitude might stem from the fact that all of our testers come from the IT sector, either in their line of work or as students. It is worth noting that we had only one tester that has UX design as their main occupation among the testers. This unfortunately leads to a lack of representation in our test group for the people who will be the most affected by the widespread use of a product similar to our prototype.

One important question that needs to be raised is whether a tool like this would reduce the demand for UX designers or even replace them entirely. Most seemed to think a UX designer would be necessary at some point in the process but it would be interesting to know if UX designers would share the concern of one of our testers who expressed that a further developed tool could potentially replace UX designers. The general consensus seemed to be that a tool similar to the prototype would benefit the UX profession since prototyping and getting inspiration when designing a user interface would be easier.
9 Conclusion

The goal of this project was to evaluate the use of generative AI when creating user interfaces. This was done by creating an AI capable of generating pictures of user interfaces and using it to conduct user testing. The resulting model was able to create images of login screens which were deemed to be of sufficient quality. The model was, however, very inconsistent and roughly 80% of the output had to be discarded while generating images for the user tests. A prototype was created in Figma using the images produced by the model and used to conduct user tests. The feedback received from the user tests displayed a positive attitude toward the prototype and the use of generative AI as a tool when designing user interfaces.

The user test results indicated potential use for a generative AI in the UX design field. More user tests with a higher fidelity prototype are required in order to thoroughly investigate this area of potential use.

Based on these results we conclude that an answer to the research question “Can generative AI be used as a tool in the process of UX design?” is that generative AI has the potential to be used as a tool in the process of UX design given further development according to the impression our testers got from using the prototype. The testers found other potential areas of use than simply as inspiration during the design process. For example, it could be used to quickly make a prototype when planning the design of an application with a stakeholder.

As an answer to research question 2, “What do people think about the widespread use of a generative AI will affect the UX profession?” It is hard to say based on our relatively small test group, but in that group, the testers mostly thought of it as a tool that could aid designers rather than replace them.
10 Future Work

The most obvious improvement that could be done to the current fine-tuned model would be to improve the consistency of the generated output. Currently, multiple outputs of the same text prompt are needed to reach a satisfactory result. Many of the produced images do not meet our standards, and can in some cases not even be considered user interfaces. For examples of substandard outputs, see figure 23a and figure 23b.

(a) Image generated by the latest version of the fine-tuned model, the result of the prompt: "Vivid lgnscrn with gray colors and animals in the background"

(b) Image generated by the latest version of the fine-tuned model, the result of the prompt: "Minimalistic lgnscrn with blue colors and a futuristic city in the background"

In figure 23a no necessary properties for a login screen can be seen, it just appears to be an image depicting animals not known to man. Figure 23b displays some features that could be interpreted as a company logo and some type of text boxes. The request for a futuristic city has been completely neglected and the image could be considered messy.

It would be desirable that every generated output at least contain some basic elements necessary for login screens such as fields to fill with user data required to log in and that the output always contain the requested features such as color and theme.

We believe that the performance of the model could be improved with a larger dataset and more training steps which could be attempted in the future. The dataset that we have
Future Work

Fine-tuned the model with currently consists of 60 images of login user interfaces, labeled in a specific way to maintain consistency in the dataset. If we were to expand this dataset the output might become more consistent. Furthermore, increasing the training steps, meaning that the model loops through the dataset might also improve the output. It is worth considering that there might be a more optimal way to caption the images than how it was done in this project. As earlier mentioned information regarding how to do this was scarce, and it is possible that some ways to construct sentences are more effective than others in text-to-image conditioning. Further research on this subject would be positive in future work.

The fine-tuned model could be adapted to accommodate people with visual impairments such as color blindness and poor vision.

It could also be adapted to generate images of user interfaces for mobile applications.

Another idea for future work is to change the output from images to Figma files. Instead of receiving an image of a user interface, the user could get a .fig file that can be opened in Figma, containing the design of the user interface with movable components and all other features that Figma offers. This would shorten the path between getting a mock-up of a design and to start working with the mock-up. A further development could be that the user could get a user interface that is built with a framework like React or Flutter. By describing the desired functionality and design of the interface, the AI could generate a complete user interface with functional components, styles, and behavior. This could reduce the time and effort required to design and develop a user interface and could open up new possibilities for quick and efficient prototyping. With such a tool designers and developers could focus more on the high-level conceptualization of the user interface, while leaving the implementation details to the AI. Whether this could contribute to a bright and optimistic future where the developers use AI as a tool to become more efficient in their work or a cynical dystopia where people get replaced with AI is currently a hot topic, but it is nonetheless one route that further development could go.
References


[18] “We actually used 256 a100s for this per the model card, 150k hours in total so at market price $600k.” [Online]. Available: https://twitter.com/EMostaque/status/1563870674118320667?sf=20


A  User Stories

**User story 1:** As a product owner at a small company, I want to use an application to generate an image of a login user interface, so that I can use it as a first draft to draw inspiration from for the hiking community website that I am designing.

**User story 2:** As a UX designer, I want to use an application to generate an image of a login user interface, so that I can use it as a first draft to draw inspiration from for the Bahamas travel guide website that I am designing.

**User story 3:** As an IT engineering student, I want to use an application to generate an image of a login user interface, so that I can have a personalized and visually appealing mock-up login screen for the accommodation rental web application that I need to create for a course.

B  User Survey Results

These are the results from the survey used in the evaluation.

**Figure 24** Do you as a tester understand that you will be anonymous if your answers in this test are used for the project report?
B User Survey Results

Do you as a tester consent to that the results of your test might be used in a project report that is publicly available?

9 responses

Figure 25 Do you as a tester consent to that the results of your test might be used in a project report that is publicly available?

The information you have provided will not be used outside of the project that the test is about. Do you agree to that we save the information you have provided us for the purpose of this project?

9 responses

Figure 26 The information you have provided will not be used outside of the project that the test is about. Do you agree to that we save the information you have provided us for the purpose of this project?
Figure 27 Testers occupation (Work/school/etc). There were two testers that had the occupation “Student”.

Figure 28 How familiar are you with designing user interfaces?
B  User Survey Results

Figure 29 Are you familiar with the expression "generative AI"?

Figure 30 If yes: have you ever used a generative AI and in what context?
Figure 31 Continuation on previous question

Questions after the test has been done:

What was your general impression of the prototype that you tested? (Fun, boring, hard, unclear what to do?)

9 responses

- At first glance it was a bit difficult to understand how many options you would need to choose, but after the first try it was easier to generate a requestet login screen.

- Generally I had no issues with testing of the prototype, the instructions were quite clear and easy to get a hang of.

- Well the test was pretty barebones and all the user stories were really obvious to complete. And it kinda hurt the showing of the prototype since the login screens where all pre-generated beforehand and pretty basic. That is the limitations of figma of course. Maybe I know too much about the process so im the wrong person to ask. Im guessing the next prototype is going to be the real deal.

- It was fun and impressive

- Fun! It was enjoyable to click around and see the different results you could get.

- Very clear UI that was super easy to understand. The results fit quite nicely with my choices as well, however especially for the vivid ones the color combinations are not always the best.

Figure 32 What was your general impression of the prototype that you tested? (Fun, boring, hard, unclear what to do?)
B User Survey Results

Figure 33 Continuation on previous question

Feels like this can work really well as a starting block.
Quite easy to understand with buttons for every keyword. Buttons with the specific keywords are quite restrictive in the creative process, but I guess a future work would be to allow for free text.
Cannot say I found it either fun or boring, but with a more open result it could prove to be fun and allow for creativity.
A bit restrictive in available choices.
Fun and straightforward!

Do you think that the resulting images in the prototype could be used as inspiration when designing a login screen?
9 responses

Figure 34 Do you think that the resulting images in the prototype could be used as inspiration when designing a login screen?
**B User Survey Results**

**Figure 35** Do you feel like the choices you made when picking properties for the login screens had an impact on the final images?

**Figure 36** Was there anything particular that gave you a positive impression, and if so, what?
B  User Survey Results

I very much like the idea that you can quickly get some example designs and general feeling based on some inputs. I think this could be a great tool for designers and concept workers in the future.

Easy to use and distinct correlation between choices and results.

Distinguish quite good between background and foreground components, such as the nature, blue and minimalistic design.

Also produces recognizable login screens.

The layout of the generated user interfaces was surprisingly good in some cases, e.g. positioning of UI elements, size, etc.

The designs were quite fun. Liked that it incorporated real images as backgrounds.

Figure 37 Continuation on previous question

Was there anything particular that gave you a negative impression, and if so, what?
9 responses

not necessarily

It strikes me that scaling up the tool might be difficult as the amount of colors, themes and styles grow over time. The main benefit I see to this is that I quickly can get a easy, but "dirty" image in my hands and adding enough steps and customizations to the generative questions to the users might prove detrimental to that point. Important to keep ease of use in balance with wanted complexity in generation. E.g. having the option whether I would like to avoid spiders or snakes might be great for accessibility reasons in my drafts, but not a question that is relevant often.

They were kinda hard to see.

The text in the interfaces was a bit "annoying" since it looked a bit cluttered

I would like to be able to have the option to influence the actual login parts of the screen too (i.e number of fields, location of the login form etc.)

Limited options, somewhat generic layouts and colors felt a bit off from time to time.

Figure 38 Was there anything particular that gave you a negative impression, and if so, what?
The text seems to be just some mumbo jumbo where the designer has to guess what it is supposed to represent for further design.

Sometimes it generated UI elements that made no sense in context, e.g. when a login screen was generated there was a slider element in the login container. Color choices were sometimes odd, like there was no consistent theme to it.

The "Vivid tropical grey" gave me a very colourful image which was a bit strange.

**Figure 39** Continuation on previous question

Do you have any ideas for improvement of the prototype?

9 responses

More options to choose from

A re-do-based-on-the-same-prompt-button could be useful to generate many images quicker.

If at all possible, the ability to fine-tune a result with added details like "make the background less saturated" or other changes to a generated image would be immensely helpful.

Wish I could have zoomed in on one of the two chosen designs.

Maybe skip text

Having the option to influence layout of the page too and the general shape of it would be really nice. Being able to pick an initial generated design that I liked and then iterate on that would also be of high value.

More options to put your personal touch on the ideas. Modern / retro for example.

Allow for free text inputs and a result with clearer text, does not have to be perfect but recognizable.

**Figure 40** Do you have any ideas for improvement of the prototype?
I think having a selection of "base" choices is good. The interface that the designer interacts with (when selecting generation options) could have UI components that let you select or describe the context of your user interface. For example, if I as a designer need to design a login screen, then I can select "Login screen" from some predefined list of selection (or alternatively describe the purpose of the UI. This could extend to other parameters that the AI uses. Color, style, theme could all have a significant amount of more options.

Also, allowing iteration on a previously generated image would probably be extremely useful. A designer could generate an initial UI mockup and then communicate to the AI "change the color of this specific modal to a light blue" or "add padding to this container" and then the AI generates a new version where those changes are applied.

More options!

Figure 41 Continuation on previous question

If the prototype was further developed, could you see yourself incorporating it in your line of work or for other purposes?

9 responses

Figure 42 If the prototype was further developed, could you see yourself incorporating it in your line of work or for other purposes?
If yes: Please tell us how you would use it.

8 responses

I could use it to get inspiration on different login screens and compare different results with each other.

As a time-saver in prototyping stages of projects. Sometimes I don't have the time to throw together a working design and rather than making something half-assed, it is sidpped entirely until later in the project. Using, and explaining what was used, AI tools for this tasks could ensure that I don't spend hours on something that gets thrown into the bin, while also not having the anxiety connected to harshly criticizing a person's work.

Well it would have to be more than just login screens, which there are plenty fine standard designs already. It would have to be other screens and components like specific menus and header and navbars. It would be great if you could generate the code from whatever you build directly.

I would use it like in this example, give the program some keywords relating to my expected outcome.

It would speed up the concept artists or designers work, it would make it possible to quite quickly present multiple ideas in order to pick one to refine. This would make it less costly and time-intensive to produce a design that is likely to be used.

**Figure 43** If yes: Please tell us how you would use it.

If it produces a useable UI quite fast, I would use it as a first draft. Maybe fix some issues I feel it might have, then I would have a first user experience test, which hopefully speeds up the process.

Quickly generating ideas for user interfaces

For brainstorming and inspiration

**Figure 44** Continuation on previous question

If no or maybe: Why?

1 response

I might use it as an inspiration for a personal blog / website. However for me to use this instead of a full websitebuilder with premade themes I would like to see it incorporated into a similar product. To clarify I would gladly be able to prompt a website-builder how I want my site to look and feel and then have the whole framwork being generated in that style.

**Figure 45** If no or maybe: Why?
Please elaborate briefly on how you think the widespread use of a tool similar to this prototype would affect the UX-designing profession as a whole?

I think it will be in favor of UX-designers because this tool helps to provide them with a quick prototype to get inspiration from.

Knowing little about how well-schooled and well-researched UX-designers work, it is quite possible that the generated images will have too hard a time generating examples that adhere to design principles. That said, I imagine the ease of absolutely blasting out examples in early stages in order to get a feel for a customer's taste could be incredibly useful. I imagine a designer in a very early stage being able to put generated images in the hands of a stakeholder and using it as a vehicle to formulate graphic design input ought to work better than using words.

Could make it easier for people with little to no background within UX to start their project. Its of course not going to be able to take over the job of the UX-designer but as a supplement it could be helpfull.

It could really be used for giving non-designers a bit of a head start in the designing process.

It would speed up concept-working and allow designers to spend more time actually refining an already liked design instead of having to do multiple iterations of concepts. It could likely affect people currently working as concept artists who would have to adapt or slightly pivot their own role. I could also see that this opens up for people with minimal design knowledge to produce a concept they like and a feel they want to go for which they can present to design consultants if they lack that skillset or need to prototype for a pitch meeting etc.

Got the potential to cut down iteration and brainstorming sessions by quite a lot since you can get a large umbers of prototypes in a very short amount of time.

It will however be very dependent on the quality and “responsiveness” of the tool in my opinion. It needs to be able to give the designer a wide variety of outputs for the same / similar prompt.

If it is user friendly, easily accessed and produces OK results it would probably be quite well used. However, if it produces too good results, it might also crumble the whole UI/UX design area of work. :)

This would be a tool to accelerate the UI/UX design process, allowing designers to use this as a sort of “base” to build upon.

**Figure 46** Please elaborate briefly on how you think the widespread use of a tool similar to this prototype would affect the UX-designing profession as a whole?

**Figure 47** Continuation on previous question
Figure 48 Continuation on previous question

Could acts as a great starting point when brainstorming designs!