HAMK
Hämeen ammatti-
korkeakoulu

Utilizing AI in Design

Fine-tuning a Generative AI Model for Emil Aaltonen Museum

Bachelor's thesis
Smart and Sustainable Design
Spring 2024
Iris Kinni, Maija Lumiaho

HAMK
Hämeen ammatti-
korkeakoulu

Smart and Sustainable Design | Abstract

Authors Iris Kinni, Maija Lumiaho | Year 2023

Subject Fine-tuning a generative AI model for Emil Aaltonen Museum

Supervisors Helena Leppänen, Ville Siipola

This thesis documents the process of fine-tuning a generative artificial intelligence (AI) model for a client, Emil Aaltonen Museum, by adding an additional layer of training to an existing generative AI model with a personalized dataset to generate new images in a specific style and quality, in this case, realistic footwear images in the style of a Finnish footwear manufacturer Emil Aaltonen.

This thesis is commissioned by Emil Aaltonen Museum with the head of the museum, Mika Törmä, as our partner in this collaboration. As a result of this thesis, we will present a fine-tuned AI model for Stable Diffusion, and it will be integrated in an interactive website built for an exhibition held in the museum in January 2025 accompanied by a user manual to help people generate images. The model will be open for the museum's visitors to generate footwear images with, using their own prompts. The model will be trained using footwear images of Emil Aaltonen so that the resulting images will be imitating the design style of Aaltonen. This thesis will only document the process of training the AI model, and the website will be built, and manual will be written after the official thesis work is finished in order to delineate the amount of work that can be done within the time limits of this thesis.

For this thesis we will interview people working in the field of design who have utilized AI in their work. In the interviews, we will ask the interviewees about how their work relates to AI, the implementation and value of AI in a design process, the impacts of AI on sustainability, product development and production chains in the design field. The interviews are semi-structured, and the results are analyzed and presented in this thesis.

The thesis will be divided into three parts: first theoretical background information about AI and the methods and programs needed for fine-tuning is presented. The second part covers the practical process of fine-tuning an existing Stable Diffusion model, for which the training data is provided by Emil Aaltonen Museum. In the last part the use of AI in design is explored with the help of five interviewees and the results of the interviews and the whole thesis are presented. The purpose of the last part is to tie down the results of the practical work and the interviews to deepen the understanding of utilization of AI in design process as well as to answer the research questions set for this thesis.

Keywords artificial intelligence, fine-tuning, footwear design

Pages 51 pages and appendices 1 pages

HAMK
Hämeen ammatti-
korkeakoulu

Älykäs ja kestävä mutoilu Tiivistelmä
Tekijät Iris Kinni, Maija Lumiaho Vuosi 2023
Työn nimi Fine-tuning a generative AI model for Emil Aaltonen Museum
Ohjajaat Helena Leppänen, Ville Siipola

Opinnäytetyö dokumentoi generatiivisen tekoälyn (AI) kustomointia opinnäytetyön asiakkaalle, Emil Aaltosen Museolle. Generatiivinen tekoälymalli kustomoidaan kouluttamalla jo olemassa oleva AI malli yksilöllisellä datasetillä, jonka tuloksena voidaan luoda kuvia datasetille ominaisilla yksityiskohdilla. Tässä tapauksessa AI malli kustomoidaan luomaan realistisia jalkinekuvia Emil Aaltosen desgintyylillä.

Opinnäytetyön tilaaja on Emil Aaltosen Museo, ja yhteyshenkilönä toimii museonjohtaja Mika Törmä. Työn tulos on kustomoitu Stable Diffusion AI malli, jota käytetään sille omistetun nettisivun kautta museon järjestämässä näyttelyssä 2025. AI malli on vapaana käytettävissä näyttelyn aikana, ja käytön helpottamiseksi luodaan käyttöohjeet näyttelyn vieraita varten.

AI malli koulutetaan Emil Aaltosen valmistamien jalkineiden kuvilla, jotta mallilla luodut kuvat imitoisivat Aaltosen muotoilutyyliä. Opinnäytetyö dokumentoi ainoastaan AI mallin

kouluttamisen; nettisivujen rakentaminen ja käyttöohjeiden tekeminen tehdään virallisen opinnäytetyöajan ulkopuolella työmäärän rajaamiseksi.

Opinnäytetyössä haastatellaan myös muotoilun alalla toimivia ammattilaisia, jotka hyödyntävät tekoälyä työssään. Haastatteluiden tarkoituksena on selvittää tekoälyn tuomaa arvoa ja hyödyntämistapoja muotoiluprosessissa sekä tekoälyn vaikutuksia kestävyyteen, tuotekehitykseen ja tuotantoon muotoilun alalla. Haastattelut toteutetaan puolistrukturoituna, ja niiden tulokset sekä analyysit esitellään opinnäytetyössä.

Opinnäytetyö koostuu kolmesta osasta: ensimmäinen osa sisältää teoreettista taustatutkimusta tekoälystä sekä AI mallin kouluttamiseen tarvittavista ohjelmista ja metodeista. Toisessa osassa esitellään käytännön työ eli AI mallin kouluttamisen työprosessi. Kolmas osa sisältää haastattelut ja niiden tulokset sekä koko opinnäytetyön tulosten analysoinnin. Viimeisen osan tarkoituksena on sitoa yhteen käytännön työ ja haastatteluista saadut tulokset sekä syventää ymmärrystä tekoälyn hyödyntämisestä muotoiluprosessissa. Tämä osio vastaa opinnäytetyön alussa asetettuihin tutkimuskysymyksiin.

Avainsanat tekoäly, jalkinemuotoilu

Sivut 51 sivua and 1 sivu liitteitä

Contents

1 Introduction 1

2 Research 3

2.1 Information acquisition methods 3

2.2 Process description 4

2.3 Frame of reference 6

3 Background information 7

3.1 What is artificial intelligence (AI)? 8

3.2 Generative AI 9

3.3 Fine-tuning 11

3.4 LoRA, DreamBooth, textual inversion, hypernetworks 12
3.5 Prompting 14
3.5.1 Parameters 16
3.6 Hardware, code, and software 18
4 Requirements for fine tuning with LoRA and DreamBooth in Google Colab 20
4.1 Datasets 20
4.2 Preparing the data 21
4.3 Fine-tuning parameters 21
4.4 Code notebook in Google Colab 21
5 Fine tuning a generative AI for Emil Aaltonen Museum 22
5.1 Preliminary tests 22
5.2 Regularization images 24
5.3 Preparing image data 25
5.4 Captioning training images 27
5.5 Building LoRA model 27
5.6 Results of the LoRA model 33
5.7 Manual 35
6 Theme interviews 37
6.1 Sustainability 40
6.1.1 Environmental sustainability 40
6.1.2 Social and cultural sustainability 42
6.2 AI as a creative tool 44
6.3 Value of AI 46
6.4 Additional interview with Nicoline Van Enter 48
6.4.1 Implementation of AI in the design process 48
6.4.2 Value of AI in the design industry 49
6.4.3 The impacts of AI on product development and production chains 50

7 Conclusion 51
7.1.1 Sustainability of creating the customized AI generator 51
8 Reflection 53
Sources 55

Pictures and figures

Figure 1 Visual representation of thesis process description 6

Figure 2 Visual representation of frame of reference 7

Figure 3. Visual representation of diffusion process (*O'Connor, 2022*) 10

Figure 4. Visual representation of diffusion process 10

Figure 5. Refiner (*Stability AI, n.d.*) 17

Figure 6. Software and hardware needed for making a LoRA model 19

Figure 7. setting up eam_style LoRA model for Stable Diffusion 36

Figure 8 Affinity diagram from the interview data 39

Picture 1. A photo of dyne shaped sneakers generated with Stable Diffusion 15

Picture 2. A photo made with fine-tuned model, without regularization images 23

Picture 3. A photo made with fine-tuned model, with regularization images 23

Picture 4. A regularization image created with the basic version of Stable Diffusion ... 24

Picture 5. a regularization image created with Realistic Vision V5.1............................ 24

Picture 6. A footwear image from EAM before editing 26

Picture 7. A footwear image from EAM after editing 26

Picture 8. A highly descriptive caption for the prototype 1 training image 28

Picture 9. An image generated with the prototype 1 28

Picture 10. Updated caption for a training image for prototype 2. 29

Picture 11. An image generated with the prototype 2. 29

Picture 12. An image generated with the third LoRA model 31

Picture 13. An image generated with the third LoRA model 31

Picture 14. A control image without any additional parameters 32

Picture 15. An image generated with increased sampling steps................................. 32

Picture 16. An image generated with refiner 33

Picture 17. An image generated with hires.fix 33

Picture 18. Comparison between Aaltonen’s designs and images created with eam_style 34

Picture 19. An image generated with the prompt “a photo of a woman” 43

Picture 20. An image generated with the prompt “a photo of a Pakistani woman, traditional clothing, short hair” 43

Picture 21. Design from the AI Fashion Week 2023 (Prego, 2023) 45

Picture 22 A picture of the characters in Inworld Origins (*Steam, n.d.*)........................ 46

Picture 23. New.Arch.ai footwear sketch to AI generated images 49

Picture 24. Rejected regularization image generated with Stable Diffusion 53

Picture 25. Rejected regularization image generated with Stable Diffusion 53

Appendices

Appendix 1/1 Data management plan

1 Introduction

Fashion design is going through a massive change as the recent developments with artificial intelligence has enabled the exploration of AI assisted design. Designers and people working in the fashion industry are now taking advantage of AI by using it in the creative process and, in consequence, exploring whether it's possible that instead of replacing the role of designer, AI can be used to expand creativity and ideation based on designer's original ideas.

Most text-to-image AI generators are open for everyone and can be used for visualizing any creative ideas the user has, but sometimes the quality of the generated images can be questionable and therefore not suitable for professional design related tasks. However, generative AI models can be customized for specific tasks by utilizing a fine-tuning method, improving the quality and accuracy of the generated images as well as imitating a style of the images that the AI model is trained in, thus preserving, for example, a design style of any individual. This thesis offers an overview on how to fine-tune a generative AI model, what impacts generative AI has on the design process and explores the possibilities and threats AI poses in creative fields with an emphasis on footwear design.

The aim of this thesis is to create a customized AI model for Emil Aaltonen Museum. An existing Stable Diffusion model will be fine-tuned to generate images of footwear in the design style of Emil Aaltonen, a Finnish footwear designer. This thesis was commissioned by Emil Aaltonen Museum and the resulting fine-tuned AI model will be displayed in an exhibition held in the museum in January 2025.

For gaining knowledge on how generative AI can be utilized in the design field, the documenting of the practical work of fine-tuning an AI model will be accompanied by comprehensive background research on the topic of generative AI and interviews with five professionals who utilize or research AI in their work in the field of design.

The topic is selected based on the authors' shared interest and former studies on AI and design. This thesis is made without much prior coding knowledge, therefore video tutorials and online sources about coding will be utilized. Additionally, online code repositories, such as GitHub, will be used to find suitable Google Colab notebooks to run the training code and to run Stable Diffusion WebUI (web user interface). Stable Diffusion WebUI will be used to utilize the finished AI model and generate images. Hugging Face, a machine learning and data science platform will be used to store and access the final AI model.

Collaboration with Emil Aaltonen Museum

The client of this thesis is Emil Aaltonen Museum. Emil Aaltonen Museum is a Tampere-based Museum, displaying the footwear designs of Emil Aaltonen, among famous Finnish artwork, photography and Aaltonen's personal collections. Aaltonen was a footwear industry professional and the owner of Aaltosen Kenkätehdas Oy which manufactured timeless footwear designs through the 20th century. (*Pyynikinlinna, n.d.*) The museum holds several photographs of Aaltonen's footwear designs which will be used as training data for fine-tuning the resulting AI model.

The details of this project were discussed in a meeting with the head of Emil Aaltonen Museum, Mika Törmä. It was the client's wish that the resulting AI model would be able to preserve the footwear design heritage of Emil Aaltonen and if possible, portray images generated in Aaltonen's style in a futuristic manner. It was agreed that the model will be operated in English, i.e. the prompts to generate images with the model must be written in English, but to assist visitors in the use of the model, an instruction manual will be written in Finnish. The manual will be a simple signboard next to the project with instructions for using the model and constructing prompts. A computer and a screen will be provided by the museum operating as the necessary platforms for using the model during their exhibition as well as displaying the generated images for the visitors.

Another important takeaway from the meeting was that the images, that are the property of the museum, cannot be made publicly available, so necessary steps will be taken to prevent that data leaking to public through any website that might be used in the process of fine-tuning the model.

The result

The result of this thesis is a LoRA model, trained to replicate the design style of Emil Aaltonen. A website and a manual for the LoRA model will be done separately from the thesis work. Additionally, professional insight on the possibilities and impacts of generative AI in creative processes will be gained alongside answering the research questions described in the following chapter.

2 Research

This thesis is heavy on technical information and background information. It is written from the point of view of a beginner and attempts to open the concepts and terminology thoroughly so that someone who has no previous knowledge on the topic will be able to understand it. To combat the risk of approaching this thesis with too broad a lens, the topic has been narrowed down with the help of one main research question and three sub-questions. The research question and its sub-questions are:

1. What is the process of making a customized AI model for Emil Aaltonen like
    a. What is generative AI?
    b. How can generative AI be used as a tool in the design process?
    c. What kind of impacts generative AI has on the design field?

2.1 Information acquisition methods

This is a practise-based thesis consisting of practical work and theoretical background research. Background information is gathered from expert interviews and existing written sources such as web articles, professional literature, and academic papers. Research through the practical process, written sources and expert interviews aim to answer the research questions of this thesis.

The expert interviews are done with professionals working in AI and design: Gabriele Moschin, Andrea Filippi, Christian Guckelsberger, Petra Jääskeläinen and Nicoline van Enter. They all work or make research, in one way or another, with AI and creativity. Interview is chosen as a research method because as generative AI develops constantly, written sources become outdated fast. Interviews can provide a more current view of the topic as well as offer perspective from a more personal point of view.

Because the topic of AI design can be quite ambiguous, the interviews will be executed as semi-structured theme interviews, which allows broad discussion. According to the book

Tutkimushaastattelu - teemahaastattelun teoria ja käytäntö (Hirsjärvi & Hurme 2022) an interview is classified as semi-structured, when one part of the interview is fixed, but not all. For example, the interview questions can be the same for all interviewees, but the interviewer can vary the order of the questions. A semi-structured theme interview means that the interview proceeds according to certain themes instead of fixed questions. This leaves room for the interviewee's interpretation of the topic *(Hirsijärvi S. & Hurme H., 4.2.3. 2022)*.

The interviewees will be presented with a set of pre-prepared questions in advance to guide the conversation. Rather than strictly following the questions, the interviews will be adapted to the interviewees' own interests and knowledge. The themes of the interview are as follows: application of AI for design processes, the value of AI for the design industry and the impacts AI has on sustainability, product development and production chains in the design field. The interviews will be recorded as audio and a data management plan (*appendix 1*) is made to secure the processing of the interview data.

The interview materials are transcribed with Microsoft Word's transcribing tool and then cleaned with removing repetitive or irrelevant material. Then the materials will be analysed, and an affinity diagram will be made from the resulting data. The interview data of this thesis is qualitative, so the analysis will concentrate on finding commonly arising themes and connections between different subjects and phenomena. (*Hirsijärvi S. & Hurme H., 7.4 2022*)

2.2 Process description

The starting point for this research is a request for a customized generative AI model for Emil Aaltonen Museum. The process begins with background information acquisition of AI and the client. A meeting with the head of the museum is arranged for collecting information about the client's needs as well as footwear image data needed for the fine-tuning process.

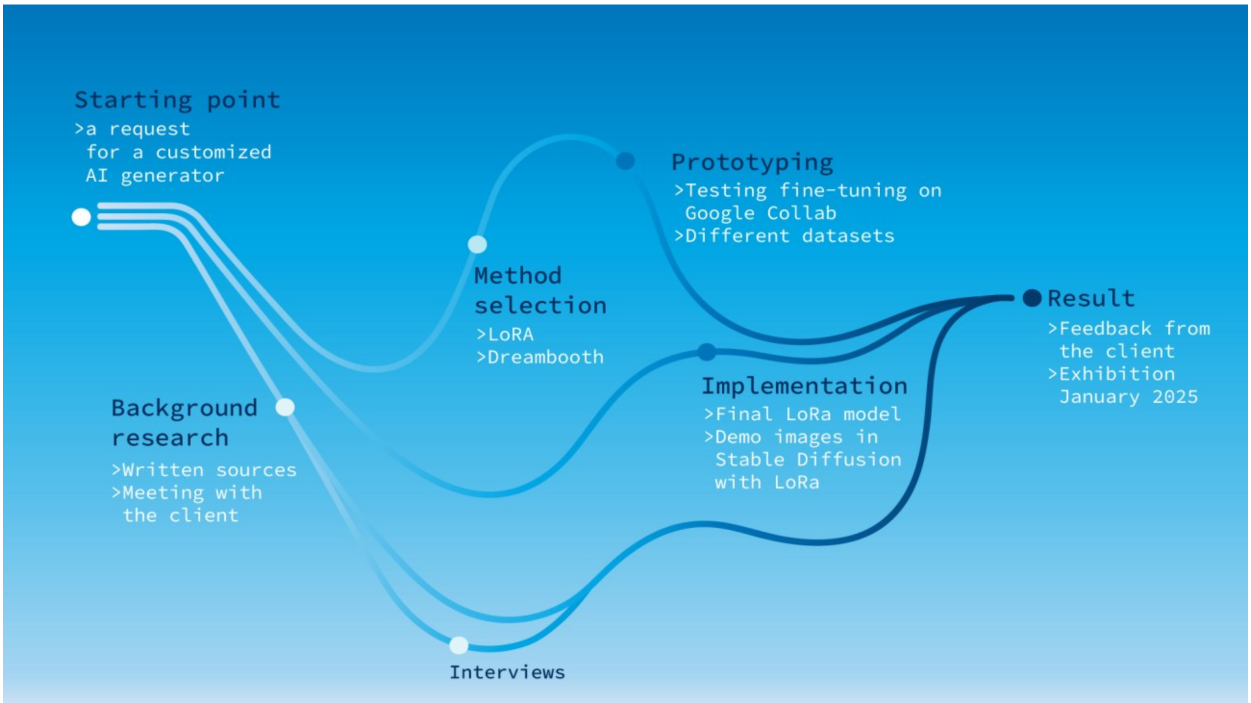
Then, expert interviews are done and recorded as part of background research and the recordings are transcribed and analysed afterwards.

Next, four different fine-tuning methods are researched and the most suitable one for this thesis chosen. LoRA, DreamBooth, Hypernetworks and Textual Inversion methods will be researched for this based on the amount of research and instruction material available online.

Prototyping of the fine-tuned AI model is done with testing the chosen fine-tuning method on Google Colab with a few different datasets. First tests are made with a with footwear images of a popular sneaker brand and the following tests with the data from Emil Aaltonen Museum. First phase of prototyping requires processing of the image data, which means captioning each image with a descriptive text. This is done automatically with a function called BLIP captioning after which they are adjusted manually. Next, regularization images are created with Stable Diffusion. Last phase, before actual fine-tuning can be started, is to set optimization settings, which affect the quality of the image generator.

Once the most optimal way of fine-tuning an AI model is found, and settings are correct, the final version of the model is trained and compiled into a repository. Example images are generated with the final model in Stable Diffusion and sent to the client for feedback. Adjustments are done accordingly. A visual representation of the process description can be seen in *figure 1.*

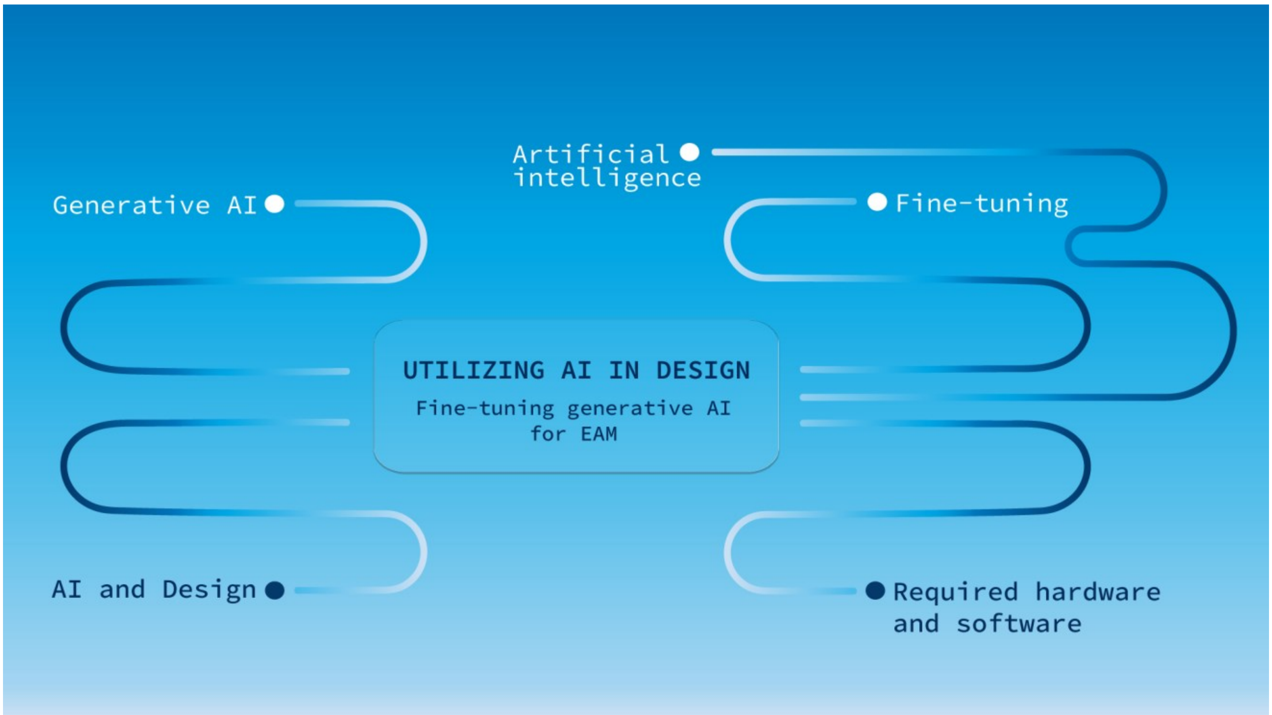
Figure 1 Visual representation of thesis process description



2.3 Frame of reference

Frame of reference for this thesis consists of theory on AI, generative AI, fine tuning, AI in design and basic information about the required software and hardware for fine-tuning. It will also include technical information about different fine-tuning methods with an emphasis on LoRA and DreamBooth methods. An overview of generative AI in design will be presented. A visualization of the frame of reference is pictured in *figure 2*.

Figure 2 Visual representation of frame of reference



3 Background information

Background information about AI is needed to understand the work that is being done for the client and this thesis. In these chapters some of the concepts such as AI, generative AI, what are prompts and what is fine-tuning are explained as well as the different methods of fine-tuning and their pros and cons, and the reasoning for choosing a fine-tuning method for the model. These chapters also include information about the software, hardware, open-source code, and cloud-based applications that are used to achieve the resulting fine-tuned model.

For the purposes of this thesis, we will concentrate on narrow AI i.e., the only artificial intelligence that currently exists. We will not be speculating on the topic of artificial general intelligence (AGI) which is a theory of a general artificial intelligence that most closely simulates human intelligence by possessing qualities that no current computer system can have, such as common sense, abstract thinking and cause and effect. (*Hashemi-Pour, n.d.*)

3.1 What is artificial intelligence (AI)?

Artificial intelligence as a term has many definitions and it has no exact meaning that is widely accepted. (*Wang, 2019, p.1*) Since the introduction of the term artificial intelligence at the 1956 Dartmouth Summer Research Project for Artificial intelligence, (*Dartmouth, n.d.*) the field has evolved into many different sub-fields of what we consider as AI, such as reasoning, planning, machine learning, vision, and natural language processing. (*Wang, 2019, p. 7*). Often what is presented as AI turns out to be one of the many sub-fields of AI instead, such as machine learning. (*Laskowski, n.d.*) Another reason for the difficulty to define AI is that our idea of intelligence changes over time and something that was previously thought of as AI is not necessarily considered intelligent anymore e.g., the first versions of AI that relied on human-input and rules. (*Petersson, 2023*) AI can be thought as an umbrella term to describe different methods, algorithms, code, and technologies used to create a machine that simulates human intelligence.

The technologies utilized in AI include machine learning, deep learning, and neural networks. (*Petersson, 2023*) Deep learning, neural networks and machine learning are all terms that sometimes are used interchangeably to describe AI but there are important differences between them. (*IBM, n.d.*)

Machine learning is a subset of AI, and it focuses on teaching a machine to mimic human intelligence. Essentially, it describes algorithms that allows the AI being trained to learn from previous data: it makes training AI simpler since it eliminates the need for human input for every variable of every decision the AI can take in any possible scenario. (*Petersson, 2023*) Machine learning makes many nowadays ordinary features of technology possible such as speech recognition and online recommendation engines, but it also has applications in radiology imaging in healthcare. (*IBM, n.d.*)

Neural networks are models of machine learning. As the name suggests, they are designed to mimic human brain and the connections within. (*Yasar, n.d.*) They consist of thousands to millions of processing nodes which are connected to each other and are assigned a number called “weight”. The nodes are arranged as layers: one input layer, one or more layers in between and lastly an output layer. The information from one layer to other moves usually only in one direction. (*Hardesty, 2017*) The nodes in the input layer are connected to nodes in the following layer which then in turn are connected to the nodes in the following layer until the output layer is reached. (*Yasar, n.d.*) Neural networks work by receiving raw data in the first input layer and then passing it along to the next layer if certain conditions are met e.g. if the previous node has a weight surpassing the decision-making node's weight, the data gets sent forward. If the weight is lower, it doesn't get sent forward. (*IBM, n.d.*) The weights determine the final output. The input is the raw data which is sent to the next layer where it gets transformed and multiplied until the output is produced, which is different from the raw data. (*Hardesty, 2017*)

Deep learning is a subset of machine learning, and it was developed based on neural networks. (*Petersson, 2017*) The word "deep" in deep learning refers to the depth of the network in neural networks (*IBM, n.d.*) i.e. how many layers there are in a neural network and through how many nodes the raw information passes before reaching output layer. The more layers a neural network has, the more accurate the output is. (*Saturn Cloud, 2023*)

The importance of defining AI for the purposes of this thesis is significant since throughout it, working with AI in practice as well as examining the role of AI in in design is done. It helps to achieve a precise defining of the topic and gain real insight about the applications of AI in the real world and within the topics we have chosen for our interviews. For the purposes of this thesis, we will use the term AI to mean computer models or programs that are programmed and trained to execute tasks that require human-like vision, language, and learning. (*Tarud, 2023*) More specifically, this thesis focuses on generative AI and the possibilities of personalizing it by using fine-tuning.

3.2 Generative AI

Artificial intelligence models that generate new content like images, text, audio or even code are called generative AI models. *(McKinsey & Company, 2023)* This is done by utilizing machine learning, that focuses on training a computer with specific algorithms to learn from past processed data without specific instructions to do so. The algorithm used in machine learning uses its experience to learn and improve in the task it is meant to execute. *(Rackspace Technology, 2023)*

Some of the most famous examples of generative AI are Open AI's text-to-image generator DALL-E 3 (openai.com/dall-e-3) and text-to-text generator ChatGPT (openai.com/chatgpt) and Stability AI's open-source text-to-image generator Stable Diffusion (stability.ai/stable-image). Stable Diffusion is used in this thesis, since it is open-source, and it supports most of the fine-tuning methods explored in this thesis.

Stable Diffusion, developed by Stability AI, is trained using a specific diffusion model, Latent Diffusion Model. This means that the AI is trained to generate compressed (latent) versions of the images, since it reduces the memory and computational time used for training and using the model. (*Onkar, 2023*) Diffusion models go through millions of images with an accurate text label to learn the relationship between words and images. *(Hypotenuse AI, 2022)* They work by destroying the data used to train it, which happens by progressively adding gaussian noise to an image until the image is completely unrecognizable by the end of the process. The purpose of adding noise to an image is to then learn the reverse diffusion, i.e., denoising the image, *figures 3. and 4. Introduction to Diffusion Models for Machine Learning (O'Connor, 2022),* that can then be used to generate new data. *(O'Connor, 2022)* To customize the results generated by the AI, we will be utilizing an existing method called fine-tuning.

Figure 3. Visual representation of diffusion process (*O'Connor, 2022*)

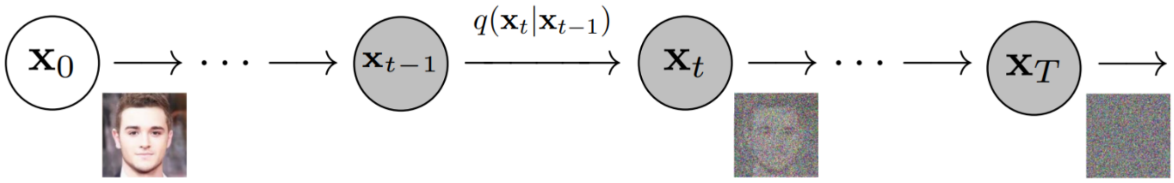
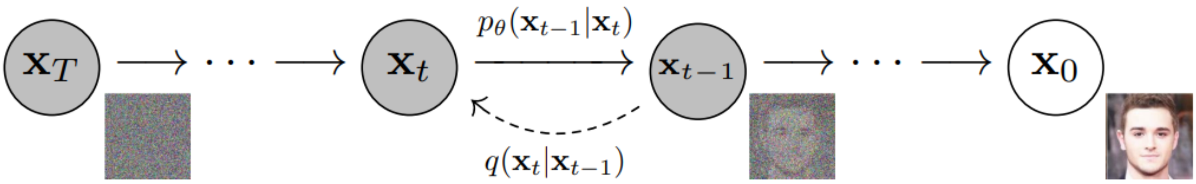


Figure 4. Visual representation of diffusion process



3.3 Fine-tuning

Fine-tuning is a process applied to existing generative AI models. It can be done with all kinds of generative AI models such as large language models (*Replicate, n.d.*), but the theory of this thesis will focus only on fine-tuning text-to-image models.

The aim of fine-tuning is to customize an AI model for a specific task by re-training it with custom datasets. The custom properties that can be achieved by this process can vary from style and tone to generating results of a specific object or person. (*Replicate, n.d.*) This means that the fine-tuned AI model is able to generate varying results of the object or style it has been trained with. (*Ruiz et.al., 2023, p. 2*)

Fine-tuned AI models can generate higher quality and more accurate results with less examples required in the prompt fed to the generative AI model. (*OpenAI, n.d.*) Prompts are textual descriptions of the wanted result, this is described in detail in chapter *3.5 Prompting*. Instead of a long and detailed description in the prompt, only one word assigned to the fine-tuned object is needed to describe that particular object. This word can be called a unique identifier, an activation word or a trigger word, depending on the source or codebase used for the fine-tuning process. Usually, it is recommended to use a non-existing English word as the activation word, so that the AI model is not confused to connect the word with an existing subject. (*Ruiz et.al., 2023, p. 4*)

The original AI model to which the fine-tuning process is being applied to, is called a pre-trained model (sometimes referred as base model). Fine-tuning allows the AI model to maintain previously learned skills of the pre-trained model but guides it to generate outputs from the smaller dataset it has been re-trained with. Since the process uses the pre-trained AI model as a base of the training, it requires less computer power than training an AI model from scratch. (*Srivastava, 2023*) One of the original research projects of fine-tuning was made successfully with a small dataset of five images (*Ruiz et.al., 2023, p. 3*), but it can be done also with thousands of training images, for example Realistic Vision, a text-to-image AI model fine-tuned for generating photorealistic images that has been trained with around 3000 custom images. (*Civitai, 2024*)

There are several different methods for fine-tuning generative AI models that differ in ways in which the custom data is being processed in the pre-trained model. This also changes the final form of the fine-tuned model: it can be a smaller version of the pre-trained model that is used in combination with original pre-trained model, or a copy of the pre-trained AI model called checkpoint model that can be used on its own. The fine-tuned models that are used in combination with pretrained models can, for example, be uploaded and run in Stable Diffusion. (*Stable Diffusion Art, 2023*) For the purposes of delineating the subject, four most used and researched methods, which are LoRA, DreamBooth, Textual Inversion and Hypernetworks, are described in the following chapter. They are all supported by Stable Diffusion.

3.4 LoRA, DreamBooth, textual inversion, hypernetworks

LoRA

LoRA (Low Rank Adaptation) is a fine-tuning method developed by a research group in Microsoft. In this method the fine-tuning process is being applied to only specific layers of the neural network of the pre-trained model, and the layers that are not trained, are kept frozen. (*Hu et al., 2021, p. 2*) This is done to reduce the training time and file size of the fine-tuned. LoRa models are trained with only feeding the pre-trained model a new, custom, dataset. (*Falbel, 2023*)

The result of LoRA is a small version of the pre-trained model that is used in combination with the pretrained model. LoRa models are computationally lightweight, usually 2-200MB in size (*Stable Diffusion Art, 2023*) compared to for example checkpoint model Realistic Vision V6.0 which is 1.99GB. (*Civitai, 2024*)

DreamBooth

DreamBooth is a fine-tuning method developed in collaboration with Google Research and Boston University. It is a method that utilizes a custom dataset paired with a unique identifier and a regularization image dataset *(in the original research this dataset is referred as a class-generated samples, but most codebases currently use a word "regularization image")*. The AI model learns to connect the unique identifier as the subject of the custom images. Regularization images prevent language drift, a common problem among fine-tuning methods, where the pre-trained model loses semantic meaning of words as new data is applied to it. The DreamBooth fine-tuning method trains all layers of the pre-trained model.

(*Ruiz et.al., 2023, pg. 4-5*) The result of DreamBooth training is a checkpoint model that can be used on its own, but it is also a large file that is computationally quite heavy. *(Antalpha.ai, 2023)*

Textual inversion

Textual inversion is a similar method of fine-tuning as DreamBooth. The goal is to teach the AI model to associate new “pseudo words” with the specific images it is given. For example, the developers of the textual inversion use a non-existing word S* to indicate certain images. The main difference to DreamBooth is that this process inserts new embeddings in the text encoder space of the diffusion model rather than modifies the original structure of the model *(Gal et.al., p. 2, 2022)*. Meaning that the result is not a completely new AI model, like in DreamBooth method, but a much smaller embedding file that can be placed into pre-trained AI models. In a user study made by the developers of the DreamBooth method, the quality of Textual Inversion and DreamBooth methods were compared, and it was found that DreamBooth method was preferred. *(Ruiz et.al., p. 7, 2023)*

Hypernetworks

Hypernetworks is fine-tuning method in which an additional neural network (called hypernetwork) is applied to a pretrained AI model. The custom data is processed in the hypernetworks, which then generates data for the pre-trained model. *(Ruiz et.al., p. 3, 2023).* Like in textual inversion and LoRA, the result of hypernetwork fine-tuning is a relatively small file as it does not change the entire pretrained model.

Conclusion

In DreamBooth method the fine-tuning process is being applied to the full AI model *(Ruiz et.al.., 2023, p. 4)* whereas LoRA, Hypernetworks and Textual inversion work only on specific neural network layers *(Antalpha.ai, 2023).* Therefore, DreamBooth can be considered a high-quality but time- and GPU-consuming option and the three other methods as fast and efficient ones. DreamBooth method has the benefit of using regularization images in order to prevent language drift. To achieve both the benefits of DreamBooth method and the efficiency of others, these methods can be combined. Open-source code is available for a combination of LoRA and DreamBooth fine-tuning, which was chosen as a fine-tuning method for this thesis work. Although the technique used in this thesis includes both LoRA and DreamBooth, the resulting AI model will be referred as a LoRA model, since it will be uploaded and run in Stable Diffusion as a LoRA model.

3.5 Prompting

Text-to-image generators generate outputs mainly with a text prompt and in addition, there are few parameters that can be used to enhance the results in different ways. A prompt is a textual description of the desired output, and they can be structed with a varying set of keywords. The keywords can be categorized, for example, by subject, medium, environment, lighting, colour, mood, resolution, and composition. Based on the authors' experience, every text-to-image generator works with slightly different prompting techniques, and these chapters only cover what kind of prompts are the most effective for Stable Diffusion.

Stable Diffusion is built with a dataset containing 2 billion text labelled images taken from publicly available internet. It can recognize most of the common words from the prompts, but some specific subjects that appear less in the image data, are not recognized as well *(Muppalla V. & Hendryx S. 2022)*. This is why fine-tuning method has been developed for teaching the AI new, specific, concepts.

Prompt structure

For good results, the prompts need to be clear and well-defined. The prompts are started with assigning a subject, which will be the focus of the image. Other keywords will guide the details and style of the output. For instance, the type of the image can be described with medium *(a photograph, an illustration or a digital painting)* and composition *(a portrait, a close-up or a wide-angle view)*. Colour *(saturated, pastel, muted)*, lighting *(studio, iridescent, cinematic)* and mood *(energetic, spooky, dreamy)* can be used to dictate the tone of the output. Resolution keywords can be added to the prompt to increase the sharpness of the images, for example: *highly detailed, 8k, sharp focus*.

Negative prompts

Negative prompts can be used for excluding certain features from image for example *poor anatomy* and *low quality* or other more concrete features such as *legs*. Negative prompts *lowres* and *worst quality* are typical for increasing the image quality. In general, text-to-image generators do not interpret text well, so including “text” in the negative prompt might prevent confusing text-like results. NSFW (not safe for work) can be used as negative prompt to filter out disturbing content or such that violates human dignity.

Example

To exemplify how prompts work, an example image *(picture 1)* is generated with Stable Diffusion text-to-image AI generator, with the following prompts:

Positive prompt: *A photo of dyne shaped sneakers in a desert, sand colours, golden sunlight, photorealistic, highly detailed*

Negative Prompt: *lowres, bad anatomy, cropped, worst quality, illustration*

Picture 1. A photo of dyne shaped sneakers generated with Stable Diffusion

Weights

Individual keywords can be emphasized with assigning the words a weight with specials characters. The weights are assigned differently in different ai engines. In Stable Diffusion, the keywords are weighted with (keyword: factor). The factor dictates how strongly the keyword is emphasized. The factor is a value of +/-1: values above 1 will give more emphasis and values under 1 will give less emphasis for the word. For example, the weight (sneaker: 1.6) will emphasize the word “sneaker” in the output.

Prompt scheduling

Keywords can be blended with prompt scheduling to get combined generations of two separate subjects. This method is used with the keyword weights (keyword 1 : keyword 2: factor). For example (sneaker : heel 0.5) would result in the AI model to switch generating an image of a sneaker to a heel at a half way of sampling steps. The result would in that case have equal features of both, a sneaker and a heel.

3.5.1 Parameters

Seed value

The seed is a value that can be used for varying the results of the same prompt. Changing the value of the seed but keeping the prompt set, will generate slightly different results. *(Muppalla V. & Hendryx S. 2022).* With same prompt and seed value, the result will always be the same. Seed value of –1 will generate random results. *(Andrew, 2022)*

Sampler methods

Sampler is an algorithm of a diffusion model that is responsible of the denoising process *(Neuroflash, 2023)*. Using the same prompt and other parameters but changing the sampling method would result in slightly different results. Stable Diffusion has plenty of different options of what sampling algorithm to use in generating the images for example, Euler, DDMI and KDPM Karras.

Sampling steps

Sampling steps is a parameter that controls how many iterations the AI model takes to generate an image in diffusion process. Higher number of sampling steps equals higher number of iterations and therefore a longer generation time but usually more high-quality image. *(Bagu, 2023)*

CFG scale

CFG scale controls how strictly the AI model follows the given prompt. The default value of this scale is 7 in Stable Diffusion: lower values give more “creative” results whereas setting a higher value generates more exact images. *(Andrew, 2022)*

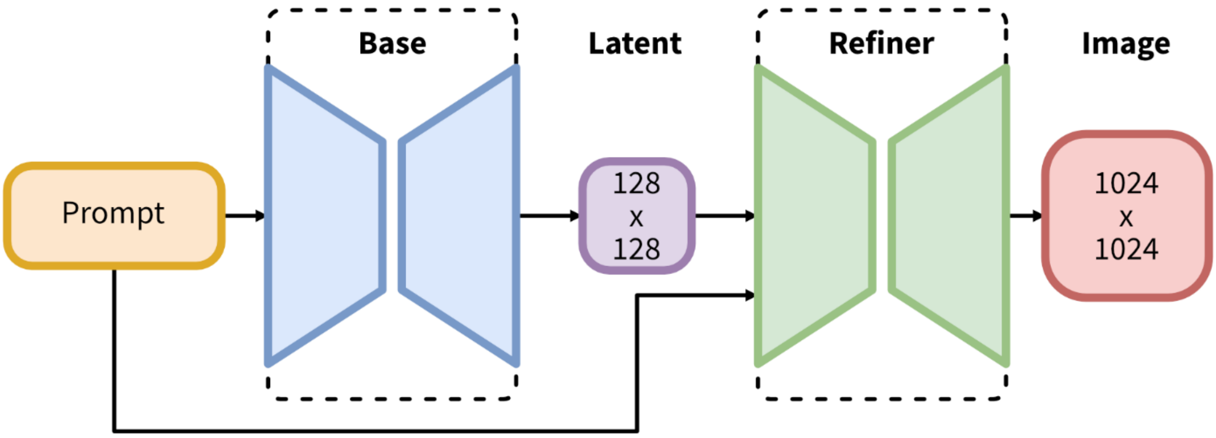
Hires.fix

Hires.fix is a parameter used for increasing the resolution of the generated image. Stable Diffusion 1.5, which is used in this thesis, can create images only with the resolution of 512x512, but using hires.fix can be used to upscale the image. *(Barretto, 2023)*

Refiner

Refiner is an extension that is used in addition to the base AI model to enhance image quality. Refiner is given a value that indicates the point in which the refiner takes over of the denoising process. It is guided to set the value to 20% which would mean that the base model is responsible for the first 80% of diffusion process and the refiner of the last 20%. *Figure 2* shows the how the refiner is placed in the AI model. *(GitHub, 2023)*

Figure 5. Refiner (*Stability AI, n.d.*)



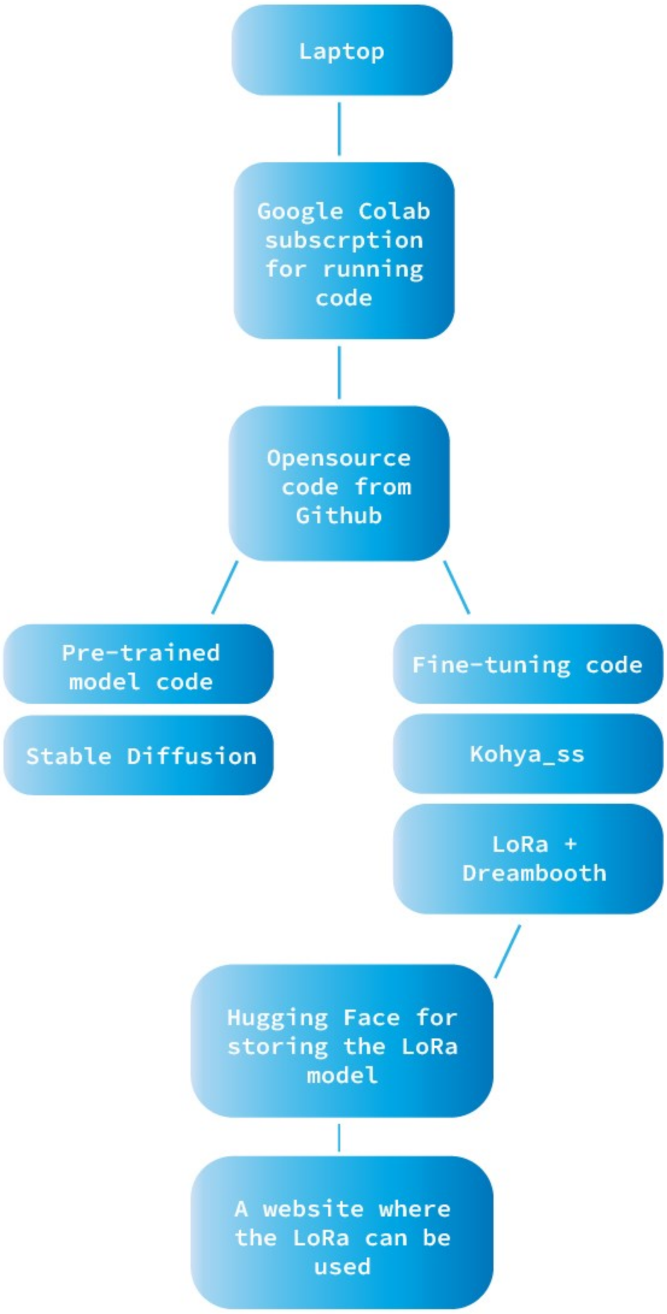
3.6 Hardware, code, and software

To train an AI model, a computer must have a certain amount of computing power. According to different online tutorials, it was determined that the recommendation is that a laptop has at least 512GB SSD memory, 32GB RAM, a recent NVIDIA GPU, 9th generation Intel Core i7 or better CPU and at least 12GB VRAM.

As discussed in the previous chapter, it was decided that LoRA and DreamBooth would be used for fine tuning. Training a LoRA model requires using a Python library called Kohya_ss, which is used for fine-tuning Stable Diffusion models *(Kohya-ss: A Friendly Stable Diffusion Fine-Tuning Library, n.d.)* Kohya_ss can be run in Google Colab, which is a hosted Jupyter notebook service which gives users access to computing resources such as GPU's, (*Google Colaboratory, n.d.)* but to work it requires a paid subscription. Kohya_ss can also be run locally and to work, it requires installation of Python 3.12.0, Git 2.42.0, Visual Studio 2015, 2017, 2019, and 2022 redistributable and PyTorch 1.12.1. Installing all these take up a significant amount of a laptop's SSD memory and when tested, Kohya_ss could not be run locally in neither of the authors' computers. Because of this issue, the best option to fine-tune a LoRA model using Kohya_ss is to train the model using Google Colab. In addition to giving access to significantly greater amounts of computing power, Google Colab is easy to use in a collaborative manner since both authors of this thesis can gain access to same notebooks through a shared Google account created for this project.

The basic version of Google Colab is free, but the work needed to be done for this project requires a paid subscription. It uses 'units' as currency to access computing power such as GPUs, and the more computing power is used through it, the more units are used. Google Colab Pro+ subscription comes with 500 units and there is an option to purchase more if needed. Additionally, Google Colab is a preferred option to use for this work because there are a number of notebooks built by an active community surrounding it: many of the notebooks used in this work can be found from users on GitHub or Hugging Face. When the final LoRA model resulting from this thesis is ready, it can easily be transferred from Google Drive to be stored in Hugging Face and be used later to integrate into a website that will be built for the exhibition for which this project was commissioned. The relations of the all the software used in thesis is shown in *figure 6.*

Figure 6. Software and hardware needed for making a LoRA model



4 Requirements for fine tuning with LoRA and DreamBooth in Google Colab

Fine-tuning with LoRA and DreamBooth in Google Colab code notebook requires the following steps: gathering datasets, preparing of the datasets, setting correct fine-tuning parameter values, and running the code cells. The next chapters describe each phase in detail, and they follow guides found from the study from Google Research and Boston University *DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation* as well as the general instructions of the open-source code notebook for fine-tuning with LoRA and DreamBooth. In addition, tutorials from online communities that are sharing code and instructions on AI related topics were utilized.

4.1 Datasets

Fine-tuning a LoRA model requires two image datasets: training images and regularization images. Training images are images of the specific object or subject that the model will learn to replicate, in this case the footwear images from Emil Aaltonen Museum. Regularization images represent the general class of the training images: footwear.

Regularization images will influence how the generated images are presented. They give the final pictures the photorealistic style, preferred angles, quality, and variety. In addition, regularization images prevent the risk of language drift. *(Ruiz et.al., 2023, pp. 4-5)* This means that the AI model learns to maintain knowledge of how general footwear looks like while it is being fed new data of the style of Emil Aaltonen footwear.

Emil Aaltonen Museum has around three hundred images that will be used as training images. The amount of regularization images depends on the amount of training images, but the exact required amount is unclear. For example, according to some sources a following formula is to be used: (number of training images) * (100)) * 1. That means the required amount of regularization images would be (300*100) * 1 images which is roughly 30 000 images. Some sources suggest using (num_epochs * num_samples), which would equate around 6000 regularization images. Despite of the unclarity, it became obvious that the number of regularization images must be significantly higher than the number of training images.

4.2 Preparing the data

Before starting the fine-tuning process, both datasets need to be prepared. This simply means clearing out low-quality images and organizing them in their own folders. Some of the images need to be edited to make sure no distractions are present in the images and won't get trained into the model. In addition, the training images need to be captioned with a description of the image and with an activation word.

The training images can be captioned with an automatic function in the notebook that is called BLIP captioning. BLIP captioning creates a text-file for each image, containing a simple textual description of the image. These automatically created descriptions usually need to be adjusted manually since they might be incorrect and insufficient. The goal of the caption is to describe the main elements of the image such as colour, material, and type of the shoe.

4.3 Fine-tuning parameters

For the AI model to learn the custom properties from the image data, there is a set of parameters in the notebook that must have correct values for the fine-tuning process. These hyperparameters are the number of images, repeats, epochs and the batch size as well as the value of u-net learning rate, text encoder and scheduler. The values assigned to these parameters dictate how the algorithm processes the image data and impact on the quality of the resulting LoRA model. The creator of the notebook has instructions on what values to use for specific cases.

4.4 Code notebook in Google Colab

The open-source code notebook for LoRA and DreamBooth fine-tuning can be found from a collective code repository called GitHub. The notebook contains all code files and instructions for running the code locally or in Google Colab cloud notebook. The notebook is used in Google Colab by inputting required datasets and editing parameter values in the notebook and running the code cells.

The downside of using open-source notebooks in Google Colab is that the notebooks can become outdated quite fast, especially because fine-tuning requires many different software installed and update to work properly. The software can be updated inside of the notebook, but it requires some knowledge of coding.

5 Fine tuning a generative AI for Emil Aaltonen Museum

5.1 Preliminary tests

Before fine-tuning the final LoRA model with Emil Aaltonen footwear images, the basic mechanisms of fine-tuning were tested using datasets consisting of images of footwear from a popular and recognizable sneaker brand. These preliminary tests were made for finding out what kind of datasets would be necessary for teaching an AI model to imitate a particular design style and how Google Colab code notebooks work.

The first test was made with only 60 training images and zero regularization images. The results were inadequate: the AI model did try to imitate the style and logo of the brand, but it was not recognizable, and the photos lacked quality and accuracy *(Picture 2).*

For the second test, 282 regularization images made with Stable Diffusion were used alongside the 60 training images, which improved the accuracy of the results. The AI model learned to replicate the logo and the style of the brand. However, the quality of the images was poor, e.g. the AI model created odd materials for the sneakers and cropped the images so that the whole shoe did not fit the picture *(Picture 3).* This is suspected to be caused by the low amount and poor quality of the regularization images.

Picture 2. A photo made with fine-tuned model, without regularization images

Picture 3. A photo made with fine-tuned model, with regularization images

With these tests, it was discovered that the regularization images highly impact the quality of the results. After this discovery, two prototypes of the model were made with a focus on the quality and the amount of regularization images. The next chapters describe what kind of datasets were made for those prototypes.

5.2 Regularization images

Preparing the datasets started with creating regularization images. The regularization images used in the preliminary tests were created using Stable Diffusion V1.5 and resulted in low quality images, which affected the model negatively. For this reason, it was decided to use a Stable Diffusion checkpoint model called Realistic Vision V5.1 for the final regularization images to achieve a high-quality dataset for fine-tuning. *Pictures 4* and *5* show the difference of images created between these two Stable Diffusion models.

Picture 4. A regularization image created with the basic version of Stable Diffusion

Picture 5. a regularization image created with Realistic Vision V5.1

In total, around 3000 regularization images were created with a variety of different types of footwear: leather boots, high heels, court shoes, sandals, sneakers, dress shoes and futuristic footwear. Different positive prompts were used for creating the regularization images, e.g. “a pair of court shoes made with leather, low heel, saturated colours, empty background, studio lighting, photorealistic”.

Sneakers and futuristic footwear images were included in the dataset to increase diversity and to give the LoRA model the capability to create different types of shoes. These images were added to training data because the client requested the model to have the ability to portray a variety of footwear in Emil Aaltonens style, even futuristic footwear.

Since generative AI models often fail at replicating anatomy correctly, it was decided to attempt to exclude legs from regularization images by using negative prompts. Negative prompts were also used to eliminate backgrounds, such as furniture or landscapes, from the generated images. Some negative prompts used for the generated regularization images were: “person, feet, low quality, text, poor anatomy, background props”.

The regularization images were generated with Stable Diffusion WebUI, and it was run with a Google Colab notebook to speed up the process of generating the images: generating one image utilizing the computing power that Google Colab offers took around 1 second – compared to 5 minutes for one image when a test was made to generate the regularization images by running Stable Diffusion locally. Creating 3000 regularization images consumed a lot of Google Colab units, in total over 500 units. The authors suspect this to be because Stable Diffusion, when launched through Google Colab, runs on a browser-based user interface called Stable Diffusion WebUI. This is a large number of units compared to running the fine-tuning code that only took around 50 units and didn’t require a user interface.

5.3 Preparing image data

Next phase was to prepare and clean both datasets (regularization images and training images). Any images with blurriness, weird cropping or unwanted details were discarded or edited to be more suitable for the fine-tuning process. Some of the images provided by Emil Aaltonen Museum had disruptive elements in them, such as open shoelaces and tags hanging from the shoes. In addition, some shoes were photographed on a patterned or otherwise complicated background. These images needed to be edited to make sure the data is clear, and that no open shoelaces or tags get trained into the model. The images were edited with Adobe Photoshop and most of the disrupting features were removed using its new AI tools such as generative fill and generative expand (example shown in *pictures 6 and 7*). Some of the images could not be edited i.e. the quality was too bad or the view for most parts of the shoe was obstructed by shoelaces or tags. 204 out of the 304 images provided by the client ended up being suitable to be used for fine-tuning.

Picture 6. A footwear image from EAM before editing

Picture 7. A footwear image from EAM after editing

5.4 Captioning training images

Next step of preparing the image data was to caption the training images with the automatic BLIP function and to manually adjust them afterwards. This part required many attempts, because the AI model is supposed to learn the content of the images based on the caption. This is described more depth in *chapter 5.5*.

In addition to teaching the model to identify the content of the image from the caption, it is supposed to learn to connect a specific word (activation word) with the style of the image. An activation word ”eam_style” was added to each caption in order to teach the AI model to make a connection between that word and the style of the Emil Aaltonen footwear images. A prefix “a photo of” was also added to each caption, so the AI model would create photorealistic images, and not for example, illustrations.

After both datasets were prepared and the training images captioned, they were placed into Google drive in their own folders called “regularization images” and “training images”. All zip-files were extracted to an open file form.

5.5 Building LoRA model

This chapter documents two different prototypes of the LoRA models that were fine-tuned with the datasets described in previous chapters. An open-source Google Colab notebook made by GitHub user Linaqruf was used to train the models. The notebook is meant for fine-tuning a pre-trained Stable Diffusion model using DreamBooth, LoRA and Kohya_ss. All the prototypes were made with default parameters found from the LoRA fine-tuning code notebook.

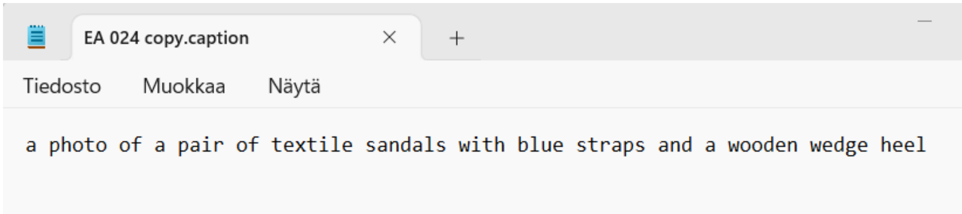
Prototype 1

Prototype 1 was made with ~3000 regularization images and 204 training images with highly detailed captions that contained all the features of the footwear images such as material, colour, type of the shoe and other important details such as buckles and height of the heel.

The fine tuning was successful in the broader sense but unfortunately the results were not what was hoped. The shoes did not have the style of Emil Aaltonen, so the images were quite regular shoe images. The problem seemed to be the captions that were very detailed and long for all the images since everything that is described in the caption do not get trained into the model.

It was discovered that the AI model learns to connect the “activation word” (eam_style) with subtle details of the image, that are not described in captions. Meaning that if the captions are too descriptive, there is nothing for the AI model to pick out as style of the image. An example of a training image caption and an image created with the prototype 1 are shown in *pictures 8* and *9*.

Picture 8. A highly descriptive caption for the prototype 1 training image



Picture 9. An image generated with the prototype 1

Prototype 2

Prototype 2 was made with the same 3000 regularization images, but the captions of the training images were edited to be less detailed, just describing the type of shoe pictured e.g. a high heeled boot. Unfortunately, since the AI technology goes forward very fast, the notebook that was used got outdated while new captions were being made and it was necessary to add additional code to the notebook manually to update some parts of the notebook. With the updates and new captions, it was possible to start training the model again, but the results were still underwhelming. The model took on some of the characteristics of the style that was hoped for but not much. For example, prototype 2 was able to imitate a particular textile texture that can be found from the training images, but otherwise the style was not recognizable. An example of a training image caption and an image created with the prototype 2 are shown in *pictures 10* and *11*.

Picture 10. Updated caption for a training image for prototype 2.

```
eam_style, a pair of sandals
```

Picture 11. An image generated with the prototype 2.

Final LoRA model

The issue in the two first prototypes seemed to be that the algorithm was still not able to connect the activation word with the content of the images. A new round of testing was done after the captions were edited further and arranged into a proper way, naming the files and folders in a coherent manner for the program to find them and utilize them as well as adding the activation word chosen for this project, eam_style, in the name of the training image files. The results of this test are significantly better, and this third model ended up being the final LoRA model. The style of Emil Aaltonen can be detected from the images and with proper prompting it's possible to generate even humorous images that combine the old-fashioned style of Emil Aaltonen with newer and more modern style of footwear *(Pictures 12 and 13).* The files of this model were uploaded to HuggingFace, so that in the future the AI model can be integrated into a website. The LoRA model is stored in HuggingFace as a private model, as the client had wished that the model or the training image dataset wouldn't be publicly available.

Picture 12. An image generated with the third LoRA model

Picture 13. An image generated with the third LoRA model

Parameter tests

When the final LoRA model is placed into Stable Diffusion, there is a of set parameters which affect the image quality and the intensity of the LoRA model. The effects of all the parameters were tested with generating images with identical prompts to find out which combinations generate best results. First a control image without any additional parameters was generated *(picture 14).*

Increasing the value of sampling steps, which controls the iterations of the denoising process *(Omniinfer, 2023),* generated the most high-quality and accurate results *(picture 15)*. Using parameters called hires.fix *(picture 16)* and refiner *(picture 17)* tremendously decreased the image quality.

Picture 14. A control image without any additional parameters

Picture 15. An image generated with increased sampling steps

Picture 16. An image generated with refiner

Picture 17. An image generated with hires.fix

5.6 Results of the LoRA model

For the most part, the final fine-tuned LoRA model was able to imitate the style Emil Aaltonen design well in terms of materials, colours, and the shape of the footwear. From most images, it can even be detected which Emil Aaltonen footwear image the AI model has used as a reference image. For example, *picture 18* shows comparison between shoes designed by Emil Aaltonen and images generated with eam_style LoRA model: pictures on the left are Aaltonen's and pictures on the right are generated with the model. The eam_style LoRA model clearly imitated the type of the heel as well as materials and colours of one Emil Aaltonen footwear design, but still added some variation to image. This was quite of an accomplishment, because it means that the LoRA model can be used to produce new designs, and to continue the cultural heritage of Emil Aaltonen.

Picture 18. Comparison between Aaltonen's designs and images created with eam_style LoRA

The LoRA model got more advanced throughout the prototyping phase, and in the end most of the disruptive qualities were managed to eliminate from the final LoRA model. For example, the image quality was improved significantly as well as the model's ability to generate photorealistic images of footwear.

Of course, some of the images generated with this AI model are flawed, because in general, shoes are difficult objects for text-to-image generators as they contain multiple small details, such as eyelets. If the prompts are too vague, the images tend not to have any custom properties.

The client was asked to give feedback on the results and especially to comment if the LoRA model can generate images that resemble the style of Emil Aaltonen well enough. The client was pleased with the results and stated that the images resemble Aaltonen's work, and the model can be thought to create a continuation of his previous collections.

5.7 Manual

At this stage, the LoRA model can only be used in Stable Diffusion WebUI, which is run through Google Colab. The LoRA model is stored in HuggingFace as private model, and only the authors of the thesis have access to it. Access to the LoRA model can be granted for the thesis coordinators if needed for the evaluation of this thesis. *Figure 7* contains guides on how to download the LoRA model and how to use it in Stable Diffusion, including technical information on which parameters and prompts are recommended.

Figure 7. setting up eam_style LoRA model for Stable Diffusion

Settings for eam_style LoRa in Stable Diffusion	
1.	Download eam_style LoRa model from the authors' Hugging Face model card.
2.	Open Stable Diffusion web UI from Google Colab.
3.	Upload eam_style LoRa to Stable Diffusion web UI from LoRa tab and activate it by clicking the appeared model.
4.	Open the settings of the eam_style LoRa, adjust the weight of LoRa to a recommended value of 1,0.
5.	Click the generation tab > a prompt <lora:eam_style: 1.0> has been added to the prompt box.
6.	**Recommended positive prompt structure:** <lora:eam_style: 1.0> a picture of a pair of court shoes, studio lighting, realistic style, high quality, product photography
7.	**Recommended negative prompt:** lowres, bad quality, feet, illustration, anatomy, background props
8.	Increase sampling steps to increase the image quality: can be set up to 150.
9.	Experiment with different samplers: Euler a, LMS, Heun and DPM2 generate good results.
10.	Using refiner and Hires.fix is not recommended.

6 Theme interviews

This part of the thesis discusses AI in a design context with the help of five interviewees: Christian Guckelsberger, Petra Jääskeläinen, Gabriele Moschin and Andrea Filippi, and Nicoline van Enter. For the sake of defining this subject, the analysis of these interview materials concentrates on topics of using of creative AI as a tool in design, sustainability issues in creative AI and what value AI brings to creative work or design, although the interviews included great discussion about the nature of general artificial intelligence as well. These expert interviews aim to clarify what is the current state of AI assisted design through a professional perspective.

All the interviewees are specialized in different areas of AI and the creative industry. The interviews are semi-structured, and the focus of the interviews varied according to the area of specialization of each interviewee. The interviewees are as follows:

Christian Guckelsberger, a Computer Scientist, Art historian and Assistant Professor of Creative Technologies in Aalto University.

Gabriele Moschin, the Head of Education of Fashion Design and Design Business at NABA.

Andrea Filippi, a Graphic and Packaging designer at Giorgio Armani.

Petra Jääskeläinen, a Researcher of Ethics and Sustainability of Creative AI/AI arts at KTH.

The questions that were sent to the interviewees in advance were following:

1. What do you do and how does it relate to AI and creativity or design?
2. How can people implement creative AI in design or creative work?
    a. What kind of creative work or design work have you recently done with the assistance of AI?
3. What value does AI bring to the creative industry or design industry?
    a. How can AI impact product development chains?
4. What kind of impact AI has on sustainability (environmental, cultural, or social) in creative or design context – is it good or bad?

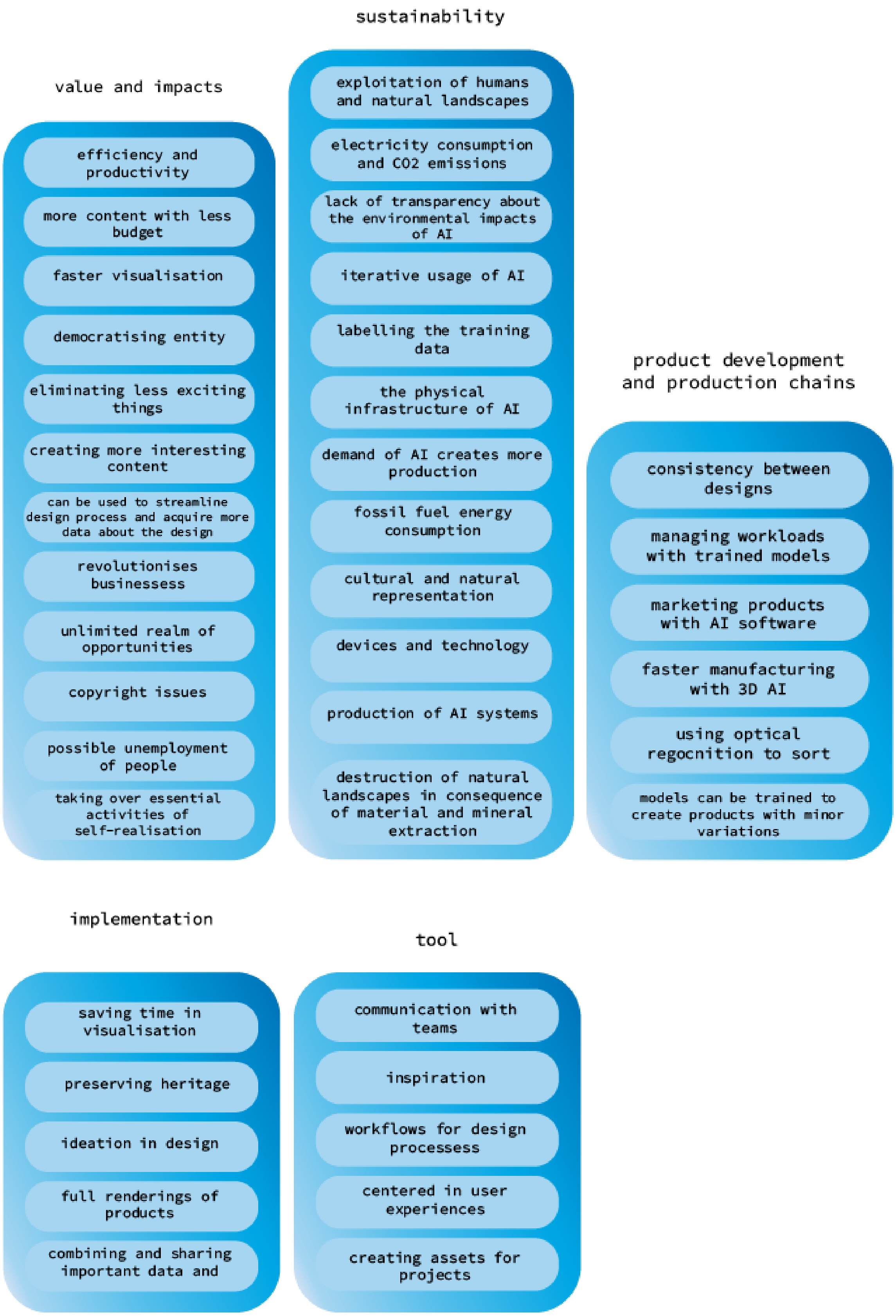
In addition, a complimentary interview was done with the co-founder and CTO of Footwearology, Nicoline van Enter, to get concrete answers on how to use AI specifically as a tool in the footwear design process. The decision of doing a fifth interview was done after interviewing the first four participants, as the data was not yet sufficient to answer all the research questions of this thesis.

Footwearology provides AI for Footwear courses and consulting for companies using AI in the design process and therefore van Enter has a lot of experience in using AI specifically in the context of footwear. The original interview questions were refined and adjusted to gain more concrete answers. Question number four from the original interview questions, "*What kind of impact AI has on sustainability (environmental, cultural, or social) in creative or design context – is it good or bad?*" was left out, since the four first interviewees answered the question comprehensively. The interview questions sent to Nicoline van Enter were:

1. What do you do and how does it relate to AI and design?
2. How can AI be implemented in a design process?
3. What value does AI bring to design industry?
4. How can AI impact product development and production chains in the design industry?

All the recorded interview data was analysed by transcribing the audio into a text format and picking up recurring themes across the interviews from that data. A figure of the answers is presented below to made to make analysing the interview data easier.

Figure 8 Affinity diagram from the interview data



6.1 Sustainability

Sustainability does not only include environmental impacts but also social and economic ones *(Penty, 2019, p. 16),* so the aim was to discuss holistically about sustainability of AI in the interviews. In our interviews, the sustainability of AI was revealed to be a complex problem that includes environmental, economic, and social aspects.

6.1.1 Environmental sustainability

The first thing Gabriele Moschin brought up in our interview was the sustainability of AI. He pointed out that during a recent workshop he and Andrea Filippi held for 50 students, around 200 images were generated per person. This means that for one workshop, 10,000 images were generated using AI. The reason he highlighted this aspect of generating new content in their workshops is because generating images with AI has a massive impact on the environment.

Other interviewees, Christian Guckelsberger and Petra Jääskeläinen, mentioned the fact that training and hosting AI is an energy intensive process and produces a considerable amount of CO_2 emissions. Additionally, Guckelsberger stated that he is not sure whether people are sufficiently aware of, or whether there even is enough transparency to make it possible to know, about the real environmental costs of AI. In addition to training and hosting AI, he pointed out, most of these systems used for generating content are used in an iterative fashion which further increases the environmental impact.

These statements are supported by recent studies. A life cycle assessment for training several AI models, performed by researchers in the University of Massachusetts Amherst, found that the training process can emit over 600,000 pounds, or 272 155 kg, of carbon dioxide. (*Strubell et al., 2019, p. 1*) Another study by researchers at Hugging Face, a platform also utilised in this thesis, found that generating one image with a large AI model consumes as much energy as fully charging your smartphone. The study shows that while training an AI model consumes massive amounts of energy, the usage (generating content) of these models is responsible for most of the carbon footprint associated with AI.

In relation to this thesis, the study also found that generating 1000 images with Stable Diffusion XL generated the equivalent of carbon dioxide that an average gasoline-powered car emits in a 4,1-mile drive. (*Heikkilä, 2023*) Even though the roughly 3000 images for our fine-tuned model were generated with a previous model, Stable Diffusion V1.5, it can only be assumed that the carbon footprint must not be much less.

Another environmental sustainability issue that Jääskeläinen talked about concerned the materials for the servers that are needed for AI applications and the origin of those materials. Extraction of these minerals and raw materials causes violence towards humans and natural landscapes at the extraction site. She says that since AI can be used with many different devices, it’s hard to locate the origin and to assess the sustainability of all the materials used for all the devices associated with AI. She said that often at the end of their lifecycle, the devices get disposed in the same low-income countries where the materials came from.

Jääskeläinen also said that consumers of AI might be under the impression that AI applications in general are based exclusively in some sort of cloud. This might furthermore push the idea that AI does not affect the real world and the environment. Unfortunately, that is not true, like Jääskeläinen points out in our interview by saying that to work, AI needs a large amount of raw resources and infrastructure in the physical world. Moschin also brought up the same topic saying that the server farms needed to train and run AI services need electricity to work and if the origin of the electricity lies in oil and carbon, it increases the pollution associated with AI. Due to the capitalist nature of our society, demand for more AI products and software creates more supply and that in turn creates more server farms and more emissions and exploitation of people and natural resources.

Jääskeläinen also points out the fact that even if customers and consumers of AI know and understand the implications of their AI usage on the natural world, they might not care about it. She asks whether it is useful to merely point out the data and show the consequences if the health of our environment is not important to users.

6.1.2 Social and cultural sustainability

In addition to issues with environmental sustainability, AI produces a plethora of social sustainability issues. One of these issues concerns labelling the training data for AI, which is often delegated to people in low-income countries: lower income allows corporations to pay people very little for this monotonous labour that is essential for the development of AI models.

The Verge published an article in collaboration with The Atlantic in which was detailed the tedious and low-paying job of data labelling in Nairobi, Kenya. The article described the job of Joe, a college graduate making extremely low wages while working the essential job associated with AI, for example earning 10 dollars for eight hours of work annotating a seconds long clip of footage. The article paints a picture of people working for a multi-billion-dollar companies with dollars a day and working a job what most people don’t even know to be a job. (*Dzieza, 2023*)

Since the basis of many artificial intelligence models is to source the training data online, it represents the content that gets uploaded the most. For example, the results of many text-to-image generators are stereotypical and rarely represent the reality. Some cultures, such as Australian or New-Zealand Aboriginal cultures, are missing almost entirely from all the representations of humans that AI makes. Jääskeläinen pointed out that even though there has been done work for removing ethnicity and gender bias from generative AI, it still often generates, for example, very binary and western representations: if a person would generate an image of a woman, the result would most likely be a long haired person, and if the image would be of a man, the result would most likely be a short haired person. This phenomenon can be seen in *picture 19* where a prompt “a photo of a woman” was used to generate an image in Stable Diffusion to see, what kind of representation Stable Diffusion has learned from the data it has been trained on. *Picture 20* is generated using a more detailed caption, specifying the ethnicity, clothing, and hair length of the subject: “a photo of a Pakistani woman, traditional clothing, short hair”.

Picture 19. An image generated with the prompt “a photo of a woman”

Without any specifications in the prompt, Stable Diffusion generates an image of a woman with very western and binary features. Only when it is specified in the prompt itself, the AI will generate a person of a different ethnicity with non-western clothing.

Picture 20. An image generated with the prompt “a photo of a Pakistani woman, traditional clothing, short hair”

It was also discussed that AI is most often trained and used with English language, which strengthens the idea that the user group intended for AI is mostly Western. One of the interviewees told that there has been development for other language models, for example Russian and Chinese AI models, and most recently Silo, a Finnish-based company that aims to build AI language models with different European languages. It was still acknowledged that even though this is a needed development, they are hardly a diverse representation of other than Western languages.

Unemployment of the people unable or unwilling to adjust to the use of AI in the workplace, was also brough up during the interviews. It was mentioned that the use of AI is not something an employee can necessarily choose. Since AI offers a major advance in competition, and our economic system often prefers efficiency and reduced use of resources over employee's wants or needs, utilising AI is the best thing a company can choose to stay in the market as profitable. The reality for many working in the design and creative fields is that they must adapt to using AI in their work to avoid unemployment. Sometimes adapting to the impact of AI might not be possible, Andrea Filippi going as far as exclaiming stock image websites to be dead – who needs to purchase background images or mock-up images for advertising if you can generate images with AI customized to you and your company's needs.

What is the impact of AI on the appreciation of good art and design. Could we get entirely new types of cultural experiences, new types of AI driven or facilitated interactive design? Most interviewees agreed that consumers have an important role in guiding the direction of the development of AI. Consumers must be educated on what is good or revolutionary design and what kind of design do our buying decisions support.

6.2 AI as a creative tool

From a creative point of view, it was described that AI could be used for a variety of creative tasks from ideation to creating actual design content.

The interviewees mentioned creating mood boards and stock images, advertisement, journalism, realising projects, 3D models, communicating with a team or a client and ideation as specific creative tasks that can be done with the assistance of AI. All the interviewees mentioned that AI can be used as source of inspiration or ideation in creative work. For example, two of the interviewees, Andrea Filippi and Gabriele Moschin, wrote an illustrated poetry book, in which the artwork is entirely created with AI. Some of the interviewees described AI as an augment or extension of human imagination rather than substitution. It was mentioned that AI could be used co-creatively as a member of a creative team and one of the interviewees described the process of working creatively with AI as merging of human and AI.

This development where AI is used in ideation to create actual design pieces can be seen for example, from AI Fashion Week that was held in the spring of 2023 *(Fashionweek.ai, n.d.)* where contestants could submit their own AI created designs and the winner creations would get turned into real garments *(Picture 21).*

Picture 21. Design from the AI Fashion Week 2023 (Prego, 2023)

For game developers, AI makes building game worlds significantly quicker than before. It also can be used as a tool to develop the games more interactive and creative than before or going as far as incorporating artificial intelligence systems as central experiences in the game. Christian Guckelsberger describes the possible implementation of AI in game development as follows:

“You could employ AI in games to drive or to guide the behaviour of our other characters that you see in game worlds.”

One example of this kind of game development where AI has a central role is a game called Inworld Origins which was developed as a case study to explore the possibilities AI brings to games in the form of non-playable-characters or NPC’s. The goal of the game is to solve a case of explosion that happened by interviewing the characters using a microphone to talk to the characters. The characters, some of which are seen in *picture 22,* then answer the player’s questions in real time and in a non-scripted fashion. (*Steam, n.d.*)

Picture 22 A picture of the characters in Inworld Origins (*Steam, n.d.*)

6.3 Value of AI

Most of the interviewees recognized that the value artificial intelligence can bring is subjective and measuring it depends heavily on the point of view adopted when the issue is discussed.

It was mentioned that AI can bring value to working life with applications in medicine, traffic planning and other areas where a machine can more accurately and rapidly consume data, process it, and make logical decisions and predictions based on that data. It was generally understood that a human cannot compete with a computer in decision-making or prediction when it comes to speed and accuracy.

However, a unanimous opinion was that artificial intelligence does make an immense difference in creative fields. It makes the generation of illustrations and mock-ups quicker and more efficient in advertising, allows creating of 3D models with reduced time and makes photo editing faster. Most of the benefits of using AI in creative fields relate to the fact that AI makes working within them faster. Whether we consider speed as valuable to creative work depends on the point of view: in a capitalist society where all our efforts are concentrated on efficiency and maximum profitability of businesses, AI's value can't be underestimated.

It also came up that the people who have spent countless of hours to first train and then work in the creative field might not agree with that. Guckelsberger said in the interview that many people working in the creative industry, whether as a fashion designer, creative writer or as a game developer, gravitated towards their respective fields for a plethora of reasons, achieving results in a fast and efficient not always being one of the reasons. Moreover, some of the interviewees saw that some people value the art and craft included in creative work over the finished product, whether it's a video game or a pair of sneakers. However, it was also mentioned that this might not be the case forever: could it be possible for people working in the creative field find value in AI assisted design processes instead? Guckelsberger suggested that good management in the workplace has a major impact on whether people who don't inherently value efficiency that comes with the assistance of AI can be encouraged to find some other aspect of AI integrated in their work meaningful.

Other benefits that arose from the discussion were the inspiration that can be drawn from and with the assistance of AI, realising projects that were not possible before, the ability to spend more time in ideation and AI assisting people in creative work who don't have traditional education in creative fields or design.

Some of the interviewees thought that the benefits AI brings us in the future is the possibility of letting AI handle all the mundane tasks humans typically reject and leaving humans to exercise their creativity and free themselves to more human connections, art, poetry and cultivating vegetables.

6.4 Additional interview with Nicoline Van Enter

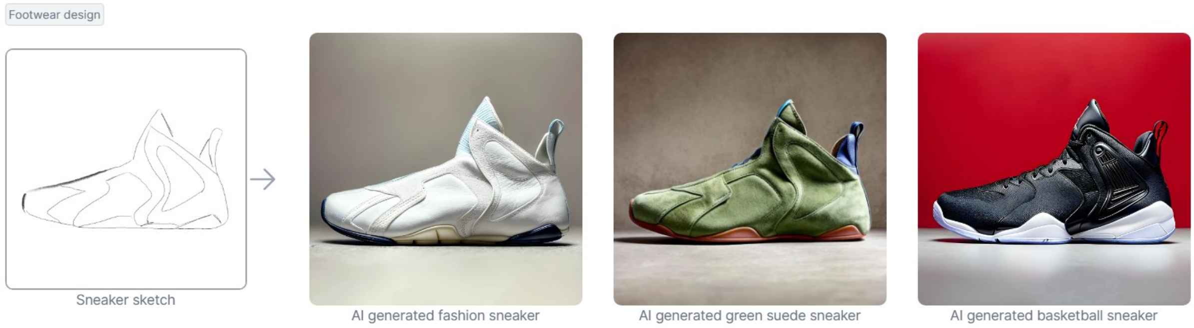
This interview was held after the four initial interviews to gain a more footwear-focused perspective for the conclusions of this thesis. In this interview Nicoline Van Enter, the co-founder and CTO of Footwearology, brings up experience-based information regarding generative AI and implementation of AI, value of AI and impacts of AI in the footwear design process.

6.4.1 Implementation of AI in the design process

When asked how AI can be implemented in the footwear design process, Van Enter mentioned that with the help of generative AI, visualisation in the design process is easier and faster. She says that since design came from artistry and one of the most important tasks of a designer is visualising, whether with pen and paper, software like Photoshop and Illustrator or 3D software like Blender 3D or Rhino, the next logical step is to use generative AI for visualisation in the design process. Generative AI is very good in visualisation, even when the user doesn't have the skills traditionally associated with designers. Experienced designers that she meets in her courses and workshops also echo this idea: the things that would take them days to do previously might only take half an hour with the assistance of generative AI.

One example of how AI could be used in visualisation of ideas, is with a generative AI tool called NewArc.ai, which is a software where the user can turn rough sketches into realistic images, and it can be used for fashion, footwear, automotive, industrial, architecture as well as interior and furniture design. Below is a visualisation from NewArch.ai website of how the tool can be used for footwear design process.

Picture 23. New.Arch.ai footwear sketch to AI generated images



Another application that van Enter mentioned was the preservation of a design style. Just like the work done in this thesis for the Emil Aaltonen Museum, companies can also fine-tune their own AI models to be able to replicate a design style and thus create a tool that is able to visualise designs consistent with the company's style fast. She mentioned a project she is working on that involves a senior designer, who has built the brand van Enter is working with, about to retire. AI models can be trained to reproduce his style so that his handwriting can be preserved in the future collections, even after his retirement.

Van Enter talks about something that could be done with AI to assist in the design process but which does not exist yet: a tool that can combine different data about the shoe such as material, sustainability, fit and customer needs. Knowing all these details is not possible for an individual designer but with the help of AI, all this data could be collected and combined for the benefit of the designer.

6.4.2 Value of AI in the design industry

When talking about the value AI brings to the design process, van Enter talked about the results of generative AI. She said that the biggest implication of AI in the design field will be in the ideation process: ideas are easy to explore by using different prompts and it diversifies the ideas a designer might have had before. She talked about the unexpectedness of the results of generative AI that are a result of the fact that generative AI is a machine that is trained but does not "see" or "think" like humans do: the model does not have any idea about the how things function in the physical world, nor does it understand the process of manufacturing shoes. This can lead for the model to present results that are not possible e.g., a shoe where eyelets are next to the shoelaces. She also mentions that this kind of unexpected visualisation of malfunctioning products might bring unexpected inspiration for the design process.

6.4.3 The impacts of AI on product development and production chains

Van Enter pointed out that AI in product development can be implemented from the beginning of the process, even before the designers start their work. Colour and trend departments can make trend stories based on the brief given by product management: these stories can be translated into AI models that are then made accessible to people (designers) who then work with those stories and utilise the models in their design process.

She also talked about the more mundane aspects of a designer's job. She pointed out that instead of working on new products and collections, designers spend most of their time working on carryover products. Carryover products are the permanent items that a company carries in their stores between collections. Training an AI to create variations of the same design with minor differences makes the design process faster and leaves the designer with more time to concentrate on new products.

It was also said that it is currently possible to achieve a very primitive 3D model of a shoe with generative AI. The models generating 3D models will continue to be developed with the development of chips and computers but for now there is still a need within companies to manually create accurate 3D models for the purposes of manufacturing. Additionally, AI allows companies to save money and labour in marketing. Visualisation is faster with AI and there is no need to wait for the finished product to photograph and market it.

In conclusion, generative AI is already being utilized by people working in different areas of footwear design. Van Enter predicted that for now, designers can avoid unemployment caused by AI, or even gain additional benefit in the job market, by learning new skills in the field of AI, but in the future, we will most likely see jobs being made redundant

7 Conclusion

The process of customizing an AI image generator For Emil Aaltonen Museum shows one example of how generative AI can be used as a design tool. As repeatedly mentioned in this thesis, the power of generative AI lies in efficient creation of ideas, and likewise, the LoRA model created in this thesis, produces new ideas by taking inspiration from past collections of Emil Aaltonen in an accurate and efficient way. It could be used as a co-creative team member in creating infinite variations of Emil Aaltonens footwear designs.

Even more importantly, customising an AI model in the style of a specific designer allows continuation of the cultural heritage that their work has offered for the footwear industry. Without the assistance of the LoRA model, it might be impossible to mimic the designer's style with precision. This way AI can be seen to bring cultural value for footwear industry. In other instances, this could bring also commercial and economical value, since customizing generative AI could be used to designing and marketing new collections based on previous ones.

This fine-tuning process gives an answer to the main research question set for this thesis: What is the process of making a customized AI model for Emil Aaltonen like. It also shows an example of how AI can be utilized as a tool in the design process, since fine-tuning can be applied to any designer’s style for ideation and creating collections based on previous designs if enough training data available in digital form.

7.1.1 Sustainability of creating the customized AI generator

This thesis reveals that utilizing generative AI in design in a sustainable way requires knowledge on many different aspects of environmental and social issues that come with using generative AI. Training new AI models always takes a lot of electricity and natural resources, and the sources can be hard to locate. The environmental impacts can also depend on the amount of time that is spent on training the AI, so in this thesis Google Colab was used to get access to better and faster GPU and to decrease the amount of time used in the process as well as avoiding the environmental impact of purchasing a new device to train the model.

There is no denying the fact that there is not a reliable source of the location, conditions, and environmental impacts of the Google Colab computing centres and the amount of electricity they use. However, using this cloud service made the training process significantly faster compared to training with the authors' low-GPU computers. For example, generating one regularization image in Google Colab notebook version of Stable Diffusion took approximately three seconds, whereas the generation of one image in Stable Diffusion run locally from the authors' computers took about 6-20 minutes. This means that it wouldn't necessarily been more sustainable, and most importantly not even possible, to fine-tune the image generator locally from the authors' computers.

In this project the pre-trained model that was used as a base for fine-tuning, was an open-source AI model, Stable Diffusion, meaning that we had no control of the original training data and its possible biases. Only thing we had control over was the datasets used in the fine-tuning process. Therefore, it was decided not to use binary terms when captioning the footwear images in order to avoid gender bias in the end product. This means that when captioning the training data, the footwear images were not divided into men's and women's shoes but instead were referred to simply as the type of shoe, e.g. “high heels” or “dress shoes”. Stable Diffusion is trained in English, so the language of the final model couldn't be changed to any other language.

The interviewees mentioned that the typical use of generative AI is iterative, and that was also the case in this thesis: a lot of low-quality images created for the project were discarded and this increased the environmental impact training the model had. Although these low-quality images were useless for fine-tuning, creative value can be found from them. They might not be completely accurate visualisations of footwear, but they can be used for ideation and they promote imagination, for example by trying to imagine how they could be turned into functional footwear. This serves as an example on how generative AI could break design bias as these kinds of low-quality or random images, such as seen in *pictures 24 & 25*, forces the designer to think outside of what is normally perceived as good design.

Picture 24. Rejected regularization image generated with Stable Diffusion

Picture 25. Rejected regularization image generated with Stable Diffusion

8 Reflection

The thesis can be considered successful, since all the research questions were answered and the main objective was completed: the commissioned LoRA model turned out fully-functioning and representing the style that the client had wished for.

It was an accomplishment to fine-tune an AI model, considering the level of skills on coding and technical aspects of generative AI the authors had before doing this thesis work. Even though the topic was familiar on a general level from former studies, this kind of work with AI requires a quite deep understanding of how artificial intelligence is constructed. Due to this, building the final LoRA model took many trials, but multiple issues were overcome and in the end a model that the client was approving of was developed.

The interviews were purposefully semi-structured, but this resulted in the interviews being too broad in some ways. On the other hand, it was interesting and insightful to hear professionals talk about AI in the context they best know it, but the interview questions were not always answered in the interviews. All five interviewees answered at least some of the questions very thoroughly so we were able to make sufficient conclusions about AI and its role in the context of design by combining the data from all the interviews. Overall, the discussions in these interviews were broad and the material acquired from them was sufficient in exploring the role of generative AI in design and what threats it poses to environment and the well-being of people.

This thesis offers a sustainability perspective on generative AI. Petra Jääskeläinen, one of out interviewees discussed the environmental and social sustainability implications of AI in depth and we briefly analysed the sustainability of our own model. Hopefully this thesis gives people interested in training their own models or using AI in their design work some insights on what to take into consideration when working with AI: this thesis should make people think more about the implications of generating content with generative AI and let it guide how they use it. Is it possible to limit the number of images generated for a project or if continuous usage is necessary, it should be considered to fine-tune a personalised AI model for the work since small fine-tuned models are practical to use and offer a more sustainable alternative to using large models.

The scheduled completion of this thesis work was delayed almost 5 months. The reasons for the delay are personal and professional, but nonetheless the collaboration between the authors was successful. The thesis was completed almost entirely remotely, both authors working on the text independently. The tasks done for the practical work e.g. creating regularization images, prototyping the fine-tuning, and captioning the regularization images were divided between the authors. Both authors were present at four out of the five interviews and both authors took part in transcribing, analysing, and summarising the interview data. As this thesis was high on technical information, the writing of background information was divided between the authors equally, and a meeting was held approximately once a week for the duration of this work to discuss the project, ask advice from one another and assign tasks to one another based on the availability of each author. In the end, the contribution from each authors for this work was surprisingly equal. This might be due to the shared goals of the authors and the fact that each authors strengths were taken into account while assigning tasks.

Sources

Andrew (28.10.2022) *Control your AI images with these must-know parameters.* Medium. https://medium.com/@andrewwongai/control-your-ai-images-with-these-must-know-parameters-e5cfbfba192f

Antalpha.ai (07.04.2023) Lora, DreamBooth, Textual Inversion in Stable Diffusion 1.5. Medium. Lora, DreamBooth, Textual Inversion in Stable Diffusion 1.5 | by Antalpha.ai | Medium

Bagu, B. (13.05.2023) Sampling steps in Stable Diffusion: finding a balance with examples. Medium. https://medium.com/@braden.bagu/optimizing-sampling-steps-in-stable-diffusion-a-practical-guide-with-examples-5716dc036fa1

Barretto, S. (23.11.2023) *How to use HiRes.fix to upscale your Stable Diffusion images in A1111.* Medium https://medium.com/rendernet/using-hires-fix-to-upscale-your-stable-diffusion-images-8d8e2826593e

Burns, E. (n.d.). *What is artificial intelligence (AI)?* TechTarget. https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence

Civitai (12.04.2024) *Realistic Vision V6.0 B1* Civitai. Realistic Vision V6.0 B1 - V5.1 (VAE) | Stable Diffusion Checkpoint | Civitai

Falbel, D. (22.07.2023) *Understanding LoRA with a minimal example*. AI Blog. Posit AI Blog: Understanding LoRA with a minimal example (rstudio.com)

Dzieza, J, (20.6.2023) *AI Is a Lot of Work*. The Verge.

https://www.theverge.com/features/23764584/ai-artificial-intelligence-data-notation-labor-scale-surge-remotasks-openai-chatbots?_hsenc=p2ANqtz-8_61uz3NEGB7cgnbfxnJLBikpmuc-2IC68FtByin77C-1cr6uYknPMSMeKYigNKPx7tBgR

Dartmouth. (n.d.) *Artificial Intelligence Coined at Dartmouth.* Dartmouth. https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth

Google Colab, n.d. *Google Colaboratory*. Google Colab https://colab.google/

Hardesty, L (14.4.2017) *Explained: Neural networks.* MIT News. https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414

Hashemi-Pour, C., (n.d.) *artificial general intelligence (AGI).* TechTarget. https://www.techtarget.com/searchenterpriseai/definition/artificial-general-intelligence-AGI

Heikkilä, M (1.12.2023) *Making an image with generative AI uses as much energy as charging your phone.* MIT Technology Review. https://www.technologyreview.com/2023/12/01/1084189/making-an-image-with-generative-ai-uses-as-much-energy-as-charging-your-phone/

Hu, E., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. & Chen, W. (16.10.2021) *LoRA: Low-Rank Adaptation of Large Language Models.* Arxiv 2106.09685.pdf (arxiv.org)

Hypotenuse AI. (11.08.2022) *AI Image Generators: How They Work and Why They Are Important.* Hypotenuse. https://www.hypotenuse.ai/blog/ai-image-generator

IBM. (n.d.). *What is machine learning (ML)?* IMB https://www.ibm.com/topics/machine-learning

Kohya-ss: A Friendly Stable Diffusion Fine-tuning Library (2023)

Kohya-SS: A Friendly Stable Diffusion Fine-tuning Library — Kohya-SS 0.1 documentation (sd-scripts.readthedocs.io)

Laskowski, N (n.d.) *What is artificial intelligence (AI)? Everything you need to know.* TechTarget. https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence

McKinsey & Company. (19.01.2023) *What is generative AI?* McKinsey & Company. https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai

Mishna, O. (n.d.) *Stable Diffusion Explained*. Medium https://medium.com/@onkarmishra/stable-diffusion-explained-1f101284484d

O'Connor, R. (12.05.2022) *Introduction to Diffusion Models for Machine Learning.* Assembly.ai https://www.assemblyai.com/blog/diffusion-models-for-machine-learning-introduction/

OpenAI. (n.d.) *Fine-tuning*. OpenAI. https://platform.openai.com/docs/guides/fine-tuning

Omniinfer. (26.07.2023) *Stable Diffusion API Playbook: Exploring Various Parameters and Comparing Image Generation Results.* Medium. https://medium.com/@admin_13769/stable-diffusion-api-playbook-exploring-various-parameters-and-comparing-image-generation-results-10ed7e723703

Penty, J. (2019) *Product Design and Sustainability: Strategies, Tools and Practises.* Routledge.

Petersson, D (14.11.2023) *AI vs. machine learning vs. deep learning: Key differences.* TechTarget. https://www.techtarget.com/searchenterpriseai/tip/AI-vs-machine-learning-vs-deep-learning-Key-differences

Prego, E. (n.d.) 41 Returns to Sun by Esther Souto Prego. Ai Fashion week.

https://fashionweek.ai/aifw23-top10/41-returns-to-sun-by-esther-souto-prego/

Pyynikinlinna. (n.d.). *Emil Aaltosen Museo, Kokoelmat.* Pyynikinlinna. http://www.pyynikinlinna.fi/kokoelmat/

Pyynikinlinna. (n.d.). *Emil Aaltosen Museo, Emil Aaltonen*. Pyynikinlinna. http://www.pyynikinlinna.fi/emil-aaltonen/

Rackspace Technology. (14.06.2023) *Understanding the Distinctions Between Artificial Intelligence, Machine Learning and Generative AI*. Rackspace Technology. https://www.rackspace.com/blog/distinctions-ai-ml-generative-ai

Replicate. (n.d.) *What is fine-tuning?* Replicate. https://replicate.com/docs/fine-tuning

Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M. & Aberman, K. (15.03.2023) DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation [Boston University, Google Research]. https://arxiv.org/pdf/2208.12242.pdf

SaturnCloud. (6.7.2023) *How to Improve Accuracy in Neural Networks with Keras.* SaturnCloud. https://saturncloud.io/blog/how-to-improve-accuracy-in-neural-networks-with-keras/#3

Srivastava, P. (22.07.2023) *Fine-Tuning AI Models: A Guide*. Medium. Fine-Tuning AI Models: A Guide. Fine-tuning is a technique for adapting… | by Prabhu Srivastava | Medium

Stability AI, (n.d.) *stable-diffusion-xl-refiner-1.0*. Hugging Face. https://huggingface.co/stabilityai/stable-diffusion-xl-refiner-1.0

Stable Diffusion Art (22.11.2023) *What are LoRA models and how to use them in AUTOMATIC1111*. Stable Diffusion Art. What are LoRA models and how to use them in AUTOMATIC1111 - Stable Diffusion Art (stable-diffusion-art.com)

Steam. *Inworld Origins* (n.d.) Steam. https://store.steampowered.com/app/2199920/Inworld_Origins/

Strubell, E., Ganesh, A. & McCallum, A. (2019) Energy and Policy Considerations for Deep Learning in NLP. College of Information and Computer Sciences University of Massachusetts Amherst. https://arxiv.org/pdf/1906.02243.pdf

Tarud, J. (23.08.2023) *AI Models: How Does It Work?* Koombea https://www.koombea.com/blog/ai-models/

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence, 10*(2), 1-37. https://sciendo.com/downloadpdf/journals/jagi/10/2/article-p1.pdf

Yasar, K (n.d.) *Neural network.* TechTarget. https://www.techtarget.com/searchenterpriseai/definition/neural-network

Appendix 1: THESIS DATA MANAGEMENT PLAN

Data collected for this thesis includes recorded interviews and image data from Emil Aaltonen Museum. The data is processed following Data managing guidelines of HAMK.

1 Management and storage of research data

Data will be stored in personal hard drive provided by HAMK, and a backup will be saved on an external hard drive. It can be accessed only by Iris Kinni, Maija Lumiaho, Ville Siipola and Helena Leppänen. The sensitive data collected from the interviews includes the names and occupations of the interviewees.

2 Processing of personal data and sensitive data

The interviews are done with professionals in the field of AI and will be semi structured theme interviews. The only personal data included in the thesis are the name and occupation of the interviewees. Interviewees will be asked if they wish to have their name and occupation excluded from the final thesis.

3 Ownership of thesis data

This thesis is owned by Iris Kinni and Maija Lumiaho. The artificial intelligence image generator that is the result of this thesis will be owned by Iris Kinni, Maija Lumiaho and Emil Aaltonen Museum. The image data from Emil Aaltonen Museum belongs to the museum and is permitted to be used only in the context of this thesis. The commissioner of the thesis, Emil Aaltonen Museum, will not have access to any interview data gathered for the purposes of this thesis.

4 Further uses of data

No personal or sensitive data will be used after the thesis is completed. The interviews are stored in personal hard drive provided by HAMK and an external hard drive and will be deleted a year after the thesis is accepted and published.

https://www.hamk.fi/wp-content/uploads/2022/02/Thesis-data-management-plan.pdf

https://hameenamk.sharepoint.com/:w:/s/opiskelijan-materiaalit/EWrtFp5uhWZCijrvRmqtmoQYbvN89nUOIAfOWF3CV6BHQ?e=41XJNv