



Satakunnan ammattikorkeakoulu
Satakunta University of Applied Sciences

CAROLINA KIVIMÄKI

Kuvien tuottaminen tekoälyllä: Stable Diffusion, AUTOMATIC1111 ja DreamBooth

Käyttöliittymän asennus, personoitu malli,
promptaus ja vastuullinen kuvien generointi

TIETOJENKÄSITTELYN TUTKINTO-OHJELMA
2023

TIIVISTELMÄ

Kivimäki, Carolina: Kuvien tuottaminen tekoälyllä: Stable Diffusion, AUTOMATIC1111 ja DreamBooth: käyttöliittymän asennus, personoitu malli, promptaus ja vastuullinen kuvien generointi

Opinnäytetyö, AMK

Tietojenkäsittelyn tutkinto-ohjelma

Marraskuu 2023

Sivumäärä: 50

Opinnäytetyön tarkoituksena oli koostaa yhteen paikkaan kaikki tarvittava perustieto, jonka avulla voi aloittaa kuvien tuottaminen tekoälykuvageneraattorilla. Työssä oli tarkoitus vielä huomauttaa kuvien tuottajan omasta vastuusta, tämänhetkisistä tekijänoikeuksista ja tekoälykuvageneroinnin takana piilevästä eettisyysongelmasta.

Opinnäytetyön tavoitteena oli yleisellä tasolla perehtyä tekoälykuvageneraattoreiden kuvan generointiprosessiin, asentaa paikallisesti AUTOMATIC1111:n Stable Diffusion käyttöliittymä ja tutustua sen käyttöön. Työssä lisäksi opetettiin Google Colabissa Dreambooth -mallilla omilla valokuvilla personoitu malli. Tätä mallia pystyy käyttämään käyttöliittymässä personoitujen kuvien tuottamiseen. Lopussa käytiin läpi yleisimmät promptaus-tekniikat, käsiteltiin tekijänoikeuksien ja eettisten kysymysten vaikutusta tekoälyllä tuotettuihin kuviin ja niiden omistajuuteen.

Työn aikana selvisi, että tekijänoikeuslain mukaan generoidun kuvan omistajuus kuuluu kuvan tekijälle. Tämä pätee niin kauan, kun kuva ei muistuta täydellisesti kenenkään toisen teosta. EU:lla on kehitteillä laki, jossa vaaditaan generatiivisten tekoälyjärjestelmien noudattavan avoimuusvaatimuksia ja taattava turvatoimia laittoman sisällön tuottamista vastaan. Promptaus on taito, joka yksinkertaisesti vaatii käyttäjältä pientä harjoittelua. Vaihtoehtona olisi löytää sen verran tehokas malli, johon riittäisi kauniin tuloksen saavuttamiseksi vain yksinkertainen kehote.

Avainsanat: tekoäly, Stable Diffusion, käyttöliittymät, valokuvat

Abstract

Kivimäki, Carolina: Creating images with AI: Stable Diffusion, AUTOMATIC1111 and DreamBooth: UI installation, personalized model, prompting and responsible image generation

Bachelor's thesis

Degree programme in Business Information Systems

November 2023

Number of pages: 50

The purpose of the thesis was to gather in one place all the basic information needed to start producing images with an AI image generator. It also was intended to point out the responsibility of the creator of the images, the current copyright issues, and the ethical problem behind AI image generation.

The thesis aimed to get a general understanding of the image generation process of AI image generators, to install locally the Stable Diffusion interface of AUTOMATIC1111, and to get familiar with its use. In addition, it was taught in the thesis how to personalize a model with one's own photos using the Dreambooth model in Google Colab. This model can be used in the interface to generate personalized images. Finally, the most common prompting techniques were reviewed, and the impact of copyright and ethical issues on AI-generated images and their ownership were addressed.

During the work, it became clear that according to the copyright law, the ownership of a generated image belongs to the creator of the image. This is the case as long as the image does not perfectly resemble someone else's work. The EU is developing a law that requires generative AI systems to comply with transparency requirements and guarantee safeguards against the production of illegal content. Prompting is a skill that requires a little practice on the part of the user. The alternative would be to find a model that is efficient enough that a simple prompt would be enough to achieve a beautiful result.

Keywords: artificial intelligence, Stable Diffusion, user interfaces, photographs

SISÄLLYS

1 JOHDANTO	6
1.1 Tekoälyn käyttö tässä opinnäytetyössä	7
2 TUTKIMUSONGELMA JA SEN PERUSTELU	7
3 TEKÖÄLYN KÄSITE	8
3.1 Tekoäly ja sen toimintatavat	8
3.2 Heikko ja vahva tekoäly	9
4 TEKSTISTÄ KUVAKSI PROSESSI	9
4.1 Kehote	10
4.2 Latentti avaruus	10
4.3 Diffuusiomalli	11
4.4 Diffuusio prosessin ehdollistaminen	12
4.5 Stable Diffusionin toiminta vaihe vaiheelta	13
5 STABLE DIFFUSION JA AUTOMATIC1111	14
6 AUTOMATIC1111:N STABLE DIFFUSION WEBUI:N PAIKALLINEN ASENTAMINEN	15
6.1 Vähimmäisvaatimukset Windows -käyttöjärjestelmissä	16
6.2 Git ja Python	16
6.3 Stable Diffusion WebUI:n asennusprosessi	17
6.4 Stable Diffusionin päivittäminen	22
6.5 Virheidenkorjaus	22
7 OMAN TEKÖÄLY MALLIN OPETTAMINEN DREAMBOOTHILLA	24
7.1 Google Colab	24
7.2 DreamBooth	24
7.3 Ohjeet omien kuvien ottoon ja käyttöön	25
7.4 Install Requirements ja Login to Hugging Face	25
7.5 Settings and run	26
7.6 Start Training	27
7.7 Valmiit T0FF1 ja 5U5KU mallit	30
7.8 Personalisoidun mallin käyttöönotto	31
7.9 Hugging Face ja CivitAI	33
8 AUTOMATIC1111 KÄYTTÖLIITTYMÄN ESITTELY	34
9 YLEISET PROMPTAUS TEKNIIKAT	37
9.1 Kehotteen rakenne	37
9.2 Sanajärjestys	38
9.3 Negatiivinen kehote	39

9.4 () ja [] syntaksit	39
10 TEKIJÄNOIKEUDET JA EETTISYYS	40
10.1 Rikotaanko tekijänoikeuksia?	40
10.2 Eettisyys	41
10.3 Käyttäjän oma vastuu	42
11 LOPUKSI	43
LÄHTEET	46

1 JOHDANTO

Viime aikoina generatiivisten tekoälyjärjestelmien suosio on ollut kasvussa. Alkuvuodesta verkossa on voinut kohdata uutisiin, joissa tekoälytaiteen eettisyyttä on kritisoitu tai videoihin, joissa tekoälyllä on luotu kappaleiden sanoista kuvia. Nyt on entisestään kehittynyt mahdollisuus saada lyhyelläkin äänikappaleella tekoäly matkimaan ihmisen ääntä ja toistamaan puhetta tekstistä. Nykyään tekstistä voidaan myös luoda jopa lyhyitä videoita.

Generatiivisten tekoälyohjelmien ajankohtaisuudesta sekä henkilökohtaisen mielenkiinnon takia lähdin työstämään aiheesta opinnäytetyötä. Halunani oli oppia käyttämään tekoälykuvageneraattoreita ilman lisäkustannuksia ja ymmärtää tekoälykuvageneraattoreiden yleistä toimintaa.

Tekoälykuvageneraattoreita löytyy nykyään lukuisia, ja monet niistä ovat tavalla tai toisella maksullisia. Promptaus-tekniikat voivat myös vaihdella riippuen käyttöliittymästä tai kuvageneraattorista. Tässä opinnäytetyössä keskitytään siis maksuttomaan kuvien tuottamiseen avoimen lähdekoodin Stable Diffusion kuvageneraattorilla. Käyttöliittymänä käytetään aloittelijoille ystävällistä AUTOMATIC1111:n kehittämää Stable Diffusion WebUI:ta. Näiden lisäksi Google Colabissa opetetaan omilla valokuvilla malli, jota voi käyttää personoitujen kuvien generointiin AUTOMATIC1111:n käyttöliittymässä.

Työ on suunnattu heille, jotka eivät ole vielä täysin perehtyneet aiheeseen, mutta ovat kiinnostuneita tutustumaan tekoälykuvagenerointiin ja oppimaan perusteista lähtien, miten kaikki toimii. Tämän lisäksi työssä huomioidaan kuvien tuottaminen mahdollisimman alhaisin kustannuksin, jolloin tämä työ sopii melkein kelle tahansa, joka on vähänkin kiinnostunut kokeilemaan kuvagenerointia.

Tämän työn tavoitteena on tutustua tekoälykuvageneraattorien toimintaan, AUTOMATIC1111 käyttöliittymän asentamiseen ja käyttöön, personoitujen kuvien generointiin, kehotetekniikoihin ja tekijänoikeuksiin sekä omavastuuseen kuvien tuotossa.

1.1 Tekoälyn käyttö tässä opinnäytetyössä

Olen tässä työssä käyttänyt ChatGPT:tä kielentarkistukseen, synonyymien löytämiseen ja ideointiin. ChatGPT:n avulla etsin sanoille tieteellisempiä termejä ja käänsin lauseiden ilmaisua tieteellisemmäksi. Kaikki ilmoittamani lähteet ovat minun käyttämiäni ja löytämiäni eivätkä tekoälyn.

Englanninkielisten artikkelien ja tekstien kääntämisessä olen käyttänyt apuna DeepL Translatoria. Englanninkielisen tiivistelmän kääntämiseen ja kieliasun tarkistamiseen olen myös hyödyntänyt DeepL Translatoria sekä Grammarlyä. Olen käyttänyt kaikkia tekoälysovelluksia vastuullisesti ja tietosuojasta huolehtien. Tässä alaluvussa 1.1. en ole käyttänyt tekoälyä.

2 TUTKIMUSONGELMA JA SEN PERUSTELU

Ongelma syntyi tarpeesta saada yhteen paikkaan tarvittava pohjatieto kuvagenerointia varten. Työn keskeinen kysymys on, kuinka kuvageneraatiota voitaisiin soveltaa paikallisesti ja maksuttomasti omia valokuvia käyttäen, ottaen huomioon myös eettiset ja tekijänoikeudelliset näkökulmat. Näiden lisäksi työssä vastataan seuraaviin kysymyksiin: ”Miten tekoälykuvageneraattorit toimivat?”, ”Miten Stable Diffusionin WebUI:ta tulisi käyttää?” ja ”Millaisia kehoitteita tulisi käyttää?”

Vastaamalla edellä mainittuihin kysymyksiin, työssä koostetaan yhteen paikkaan kaikki tarvittava perustieto, jonka pohjalta voi päästä alkuun kuvien tuottamisessa tekoälykuvageneraattorilla, varsinkin Stable Diffusionilla.

Stable Diffusionin valinta perustuu sen maksuttomuuteen ja avoimuuteen, mikä tekee siitä saavutettavan vaihtoehdon useammallekin käyttäjälle. Edellä mainittuihin kysymyksiin vastaaminen tarjoaa vastauksia ja ohjeita heille, jotka haluavat oppia tekoälykuvageneraatiosta.

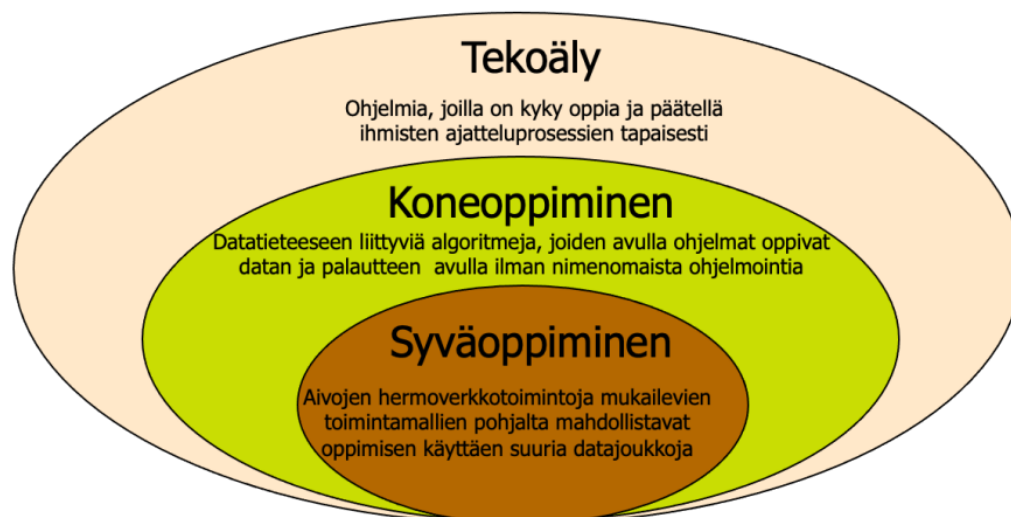
3 TEKÖÄLYN KÄSITE

3.1 Tekoäly ja sen toimintatavat

Tekoälyn eli AI:n (Artificial Intelligence) tarkoituksena on simuloida tai jäljitellä ihmisen älykkyyttä, kognitiivisia toimintoja ja ajatteluprosessia. Tekoäly-terminillä voidaan myös viitata mihin tahansa koneeseen, joka omaa ihmismielen kaltaisia taitoja, kuten kyvyn oppia ja ratkaista ongelmia. (Frankenfield, 2023.)

Tekoäly koostuu useammasta kuin yhdestä tekniikasta. Tähän lukeutuvat koneoppiminen ja syväoppiminen (kuva 1). Koneoppiminen on tekoälyn osa-alue, jossa kone kykenee itsenäisesti oppimaan sille syötetystä datasta. (Merilehto, 2018, s. 19, 27.)

Syväoppiminen puolestaan on koneoppimisen osa-alue (IBM, n.d.). Syväoppimisessa käytetään suuria neuroverkkoja, jotka koostuvat useista kerroksista. Neuroverkot ovat sarja algoritmeja, jotka imitoivat aivojen toimintaa tunnistamalla yhteyksiä suurien datamäärien välillä. Neuroverkkoja verrataan hermostoon, sillä neuroverkot muistuttavat aivoissa olevien hermosolujen ja synapsien yhteyksiä. (Chen, 2023.) Yleisimpiä sovelluksia, joissa hyödynnetään syväoppimista, ovat kuvan- ja puheentunnistuksen sovellukset (SAS, n.d.).



Kuva 1. Esimerkkikuva tekoälyn rakenteesta (Klementti, 2022)

3.2 Heikko ja vahva tekoöly

Heikko tekoöly viittaa tekoölyyn, joka on opetettu suorittamaan tietynlaisia tehtäviä. Useimmat tällä hetkellä käytössä olevista tekoölyistä toimivat juuri heikolla tekoölyllä. (IBM, n.d.) Heikkoa tekoölyä hyödyntävät esimerkiksi shakiohjelma ja virtuaaliavustajat, kuten Applen Siri (Frankenfield, 2022).

Verrattuna heikkoon tekoölyyn, vahva tekoöly on yhtä älykäs kuin ihminen. Vahva tekoöly on itsetietoinen, kykenevä oppimaan, ratkaisemaan ongelmia ja tekemään suunnitelmia tulevaisuuteen. Tänä päivänä vahvalle tekoölylle ei ole vielä olemassa käytännön esimerkkiä. (IBM, n.d.)

4 TEKSTISTÄ KUVAKSI PROSESSI

Tässä luvussa kerrotaan yleisellä tasolla, kuinka tekoölytaidegeneraattorit, kuten Stable Diffusion, taustaltaan toimivat. Tällä on tarkoitus antaa lukijalle yleiskuva generaattorien sisäisestä toiminnasta ja niiden termeistä.

4.1 Kehote

Tekoälygeneraattorit generoivat kuvia luonnollisen kielen kuvauksista eli kehotteista (prompt). Kehote, toisin sanoen tekstisyöte, on tekstiä, johon sisältyy käyttäjän kuvailema tapahtuma, väri, teema, vuosiluku tai mikä tahansa syöte, esimerkiksi ”a giraffe ice skating on ice in the 1960s”.

Jotta tekoälytaidegeneraattorit kykenevät tuottamaan hyvää tulosta ja vastaamaan käyttäjän kehotteisiin, ne tarvitsevat oppimiseen suuren ja monipuolisen joukon dataa. Tällaista dataa ovat verkosta otetut erilaiset valokuvat kuvateksteineen. (Vox, 2022.)

4.2 Latentti avaruus

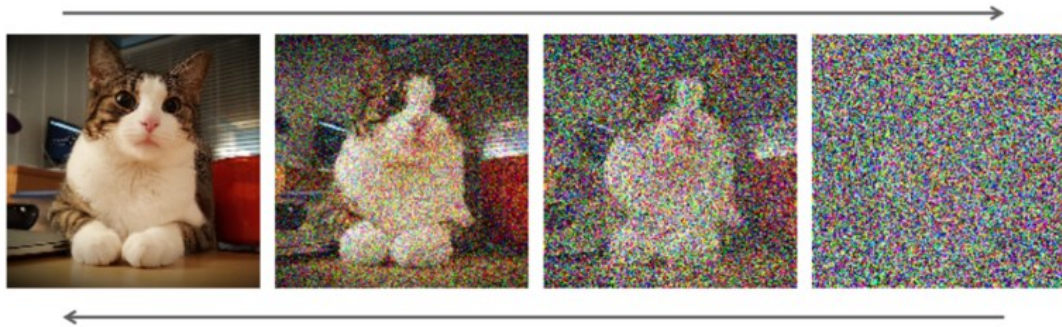
Kun generaattori tuottaa käyttäjälle pyydetyn kuvan, tulee kuva syväoppivan mallin latentti avaruudesta (latent space), joka voi sisältää jopa yli 500 ulottuvuutta. (Vox, 2022.) Latentti eli piilevä avaruus tarkoittaa yksikertaisesti avaruutta, jossa samankaltaiset datapisteet ovat lähempänä toisiaan (Tiu, 2020). Avaruudessa on ryppäitä (cluster) eli ryhmiä, jotka ovat ryhmitelty samankaltaisuuksiensa mukaan (Vox, 2022).

Latentin avaruuden ryhmittymät ovat tavallaan lajiteltu kuten eri aihepiirit käsitekarttoissa. Yksinkertaisena esimerkkinä sanoan ”mytologia” voidaan yhdistää sanoja, kuten ”kreikka”, ”jumala” ja ”jumalatar”. Edelleen jumalattariin saateen liittyy sana nainen, toisia jumalattaria ja näihin jumalattariin vielä liittyviä toisia henkilöitä tai piirteitä. Sanaan ”kreikka” saatamme yhdistää Kreikan lipun värit valkoinen ja sininen, oliivit ja niin edelleen.

Mikä tahansa kohta latentissa avaruudessa voidaan ajatella reseptinä mahdolliselle kuvalle. Käyttäjän antama kehote on se, joka toimii navigoijana. On tärkeää huomata, että eri generaattorit tuottavat samasta kehotteesta eri tuloksen, koska jokaisella on käytössä eri latentti avaruus. (Vox, 2022.)

4.3 Diffuusiomalli

Jotta kuva saadaan aikaiseksi, tarvitaan generatiivista eli uudelleen generoivaa prosessia nimeltä diffuusio.



Kuva 2. Kohinan lisäys- ja poistoprosessin esimerkki (Vahdat & Kreis, 2022)

Diffuusiomalleja opetetaan siten, että kuvaan lisätään pikseleitä, joita kutsutaan kohinaksi, tarkemmin kutsuttuna Gaussin kohinaksi. Gaussin kohina seuraa Gaussin jakaumaa eli normaalijakaumaa. Opetuskuvia korruptoidaan lisäämällä niihin kohinaa tietyn iteraatiomäärän verran ja tätä tehdään niin kauan, kunnes opetuskuva on suunnilleen puhdasta kohinaa (kuva 2). (AI Network, 2022; Logunova, 2022.)

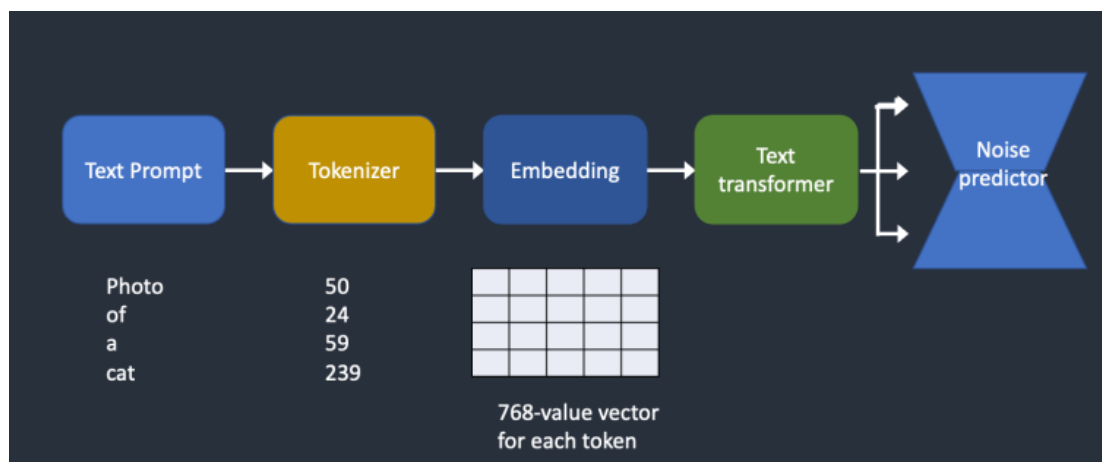
Ennustavan mallin, U-Netin, tehtävänä on ennustaa kuviin lisätyn kohinan määrää askel askeleelta (Spektor, 2022). Kun diffuusiomalli on opetettu, sitä käytetään poistamaan kohinaa asteittain ja uudelleen rakentamaan (reconstruct) kuva syntetisoimalla dataa. (Logunova, 2022.) Datan syntetisointi tarkoittaa, että datasta luodaan aivan uusi ja keinotekoinen, mutta kuitenkin alkuperäistä dataa muistuttava (Syntho, n.d.).

Opettamisen jälkeen, diffuusiomalli pystyy generoimaan kuvia satunnaisesta kohinasta soveltamalla oppimaansa kohinan poistotekniikkaa (Logunova, 2022).

4.4 Diffuusio prosessin ehdollistaminen

Tässä vaiheessa diffuusiomalli vasta luo satunnaisesta kohinasta vain satunnaisen kuvan. Jotta voimme luoda haluamme kuvan, tulee tässä vaiheessa esille conditioning eli ehdollistaminen. Prosessia ohjaa käyttäjän antama kehoite, joka on muunneltu upotusmuotoon (embedding) ja ehdollistamisen tarkoituksena on ohjata kohinan ennustajaa, U-Net mallia. (Andrew, 2023.)

Embedding, suomeksi upotus tai sisällytys, tarkoittaa matemaattista tapaa esittää tietoa (AssemblyAI, 2022). Kuva tai teksti muunnetaan sellaiseen muotoon, että algoritmi ymmärtää sen eli tässä tapauksessa vektoriksi, numeroiden kokoelmaksi. (Wayner, 2022.)



Kuva 3. Kehotteen vaiheet kohinan ennustajaan. (Andrew, 2023a)

Kokonaisuudessaan kuvassa 3 kehoite prosessoidaan ja syötetään kohinan ennustajalle eli U-Netille. Tokenisoijan (tokenizer) tehtävänä on ensin kääntää jokaisen kehotteen sana numeroksi, tokeniksi (token). Jokainen tokeni muunnetaan vektoriksi eli upotukseksi.

Kehote tokenisoidaan syväoppivassa mallissa nimeltä CLIP (Contrastive Language-Image Pre-training) (Andrew, 2023a). CLIPin tehtävänä on ennustaa ja yhdistää olemassa oleva kuva sitä vastaavaan tekstiin. Kuvaa generoidessa CLIPin päämääränä on varmistaa, että generaattorin tuottama kuva ja käyttäjän antama kehoite vastaavat mahdollisimman paljon toisiaan (AssemblyAI, 2022).

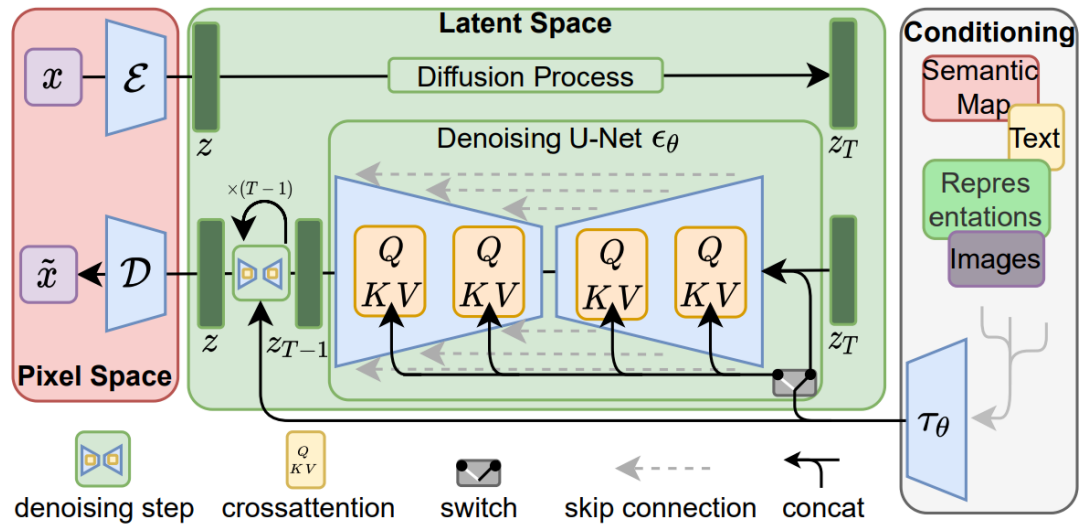
Jotta kuvan ja tekstin yhdistely pystytään tekemään, CLIP opettaa kahta enkooderia: Text Encoder ja Image Encoder (AI Network, 2022; AssemblyAI, 2022). Text Encoder muuntaa tekstit tekstiupotuksiksi ja Image Encoder muuntaa kuvat kuva upotuksiksi (AssemblyAI, 2022).

Kun tokenit ovat muutettu tekstiupotukseksi, prosessoidaan ne tekstimuuntajassa (text transformer) ja sen jälkeen upotus siirtyy kohinan ennustajalle, U-Netille, joka käsittelee tekstistä ja latentti kuvasta oikean kuvan. (Andrew, 2023.)

4.5 Stable Diffusionin toiminta vaihe vaiheelta

Verrattuna muihin kuvageneraattoreihin, Stable Diffusion on latentti diffuusiomalli. Latentti diffuusiomallissa kuvat kompressoidaan (compress) latentti avaruuteen eli erittäin pieneen kokoon. Tämä tekniikka mahdollistaa sen, että Stable Diffusion on nopea ja toimii käyttäjän omallakin laitteella. (Andrew, 2023a.)

Tekniikkaa, jota latentti diffuusiossa käytetään, kutsutaan variaationaaliseksi enkooderiksi (variational encoder). Se sisältää kaksi osaa: enkooderin (encoder) ja dekooderin (decoder). Latentti avaruudessa enkooderi kompressoii kuvan pienemmäksi ulottuuden esitykseksi ja dekooderi palauttaa kuvan latentista avaruudesta. (Andrew, 2023a.)



Kuva 4. Latentin diffuusiomallin rakenne. (Rombach ym., 2021)

Kuvassa 4 aloitetaan vasemmasta yläkulmasta, text encoderista, jossa käyttäjän kehoite muunnetaan vektoriksi. Latentti avaruudessa generoidaan satunnainen ja kohinallinen esitys tulevasta kuvasta. Latentti avaruudessa U-Net ottaa syötteenä kohinallisen kuvan esityksen ja käyttäjän kehoitteen ja ennustaa kohinan määrän. Ennustettu kohinan määrä vähennetään latentti kuvasta, ja tämä korvaa edellisen latentti kuvan. Kyseistä kohinan ennustuksen ja vähennyksen prosessia toistetaan määrätyn iteraatiomäärän verran. Lopuksi dekooderi palauttaa valmiin kuvan käyttäjälle. (Andrew, 2023; Underfitted, 2022.)

5 STABLE DIFFUSION JA AUTOMATIC1111

Elokuussa vuonna 2022 Stability AI julkaisi sivuillaan blogipostauksen, jossa kerrottiin Stable Diffusionin ensimmäisen version julkaisusta. Stable Diffusionin kehityksessä olivat mukana Stability AI:n lisäksi Runwaylta Patrick Esser ja Robin Rombach konenäön ja oppimisen tutkimusryhmästä LMU Munichista. (Stability AI, 2022.)

Viimeisimpänä versiona Stability AI:lta on ilmestynyt SDXL 1.0. SDXL:llä pystyy luomaan entistä elävämpiä ja väriikkaampia kuvia. Kontrastit, valaistus ja varjot ovat parempia kuin edeltävissä Stable Diffusionin versioissa. Kaikki

kuvat tuotetaan 1024x1024 resoluutiolla 512x512 sijaan. SDXL kykenee generoimaan konsepteja, jotka ovat perinteisesti olleet haastavia kuvamalleille, kuten kädet tai tekstit. (Stability AI, 2023.)

Tässä työssä käytetään Stable Diffusionia myös juuri siitä syystä, että sitä on mahdollista käyttää ilmaiseksi. Tietysti hintaa saattaa lisätä se, jos käyttäjällä ei ole ennestään tehokasta, edes keskivertoista, pelitietokonetta. Ilman riittävän tehokasta laitteistoa Stable Diffusionin käyttäminen paikallisesti saattaa olla haastavaa, erittäin hidasta ja jopa mahdotonta. Jos käyttäjällä ei ole aikomusta päivittää laitteistoaan, on siinä tapauksessa käyttäjälle helpompaa ja vaivattomampaa käyttää pilvimuotoisesti Stable Diffusionia. Verkosta voi löytää useampia ilmaisia tai lähes ilmaisia kuvageneraation työkaluja ja ohjelmia. Näistä mainitsemisen arvoisia ohjelmia ovat ainakin CivitAI ja Stable Diffusion Web.

AUTOMATIC1111:n Stable Diffusion on tunnettu verkkokäyttöliittymä, joka tunnetaan vain sen tekijän Github-käyttäjänimestä (Amir, 2023). AUTOMATIC1111:n käyttöliittymä toimii myös Apple Siliconilla ja Linuxilla (AUTOMATIC1111, n.d.-a).

Valitsin tähän työhön käytettäväksi AUTOMATIC1111:n käyttöliittymän, sillä se on tällä hetkellä aloittelijaystävällisin ja suosituin. Asennusprosessi on yksinkertainen, ja aloittelijan on helppo tutustua käyttöliittymän käyttöön. Netistä löytyy varmasti apua, jos asennuksen tai käyttöliittymän kanssa tulee ongelmia.

6 AUTOMATIC1111:N STABLE DIFFUSION WEBUI:N PAIKALLINEN ASENTAMINEN

Tässä luvussa opastetaan vaihe vaiheelta läpi, kuinka Stable Diffusion käyttöliittymä voidaan asentaa paikallisesti Windows -käyttöjärjestelmän omaavalle

laitteelle (Lewis, 2022). Omassa luvussaan mainitaan huomioita vaativat laitteen vähimmäisvaatimukset sekä mahdollisesti vastaan tulevat ongelmat. Lukijan on suositeltavaa huomioida, että tämä ohjeistus perustuu vuoden 2023 syksylle ja voi muuttua tulevien vuosien aikana.

6.1 Vähimmäisvaatimukset Windows -käyttöjärjestelmissä

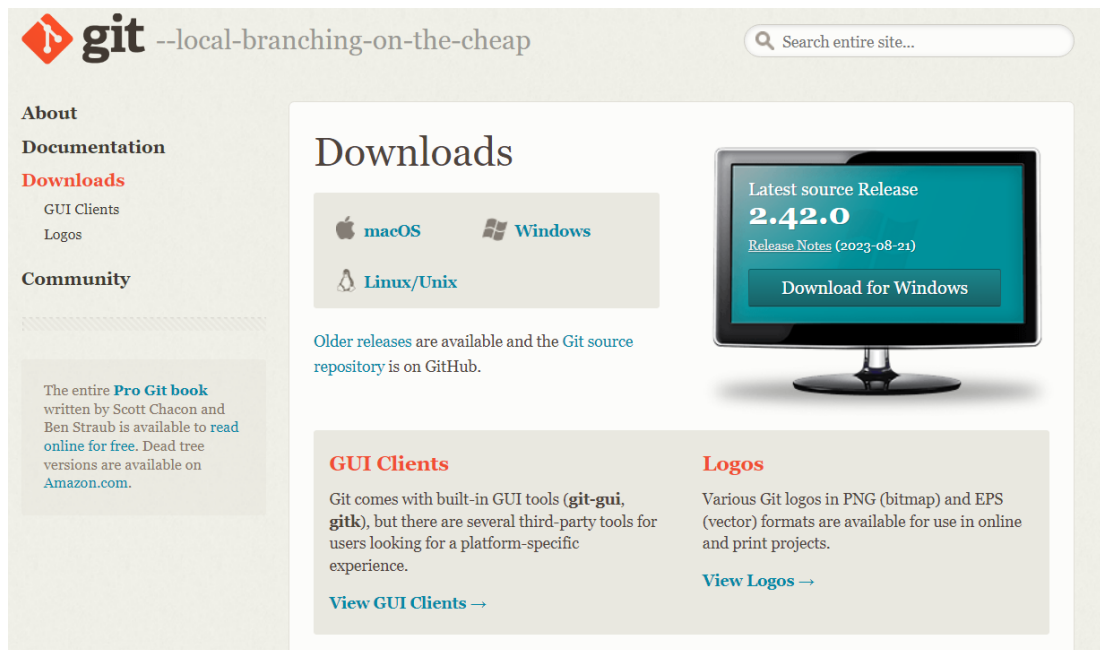
Vähimmäisvaatimuksista ei löydy yhtä ja oikeaa vastausta, mutta ainakin suuntaa antavia. Jotta Stable Diffusion toimisi moitteetta, vaaditaan laitteelta vähimmäisvaatimuksina ainakin seuraavina ominaisuuksina: 4 GB VRAMia sisältävä GPU, vähintään 10 GB levytilaa ja 16 GB RAMia (AUTOMATIC1111, n.d.-c; Andrew, 2023; Bourai, 2023; Ma, 2023).

Tätä työtä toteuttaessa on käytetty keskivertoista pelitietokonetta, jossa VRAM hipoo juuri kuutta gigatavua. Käytössä on NVIDIA 1660TI ja 16 GB RAMia. Vapaata levytilaa ennen Stable Diffusionin ja mallien lataamista oli 40 GB.

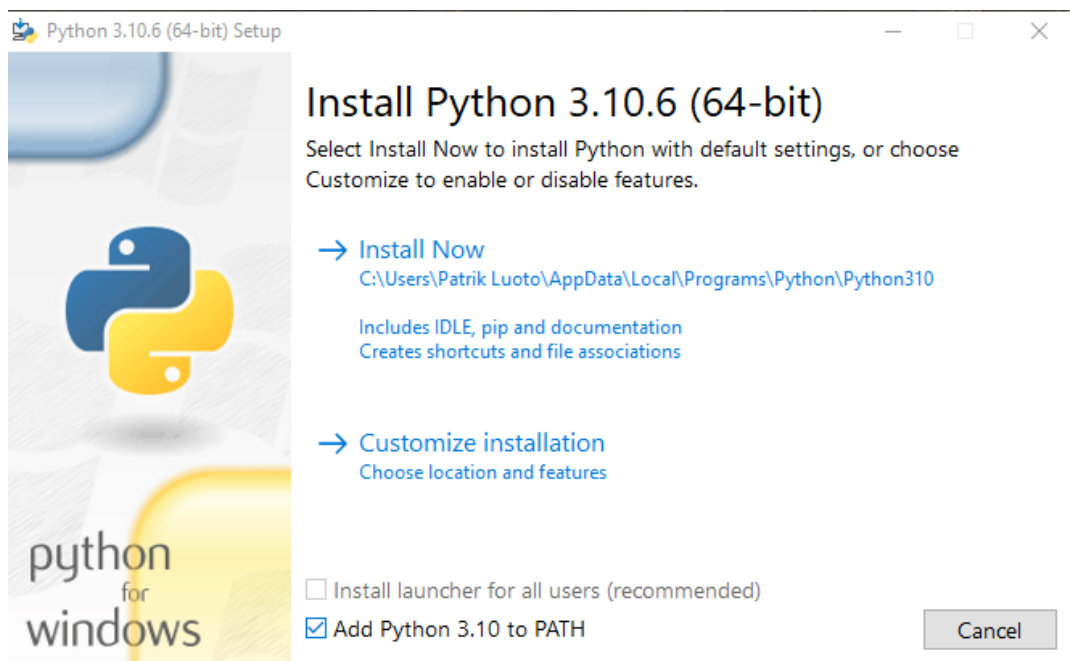
6.2 Git ja Python

Ensin laitteelle on ladattava Git ja Python ohjelmat. Git on ilmainen ja avoimen lähdekoodin versionhallintajärjestelmä (Git, n.d.). Gitin avulla voidaan ladata ja päivittää AUTOMATIC1111:n Stable Diffusion käyttöliittymä (Andrew, 2023). Työssä asennettiin kuvassa 5 näkyvä versio, mutta Gitin lopullisella versiolla ei pitäisi olla merkitystä, sillä sitä käytetään vain tiedoston poimimiseen. Gitin asennuksen voi suoraan suorittaa oletusasetuksilla.

Pythonin versio on oltava 3.10.6 eikä pienempi tai suurempi. Syynä tähän on, että AUTOMATIC1111:n Stable Diffusion käyttöliittymää on testattu versiolla 3.10.6. Jos tietokoneelle asennettu versio on eri, ohjelman kanssa voi tulla ongelmia eikä se tule toimimaan oikein. (AUTOMATIC1111, n.d.-c) Loput vaiheet Pythonista voidaan asentaa oletusasetuksilla. Kestoltaan Pythonin asennus kestää vain hetken.



Kuva 5. Kuvakaappaus Gitin verkkosivuilta Downloads näkymästä.

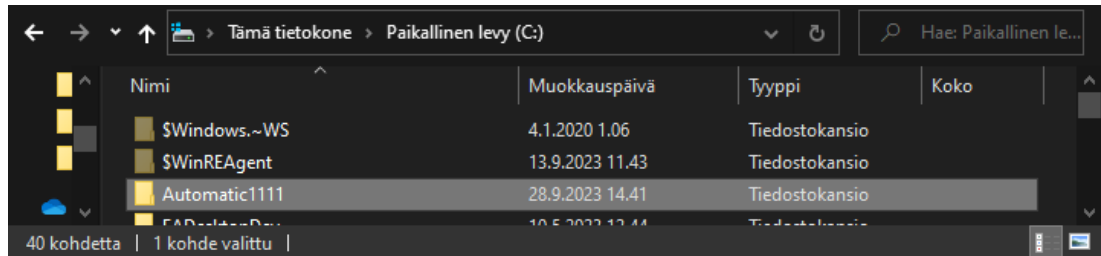


Kuva 6. Kuvakaappaus Pythonin asennusikkunasta. AUTOMATIC1111:n ohjeissa opastetaan, että Pythonia asennettaessa tulee muistaa merkitä valituksi "Add Python X.X. to PATH" ja ladattava versio on oltava 3.10.6.

6.3 Stable Diffusion WebUI:n asennusprosessi

Kun Python ja Git ovat asennettuina, ladataan Githubista AUTOMATIC1111:n Stable Diffusion käyttöliittymä.

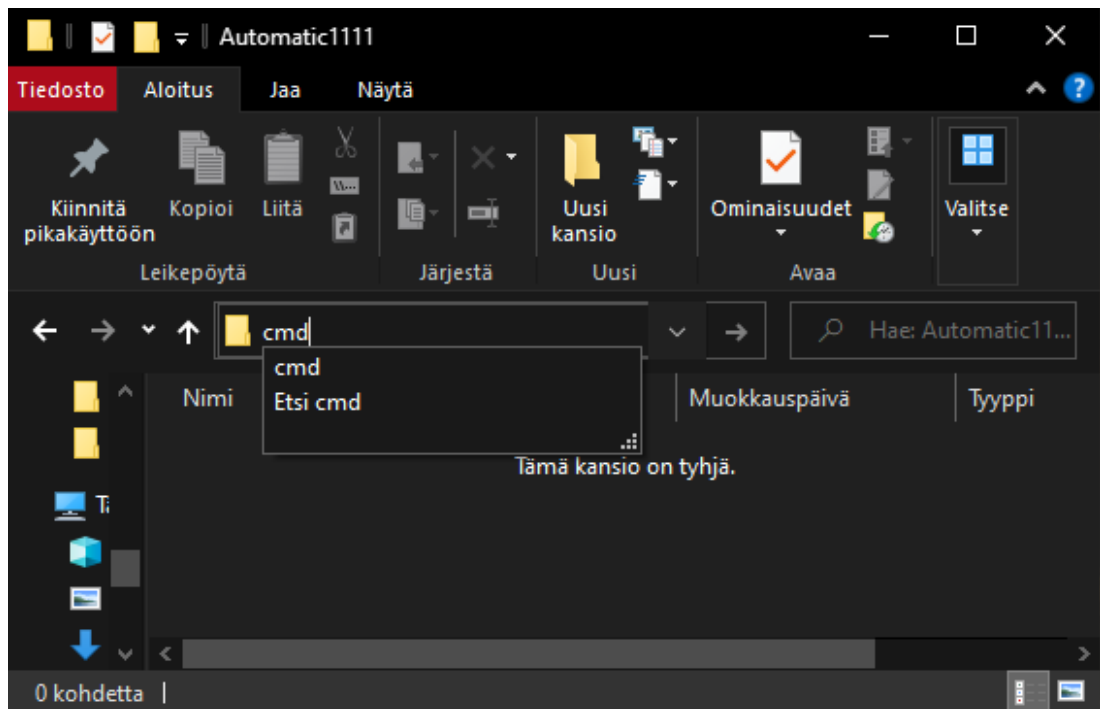
Ensimmäisenä avataan resurssienhallinta ja lisätään paikalliselle levyasemalle uusi kansio vapaaehtoisella nimellä. Selkeyttä varten nimesin kansion käyttöliittymän tekijän nimellä eli Automatic1111 (kuva 7).



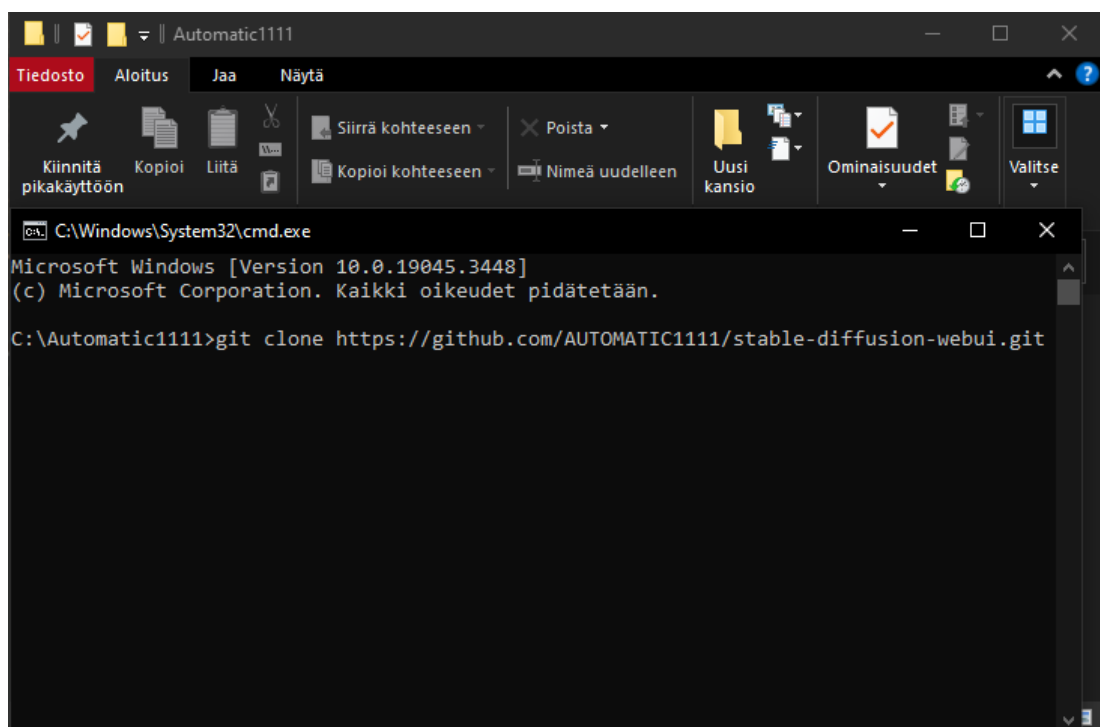
Kuva 7. Kuvakaappaus paikalliselta levyltä, johon on lisätty kansio nimellä Automatic1111.

Kansion luonnin jälkeen kansio avataan ja sen osoitepalkkiin kirjoitetaan `cmd` (kuva 8). Jatketaan eteenpäin painamalla Enteriä, ja näytölle aukeaa komentokehoteikkuna, jossa on heti valmiina sen kansion polku, jonne Stable Diffusion halutaan asentaa (kuva 9).

Polun perään lisätään `git clone https://github.com/AUTOMATIC1111/stable-diffusion-webui.git`. Git viittaa ohjelmaan, joka aikaisemmin ladattiin ja clonen avulla tiedostot kopioidaan omalle tietokoneelle AUTOMATIC1111:n Github -sivuilta.

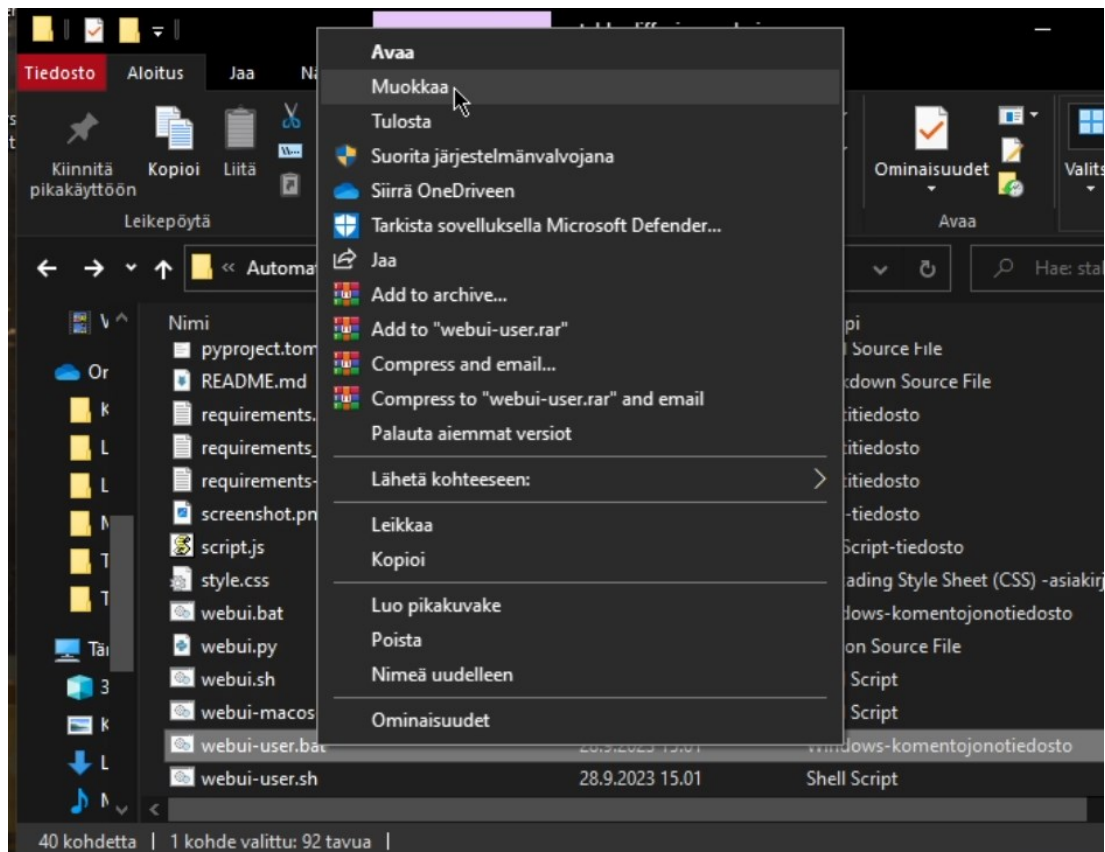


Kuva 8. Kuvakaappaus Automatic1111 nimisestä kansioista, jonka osoitepalkkiin on kirjoitettu cmd.



Kuva 9. Kuvakaappaus ruudulle avautuneesta komentokehoteikkunasta, jossa on valmiina sen kansion polku, jonne Stable Diffusion halutaan. Polun perään on lisätty Git komento, jolla kopioidaan Githubista Stable Diffusionin tiedostot.

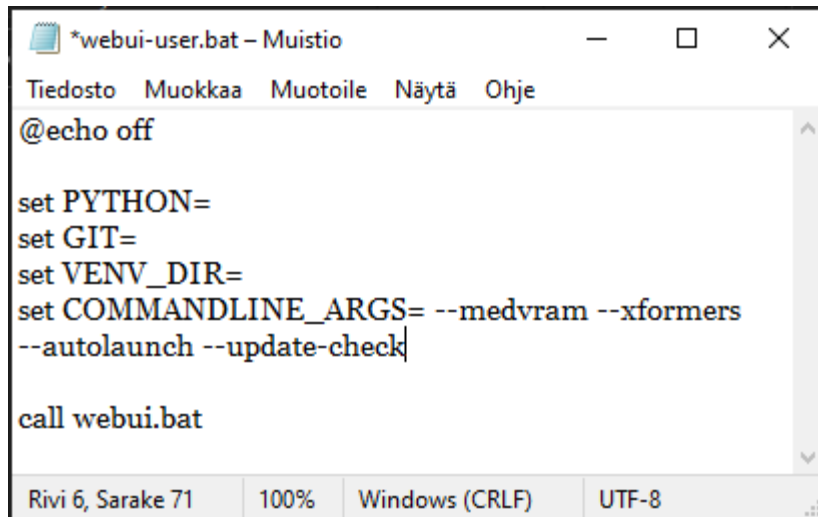
Kun asennus on onnistunut, avataan kansio nimeltä `stable-diffusion-webui`. Kansiosta löytyy `webui-user.bat` niminen tiedosto. Tätä tiedostoa klikkaamalla voidaan käynnistää Stable Diffusion. Kuitenkin ennen ohjelman käynnistämistä, lisätään kyseiseen batch tiedostoon vielä komentoja, jotka tulevat optimoimaan ja helpottamaan generointiprosessia.



Kuva 10. Kuvakaappaus `stable-diffusion-webui` kansioista.

Kuten kuvassa 10, klikataan hiiren oikealla näppäimellä `webui-user.bat` päältä ja valitaan `muokkaa`. Tällöin näytölle aukeaa muistio eli tekstitiedosto. Toinen tapa on avata uusi muistio ja raahata `webui-user.bat` tiedosto siihen.

Kohdan `set COMMANDLINE_ARGS=` perään lisätään generointi prosessia optimoivia komentoja (kuva 11).



```
*webui-user.bat - Muistio
Tiedosto Muokkaa Muotoile Näytä Ohje
@echo off

set PYTHON=
set GIT=
set VENV_DIR=
set COMMANDLINE_ARGS= --medvram --xformers
--autolaunch --update-check

call webui.bat

Rivi 6, Sarake 71 100% Windows (CRLF) UTF-8
```

Kuva 11. Kuvakaappaus tekstitiedostosta, johon on lisätty `set COMMANDLINE_ARGS=` perään uudet komennot.

Koska tässä työssä käytetyn tietokoneen VRAM on 5.8 GB, voidaan skriptiin lisätä varmuuden vuoksi `--medvram` tai `--lowram`. Näiden tarkoituksena on vähentää VRAMin käyttöä. `--xformers` lisää nopeutta generointiprosessiin. Tämän lisäksi ylimääräisinä vaihtoehtoina ovat myös `--autolaunch` ja `--update-check`. Nimensä mukaisesti `--autolaunch` -komennolla web-käyttöliittymä aukeaa käynnistyksen jälkeen selaimessa automaattisesti ja `--update-check` ilmoittaa käyttäjälle, kun web-käyttöliittymästä on ilmestynyt uudempi versio. (AUTOMATIC1111, n.d.-b)

Määritysten jälkeen tekstitiedosto tallennetaan, ja ohjelma käynnistetään kaksoisklikkaamalla `webui-user.bat` käynnistyskriptiä. Käyttöliittymällä ei voi vielä tuottaa kuvia, sillä siinä ei ole vielä mallia. Ennen käynnistystä `Stable-Diffusion` -kansioon voi jo lisätä mallin, mutta se ei ole vielä pakollista. Asennus voi kestää jopa tunnin ajan riippuen tietokoneen tehokkuudesta ja internetyhteydestä. Työssä käytetyllä tietokoneella asennus kesti noin 18 minuuttia. Asennuksen jälkeen käyttöliittymä avautuu automaattisesti selaimeen, kuten edellä `--autolaunch` komennolla oli määriteltykin.

6.4 Stable Diffusionin päivittäminen

Stable Diffusionin manuaalinen päivittäminen onnistuu lähes samalla tavalla kuin Stable Diffusionin kopiaiminen kansioon.

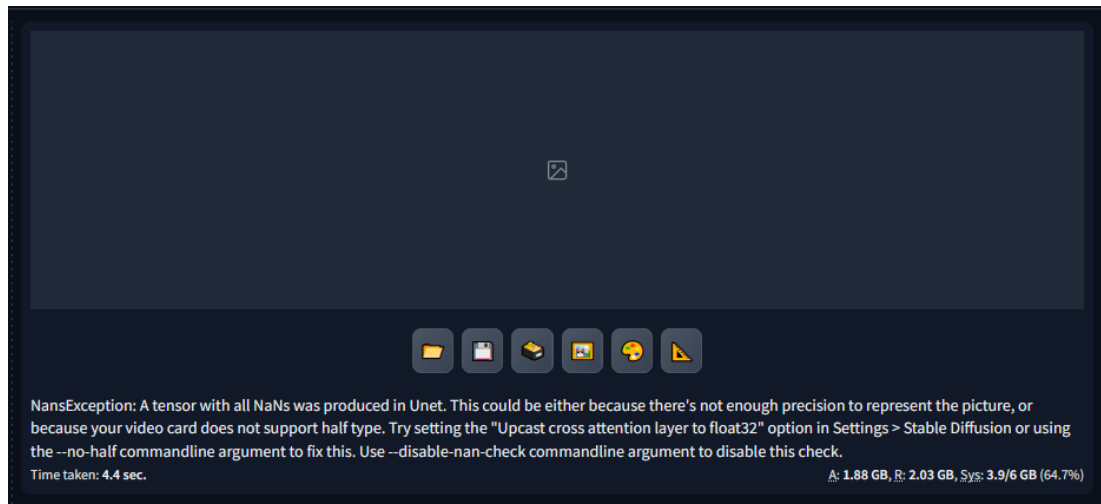
Avataan Stable Diffusion käyttöliittymän kansio. Tässä työssä polku on `C:\Automatic1111\stable-diffusion-webui`. Kansion `stable-diffusion-webui` osoitepalkkiin kirjoitetaan `cmd` ja jatkamalla enterillä eteenpäin ruudulle avautuu komentokehoteikkuna. Polkuun kirjoitetaan `git pull`. Tällä komennolla haetaan ja tarkistetaan Githubista, onko AUTOMATIC1111:n Stable Diffusionista tullut uutta versiota. (Kamph, 2023.)

Toinen tapa on määrittää Stable Diffusion päivittymään automaattisesti. Tällöin `webui-user.bat` tiedostoon lisätään ennen `call webui.bat` kehoitetta `git pull`. Kun käyttöliittymä käynnistetään, se päivittyy käynnistytksen hetkellä automaattisesti.

6.5 Virheidenkorjaus

Tämän työn aikana ilmeni yksi ongelma. Kuvaa ei generoitunut, mutta tilalle tuli kuvassa 12 esiintyvä virheilmoitus. Kokeilin asettaa ilmoituksen mukaan ”Upcast cross attention layer to float32” asetuksen päälle. Tämä ei kuitenkaan auttanut, vaan oli palattava lisäämään komentoja `webui-user.bat` tiedostoon.

Selvisi, että NVIDIA 16XX tai 10XX grafiikka korteissa pitäisi käyttää `--upcast-sampling` ja `--xformers` komentoja (AUTOMATIC1111, n.d.-c). Nämäkään kaksi eivät riittäneet. Jätin `--upcast-sampling` komennon pois ja laitoin tilalle `--precision full` ja `--no-half` komennot (kuva 13). Tämän jälkeen kuvia pystyi generoimaan.



Kuva 12. Kuvakaappaus virheestä, joka ilmeni kuvaa generoidessa.

```
webui-user.bat - Muistio
Tiedosto Muokkaa Muotoile Näytä Ohje
@echo off

set PYTHON=
set GIT=
set VENV_DIR=
set COMMANDLINE_ARGS= --medvram --xformers --
autolaunch --update-check --precision full --no-half

call webui.bat
```

Rivi 8, Sarake 15 100% Windows (CRLF) UTF-8

Kuva 13. Kuvakaappaus webui-user.bat tekstitiedostosta, johon on lisätty lopulliset komennot: --precision full ja --no-half.

7 OMAN TEKOÄLY MALLIN OPETTAMINEN DREAMBOOT-HILLA

Tässä luvussa käytämme pilvipohjaista Google Colabia ja opetamme Colabissa Dreambooth -mallin omilla valokuvilla. Työssä käytetään mallin opettamiseen kuvia kahdesta lemmikkikanista: Toffista ja Suskusta.

7.1 Google Colab

Google Colaboratory, lyhyesti Colab, on Google Researchin tuote. Colabin avulla voi suorittaa Python-koodia selaimen kautta, ja se soveltuu erityisesti koneoppimiseen, data-analyysiin ja koulutukseen. Colab on isännöity Jupyter notebook -palvelu, joka tarjoaa maksuttoman pääsyn esimerkiksi GPU:ihin. (Google, n.d.) Google Drivessa muistin käyttö on ilmaista 15 GB asti, mutta tämän jälkeen käytöstä tulee maksullista. Koska haluan ottaa mallista kaiken hyödyn irti, suosittelen henkilökohtaisesti luomaan työtä varten oman Google tilin, varsinkin, jos omalla tilillä on jo muistia käytetty riittävästi.

7.2 DreamBooth

DreamBooth on menetelmä, joka mahdollistaa tekstistä kuvaksi mallien, kuten Stable Diffusionin, personalisoinnin vain muutaman kuvan perusteella (Hugging Face, n.d.) Oman mallin kouluttamisesta, personalisoinnista, käytetään nimitystä fine tuning eli hienosäätäminen. Koneoppimisessa hienosäätö tarkoittaa sitä, että yhdellä datajoukolla koulutettua mallia mukautetaan toimimaan uudella, vastaavalla tietojoukolla (EdXD, 2022.)

DreamBooth toimii myös vaihtoehtoisesti laajennuksena (extension) suoraan Automatic1111:n Stable Diffusion käyttöliittymässä. Päädyin tässä työssä käyttämään DreamBoothia Colab-ympäristössä, koska tällöin ei tarvitse käyttää tietokoneen omaa grafiikkaprosessoria, kuten paikallisesti käytettäessä tarvitsisi.

7.3 Ohjeet omien kuvien ottoon ja käyttöön

Kuvia on suositeltavaa ottaa 512x512 pikselin muodossa, sillä Stable Diffusion mallit ovat opetettu 512x512 pikselin omaamilla valokuvilla. Puhelimen kamerasta pystyy suoraan valitsemaan kuvasuhteen 1:1. Kuvia voi myös jälkikäteen muokata Windowsin valokuva albumissakin tai verkossa, kuten Birme.net tai Paint.net -sivustoilla. Muista ottaa kuvia eri suunnista, eri valaistuksissa ja eri vaatteissa. Ota kuvia läheltä sekä kaukaa. Jos kuvien opettamiseen käytetään ihmistä, kuvissa kannattaa näyttää erilaisia ilmeitä. Yleensä noin 10–30 kuvaa on riittävä määrä opetukseen. (Cunliffe, 2022a; Render Realm, 2023; SECourses, 2022.)

Opinnäytetyössä käytetään esimerkkinä 20 valokuvaa lemmikkikani Toffista. Osassa kuvia on pientä sumeutta ja toistuvasti heinää Toffin ympärillä, mikä hieman kostahtuu lopputuloksessa. Vertailun vuoksi koulutin vielä toisen mallin 40:llä Suskun kuvalla, joista osa on terävämpiä, tausta on poistettu sekä kuvakulmat ovat vaihtelevampia.

7.4 Install Requirements ja Login to Hugging Face

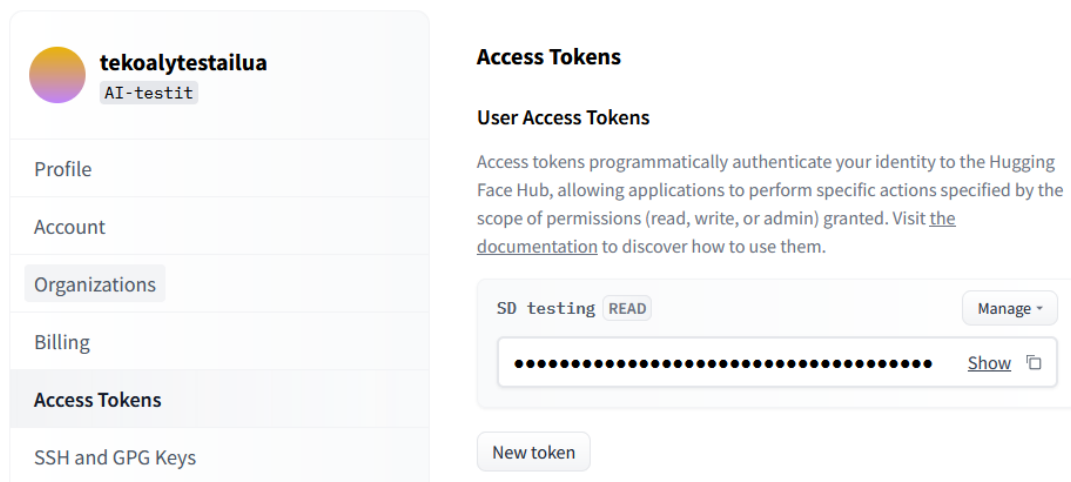
Colabin ensimmäisessä vaiheessa suoritetaan ensimmäinen solu (cell) nimeltä "Install Requirements". Suorituksen aikana voi tulla virheilmoitus liittyen PIPiin eli Python-pakettien paketin hallintaohjelmaan (W3Schools, n.d.). Virhe ei kuitenkaan vaikuta lopputulokseen, joten sen voi rauhassa ohittaa.

Hugging Faceen tulee luoda ilmainen käyttäjätili, koska Stable Diffusionin 1.5 versiota ylläpidetään siellä (Cunliffe, 2022b). Tokenin avulla DreamBoothille annetaan käyttöoikeudet Hugging Face -rajapintaan, jotta DreamBooth pystyy noutamaan tietoja Hugging Facen järjestelmästä ja kouluttamaan mallin (All Your Tech AI, 2023a).

Hugging Facessa käyttäjäasetusten kautta löytyy kohta nimeltä "Access Tokens", jossa User Access Tokensiin lisätään uusi token (kuva 14). Tokeniin

voi asettaa haluamansa mukaan read tai write. Lopuksi tokeni lisätään Colabissa sen määrättyyn kohtaan ja solu suoritetaan.

Ainoastaan Reddit-yhteisön keskustelusta löytyi tieto, ettei ole väliä kumman laittaa. Käyttäjän Remarkable-Llama616:n (Remarkable-Llama616, 2023) mukaan read riittää siten, että se antaa luvan käydä poimimassa esimerkiksi Stable Diffusion 1.5 mallin Hugging Facesta. Write mahdollistaa sen, että käyttäjä voi tallentaa ja siirtää kustomoidun mallinsa Hugging Faceen.



Kuva 14. Kuvakaappaus Hugging Face käyttäjätilin asetuksista, jossa on luotu uusi Access Token nimeltä SD testing.

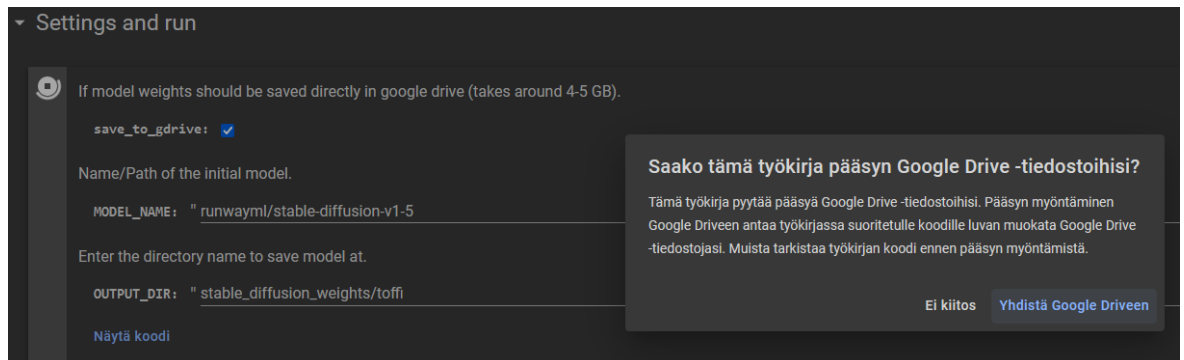
7.5 Settings and run

Alla olevassa kuvassa 15 tallennetaan malli Google Driveen eli hyväksytään save_to_gdrive ruudukko. Tällöin mallia ei tarvitse enää opettaa uudestaan ja sitä voidaan käyttää kuvageneraatioissa. (SECourses, 2022.)

Kuten kuvassa 15, MODEL_NAME pysyy samana ja OUTPUT_DIR määrää sen osoitteen, jonne malli tallennetaan, joten sen loppuun voi kirjoittaa haluamansa osoitenimen. Malli tallentuu käyttäjän valitsemaan Google Drive kansioon eli tässä tapauksessa kansioon nimeltä toffi.

Kun solu suoritetaan, Colab pyytää kuvan 15 mukaan pääsylupaa käyttäjän Google Drive kansioihin ja vaatii sisäänkirjautumista Google tilille. Kaikki

vaadittavat kohdat voidaan hyväksyä, jotta Google Driveen voidaan määritellä kaikki tarvittavat hakemistot. (All Your Tech AI, 2023a.)



Kuva 15. Kuvakaappaus Settings and run solusta, jossa tarvittavat kohdat ovat täytetty.

7.6 Start Training

Seuraavaksi aloitetaan mallin kouluttaminen (kuva 16). Start Training vaiheessa määritellään ensimmäisenä konseptilistaan instance prompt. Instance prompt on esimerkkikehote, avainsana, kutsuma nimi, sille asialle, joka mallille annetaan opittavaksi.

Nimi saa olla mikä tahansa, mutta on tärkeää ottaa huomioon, että nimen on parhaimman lopputuloksen saavuttamiseksi oltava uniikki. Esimerkiksi nimi "Emma" voi olla Stable Diffusionin mallille jo entuudestaan tuttu, koska sitä on voitu jo kouluttaa Emma nimisillä naisilla. Tällöin mallilla voi olla jo oma käsityksensä Emmasta. Sekaannuksen pystyy ratkaisemaan luomalla nimestä tarpeeksi omalaatuisen. Nimessä kannattaa käyttää, vaikka osaa omasta sukunimestä, lukujen ja tekstin yhdistelmää, keksittyä sanaa tai mitä vain normaalia erikoisempaa. Tällöin nimi on uniikki ja sellainen, jota malli ei jo ennestään tiedä. Jatkossa tätä keksittyä nimeä tullaan käyttämään, kun generoidaan kuvia haluamastamme asiasta.

Class_promptilla voidaan tarjota alkuperäisten kuvien lisäksi geneerisiä kuvia, jotka liittyvät määritettyyn aiheeseen parantaen generoitavien kuvien laatua

(SECourses, 2022). Tässä tapauksessa laitoin class promptiksi "photo of a rabbit".

Instance_data_dir tarkoittaa kansiota, joka sisältää mallikuvien eli Toffin kuvien opetusdatan. Class_data_dir on kansio, joka sisältää luokkakuvien opetusdatan eli Toffia muistuttavien kanien kuvat. (Foong, 2022). Kansioiden nimet saavat olla vapaavalintaisia.

Kun kaikki on valmista, solu suoritetaan ja suorituksen aikana Google Driveen luodaan opetusdatan tuontia varten oma hakemistorakenne. (All Your Tech AI, 2023a) Kun solu on suoritettu loppuun, siirretään opetuskuvat Colabiin ohjeiden mukaisesti (kuva 17).

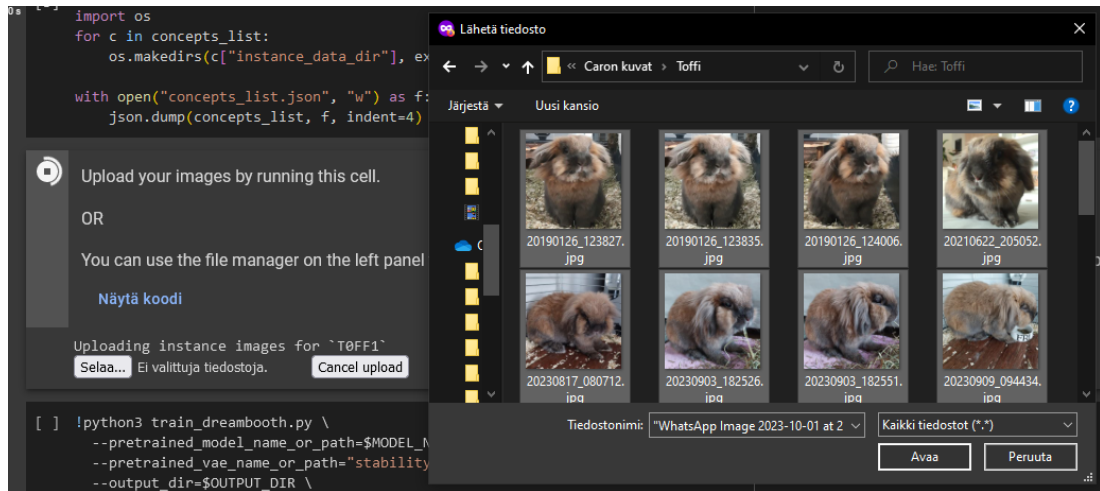
```
# You can also add multiple concepts here. Try tweaking `--max_train_steps` accordingly.

concepts_list = [
    {
        "instance_prompt": "T0FF1",
        "class_prompt": "photo of a rabbit",
        "instance_data_dir": "/content/data/T0FF1model",
        "class_data_dir": "/content/data/rabbit"
    },
    # {
    #     "instance_prompt": "photo of ukj person",
    #     "class_prompt": "photo of a person",
    #     "instance_data_dir": "/content/data/ukj",
    #     "class_data_dir": "/content/data/person"
    # }
]

# `class_data_dir` contains regularization images
import json
import os
for c in concepts_list:
    os.makedirs(c["instance_data_dir"], exist_ok=True)

with open("concepts_list.json", "w") as f:
    json.dump(concepts_list, f, indent=4)
```

Kuva 16. Kuvakaappaus konseptilistasta, jossa termit ja kuvailut ovat vaihdettu.



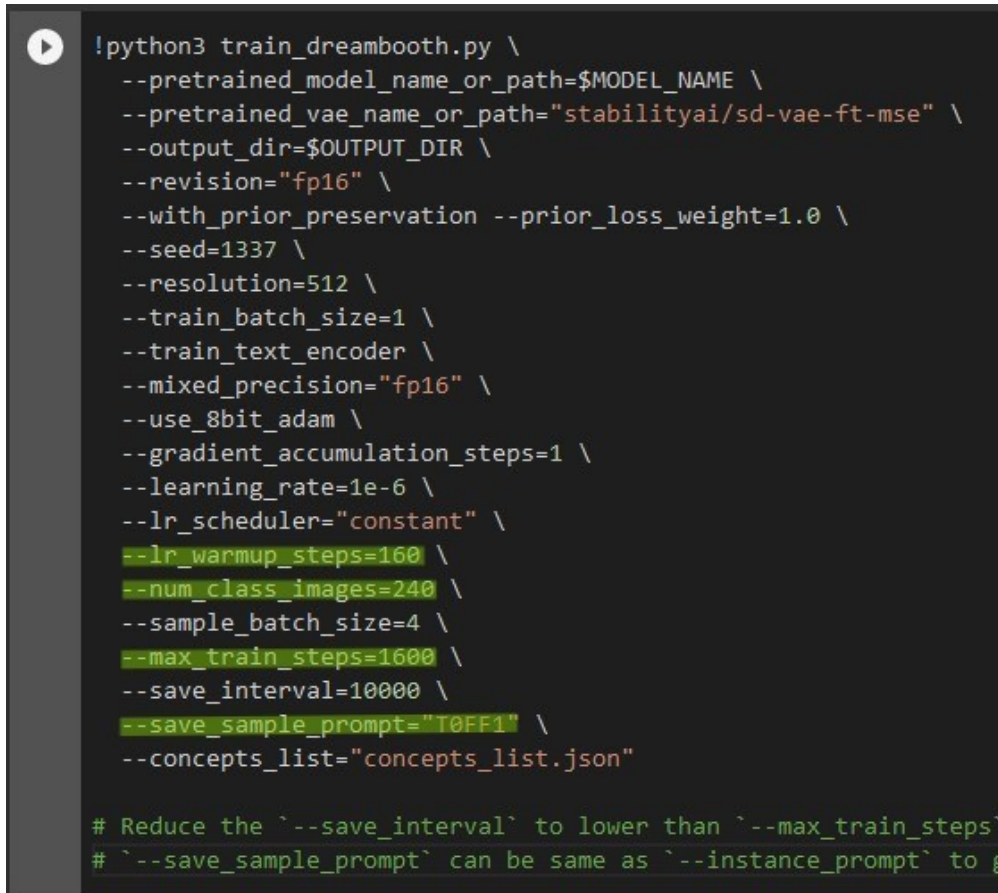
Kuva 17. Kuvakaappaus vaiheesta, jossa Toffin kuvat ladataan Colabiin.

Sillä välin, kun opetuskuvat latautuvat, voidaan muuttaa seuraavan solun parametreit. Tämä osa tehdään SECoursesin (SECourses, 2022) videon mukaisesti. Työssä opetan mallin videon mukaan, koska siinä huomioitiin enemmän kohtia kuin muissa vastaavissa ohjeissa.

Mallin opetusta varten on valittuna 20 valokuvaa. Nämä kerrotaan luvulla 12 ja saadaan num_class_imagesiin luku 240. Max_train_steps kertoo mallille, kuinka monta kertaa sitä koulutetaan. Valokuvien määrä kerrotaan luvulla 80 ja saadaan maksimi opetusaskeliksi 1600. Lr_warmup_steps on max_train_steps eli tässä tapauksessa luku 1600 jaettuna luvulla 10. Saadaan lr_warmup_stepsiksi 160. Huomioitana on, että lr_warmup_steps tulee olla integer eli kokonaisluku. Kuvassa 19, näkyvät vihreällä värillä tässä lasketut parametrit.

Näiden määritelmien jälkeen solu voidaan suorittaa ja malli aloittaa opetusprosessin. Valmis malli tallentuu aikaisemmassa vaiheessa (Luku 6.5. Settings and run) määritettyyn kansioon. Tässä tapauksessa kansioon nimeltä toffi.

Mallin oppiessa, tulee Google Colabissa olla välillä aktiivinen, koska inaktiivisuuden takia koko opetusprosessi keskeytyy ja opetus tulee aloittaa täysin alusta. (Cunliffe, 2022b) Aktiivisuus ongelman saa korjattua vain, jos hankkii Colab Pron.



```
!python3 train_dreambooth.py \
  --pretrained_model_name_or_path=$MODEL_NAME \
  --pretrained_vae_name_or_path="stabilityai/sd-vae-ft-mse" \
  --output_dir=$OUTPUT_DIR \
  --revision="fp16" \
  --with_prior_preservation --prior_loss_weight=1.0 \
  --seed=1337 \
  --resolution=512 \
  --train_batch_size=1 \
  --train_text_encoder \
  --mixed_precision="fp16" \
  --use_8bit_adam \
  --gradient_accumulation_steps=1 \
  --learning_rate=1e-6 \
  --lr_scheduler="constant" \
  --lr_warmup_steps=160 \
  --num_class_images=240 \
  --sample_batch_size=4 \
  --max_train_steps=1600 \
  --save_interval=10000 \
  --save_sample_prompt="T0FF1" \
  --concepts_list="concepts_list.json"

# Reduce the `--save_interval` to lower than `--max_train_steps`
# `--save_sample_prompt` can be same as `--instance_prompt` to g
```

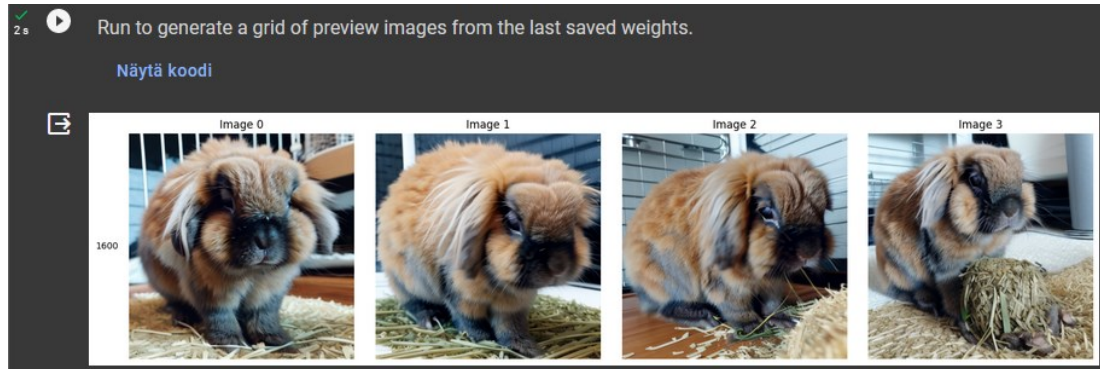
Kuva 19. Kuvakaappaus solusta, jossa näkyy vihreällä maalattuna ne parametrit, jotka määriteltiin ja muokattiin SECourses videon mukaisesti.

7.7 Valmiit T0FF1 ja 5U5KU mallit

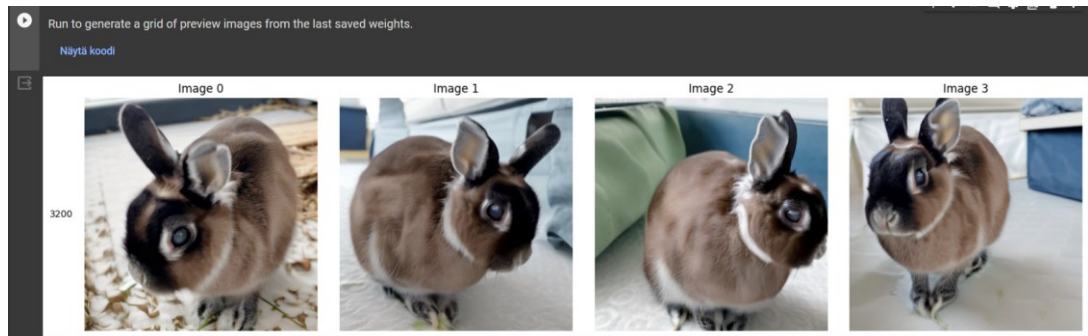
Mallin lopulliseen kouluttamiseen kului aikaa 34 minuuttia. Kouluttamisen jälkeen voi suorittaa solun, joka generoi neljän esikatselukuvan sisältävän ruudukon. Kuvassa 20, Toffin kuvat ovat mielestäni melko hyvät. Ainoat huonot puolet ovat, että opetuskuvissa oli Toffin lisäksi heinää, joten heinä on tällä hetkellä pysyvä elementti. Samoin opetuskuvissa Toffin turkki oli toisinaan sumea, joten sekin korostuu kuvissa. Kehotteiden avulla voi kuitenkin osittain opastaa ylimääräiset asiat kuten heinän ja purut pois.

Vastaavasti malli, joka koulutettiin 40:llä Suskun kuvalla, vei aikaa tunnin ja kuusi minuuttia. Kuvan 21 kuvista silti huomaa, että niissäkin on Suskun karvoissa ja päässä sumeutta ja epäselkeyttä. Susku näyttää kuitenkin enemmän itseltään kuin Toffi ja mielestäni kuvista tuli vielä onnistuneempia kuin Toffin kuvista. Vaikka Suskun kuvia olikin enemmän, ei se silti automaattisesti

tarkoittanut paljon parempaa lopputulosta. Saattaa olla, ettei Suskun kuvat olleet tarpeeksi selkeitä, niissä ei ollut tarpeeksi vaihtelevuutta tai kuvia oli mää-
rällisesti liikaa.



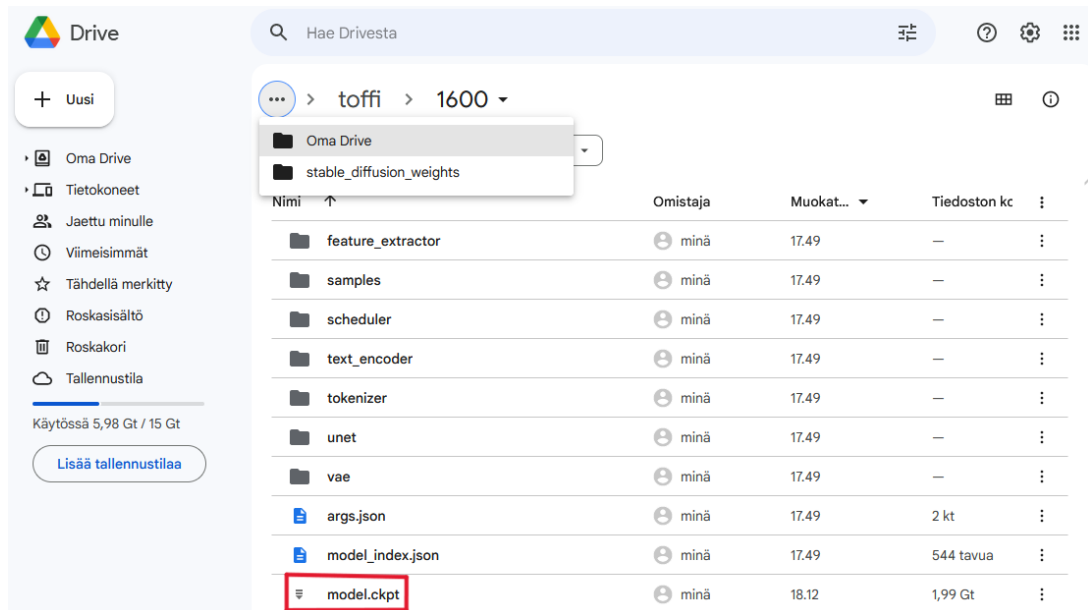
Kuva 20. Neljän esikatselukuvan ruudukko Toffista.



Kuva 21. Neljän esikatselukuvan ruudukko Suskusta.

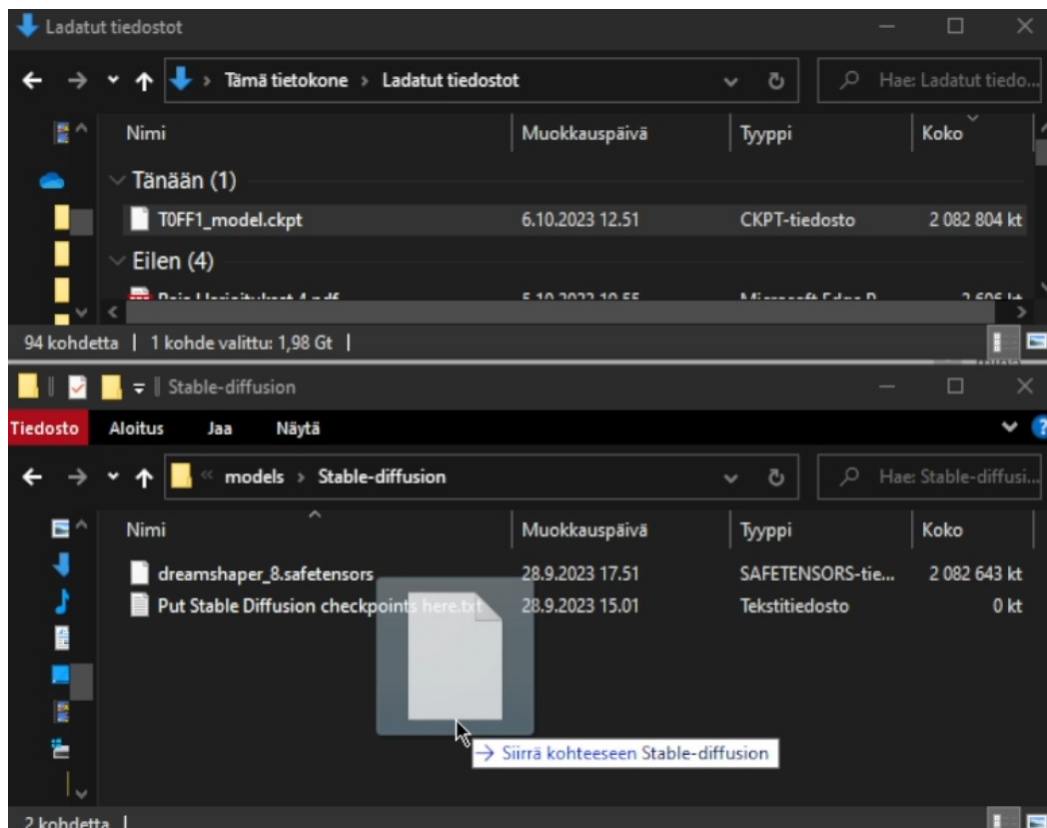
7.8 Personalisoidun mallin käyttöönotto

Kuvien tarkastelun jälkeen tulee suorittaa ruudukon yllä oleva "Specify the weights directory to use" -solu ja sen jälkeen suoritetaan "Convert Weights to ckpt to use in web UIs like AUTOMATIC1111" -solu. Fp16 saa olla valittuna. Tämän jälkeen lopullinen malli löytyy ckpt eli checkpoint tiedostona Google Drivesta. Tässä työssä malli löytyy kansio-polusta `Oma Drive\stable_diffusion_weights\toffi\1600` (kuva 22).



Kuva 22. Kuvakaappaus Google Drivesta, jossa näkyy lopullinen malli model.ckpt.

Malli ladataan Google Drivesta tietokoneelle ja siirretään Stable Diffusionin kansioopolkun stable-diffusion-webui\models\Stable-diffusion (kuva 23). Mallin nimen voi Google Drivessä tai resurssienhallinnassa uudelleen nimetä mallia kuvaavammaksi.



Kuva 23. Kuvakaappaus T0FF1 mallin siirrosta *Stable-Diffusion* -kansi-oon.

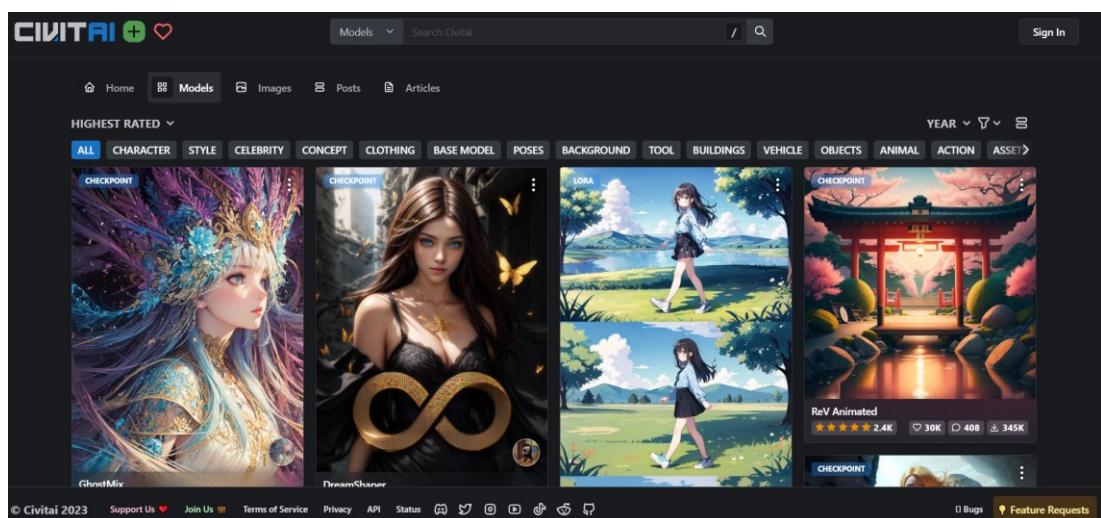
7.9 Hugging Face ja CivitAI

Edellä mainitusta Hugging Face sivustosta voi myös ladata käyttöliittymän käyttöön valmiita malleja. Hugging Facesta löytyvät esimerkiksi *Stable Diffusion*in eri versiot, kuten 1.5, 2.0, 2.1 ja *SDXL* – mallit.

Sivusto, josta voi Hugging Facen lisäksi ladata valmiita malleja, on nimeltään CivitAI (kuva 24). CivitAI on vuonna 2022 perustettu alusta, joka syntyi hienosäädön harrastajien kasvavaan kysyntään paikasta, jossa harrastajat voisivat saumattomasti jakaa teoksiaan ja olla vuorovaikutuksissa muiden tekijöiden kanssa. CivitAI:ssa voi ladata, jakaa ja tutustua mukautettuihin malleihin, joista jokainen malli on koulutettu erilaisin datajoukoin (CivitAI, n.d.).

Jos käyttäjä haluaa generoida kuvia mukaillen jotain tiettyä tyyliä, CivitAI -sivustolta on mahdollista ladata sitä vastaava malli.

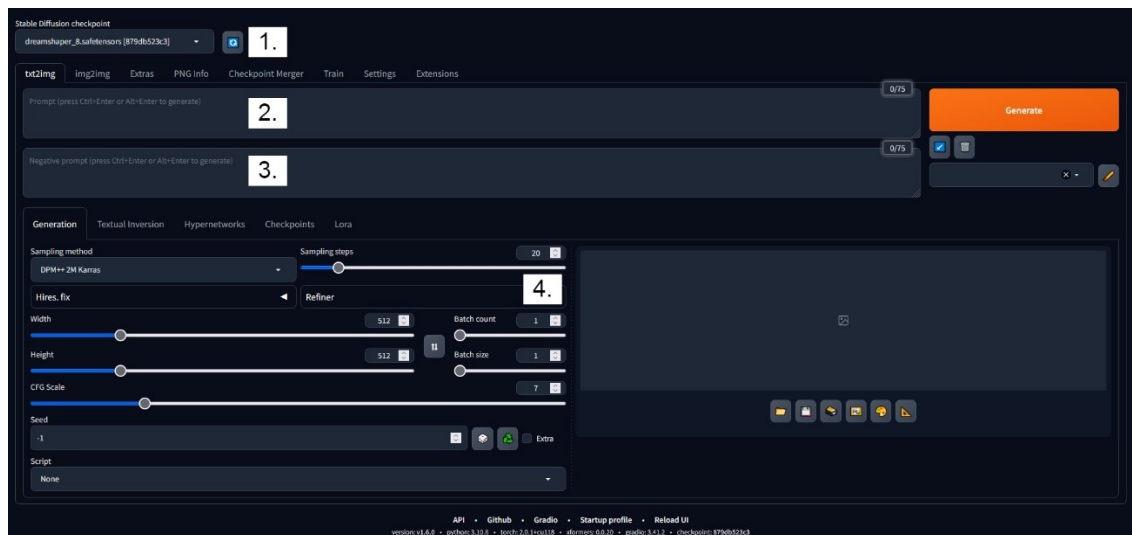
Turvallisinta on valita mallit, joissa on mainittuna sana *safetensor*, koska näihin on vaikeampi asettaa haittaohjelmia. Sivustolla voi törmätä *safetensor*in lisäksi sanaan *checkpoint*. Ero näissä kahdessa on se, että *safetensor*it ovat uudempia malleja kuin *checkpoint*it. (All Your Tech AI, 2023c.)



Kuva 24. Kuvakaappaus CivitAI:n mallit välilehdestä. Kuvassa näkyvät tämän vuoden aikana suosituimmat mallit: GhostMix, DreamShaper, Detail Tweaker LoRa ja ReV Animated.

8 AUTOMATIC1111 KÄYTTÖLIITTYMÄN ESITTELY

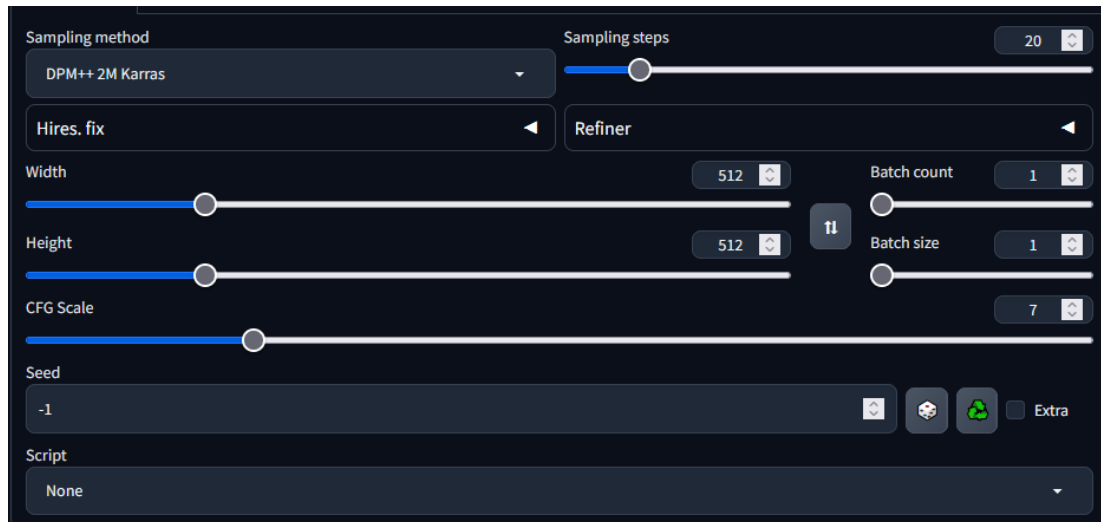
Luvussa läpikäydään AUTOMATIC1111 käyttöliittymä ja selitetään auki tärkeimmät termit. Koska tässä työssä kerrotaan vain tekstistä kuvaksi toiminnasta, perehdytään tässä kappaleessa vain käyttöliittymän ensimmäiseen txt2img välilehteen.



Kuva 25. Kuvakaappaus AUTOMATIC1111:n Stable Diffusion käyttöliittymästä, johon on numeroitu vaiheittain läpikäytävät osiot.

Kuvan 25 vasemmassa yläkulmassa on ensimmäisenä Stable Diffusion checkpointtien pudotusvalikko. Täältä löytyvät ne Stable Diffusion -kansioon lisätyt mallit, jotka lisättiin edellisessä luvussa (7.8. Personalisoidun mallin käyttöönotto). Kyseisestä valikosta pystyy valitsemaan kuvan luontia varten haluamansa mallin. Viereisestä sinisestä päivitys napista saa päivitettyä mallien listan. Tämän ansiosta käyttöliittymää ei tarvitse uudelleen käynnistää, vaikka Stable Diffusion -kansioon olisi käynnissäoloaikana lisätty uusi malli.

Toisena ja kolmantena ovat tekstikentät positiivisille ja negatiivisille tekstikehoille. Molemmat tulevat opastamaan kuvanluonti prosessia. Positiivisella eli tavallisella kehoitteella määritetään ne asiat, joita lopulliseen generoitavaan kuvaan halutaan. Negatiivinen kehoite määrittää ne asiat, joita käyttäjä ei halua kuvaansa. Neljäntenä löytyy erilaisia toimintoja, joilla voidaan säätää kuvanluonti prosessia (kuva 26).



Kuva 26. Kuvakaappaus kuvagenerointia ohjaavista toiminnoista.

Sampling Method eli suoraan käännettynä näytteenottomenetelmä. Tästä pudotusvalikosta pystyy valitsemaan erilaisia algoritmeja kohinan poistoprosessia varten. Tähän ei ole oikeaa vaihtoehtoa, joten jokaista voi mielensä mukaan koittaa. (Andrew, 2023.) Esimerkiksi CivitAI:ssa käyttäjät voivat ilmoittaa sopivasta sampling methodista mallien kuvauksissa.

Sampling stepsillä voi vaikuttaa siihen, kuinka monta kertaa kuvaa työstetään ja pyritään muuttamaan tai parantamaan. Liian pieni askellus tuottaa vain sumean tuloksen ja liian suuri askellus voi tuottaa erittäin räikeän kuvan. Huonon tuloksen lisäksi, liian suuri askellus vie yhä enemmän aikaa prosessoida. (Ekman, 2023; Kafke, 2022.)

Widthilla ja Heightilla voi hallita kuvan resoluutiota. Mitä suuremmat arvot ovat, sitä enemmän ne vaativat laitteen VRAMia. (Ekman, 2023.) Parhaimmat

resoluutiot Jim Clyde Mongen (Monge, 2023) mukaan ovat 512x512 ja 768x768 pikseliä.

Batch count eli erien lukumäärä kertoo, montako erää kuvia generoidaan kerralla. Batch size eli erän koolla määritetään generoitavien kuvien määrä. Mitä suurempi erän koko on, sitä enemmän se vie VRAMia ja aikaa generoida. (Ekman, 2023.) Esimerkiksi, jos batch size on neljä ja batch count kaksi, kuvia tulee generoitua yhteensä kahdeksan. Eli kahdessa eri erässä, joissa molemmissa on neljä kuvaa.

CFG (Classifier Free Guidance) Scale eli luokittelijan vapaa ohjausasteikko. CFG-asteikko ohjaa, kuinka tarkalleen tai täsmällisesti malli noudattaa kehoitteita. Asteikon oletusarvo on seitsemän, mutta optimaalisesti käytettävä arvo voi vaihdella malleittain. Jos arvo on alle seitsemän, käyttäjän kehotetta kuunnellaan vähemmän ja malli on luovempi. Mitä korkeampi arvo on, sitä tarkemmin käyttäjän kehotetta noudatetaan. CFG-arvon ollessa liian pieni tai suuri, ei ole myöskään mitään hyötyä, sillä silloin lopputulos voi olla erittäin sumea tai kuvan värit voivat olla erittäin räikeät. (Andrew, 2023b; Ekman, 2023; Kafke, 2022.)

Seed eli siemen. Jokaisella generoidulla kuvalla on oma siemenlukunsa. Käyttämällä tietyn kuvan siemenlukua voi generoida samankaltaisia kuvia. Vihreää kierrätys nappia painamalla saa kopioitua kuvan siemenluvun. Nopasta asetetaan tekstikenttään takaisin luku -1, jolloin kuva generoidaan satunnaisena.

Jokainen generoitu kuva tallentuu kansiopolkuun `stable-diffusion-webui\outputs\txt2img-images`.

9 YLEISET PROMPTAUS TEKNIIKAT

Selkeiden ja laadukkaiden kuvien tuottaminen vaatii käyttäjältä tietoa ja taitoa. Tässä luvussa esitellään sekä yleisiä että AUTOMATIC1111 käyttöliittymään päteviä tekniikoita, joiden avulla käyttäjä voi päästä kuvien tuotossa lähemmäs paremman laatuista lopputulosta.

9.1 Kehotteen rakenne

Kun käyttäjä on ideoimassa kuvaansa, tulee huomioida aihe, medium eli materia, jota kuva mukailee, tyyli eli kuvan taiteellinen tyyli, resoluutio, lisätiedot, väri ja valaistus. Jokaisen kohdan kategoriaa ei kuitenkaan tarvitse eikä ole pakko käyttää. Tätä listaa voi pitää muistutuksena siitä, mitä vaihtoehtoja on olemassa. (Andrew, 2023a.)

Aihe on se pääasia, joka kuvassa halutaan nähdä. Yleinen virhe aloittelijoilla on jättää kuvailematta kuvaa tarpeeksi yksityiskohtaisesti. Tällöin tekoälylle jää paljon vapautta olla luova ja lopputulos ei välttämättä ole hyvä. (Andrew, 2023b)

Medium tarkoittaa materiaa, jota kuvan on tarkoitus mukailla. Esimerkkejä tästä ovat oil painting, digital painting, 3D rendering ja pencil drawing. Tyyli tarkoittaa kuvan taiteellista tyyliä. Näihin kuuluvat esimerkkeinä surrealist, pop art, concept art, hyperrealistic, studio portrait ja fantasy. (Andrew, 2023a; anntartica, n.d.; Ma, 2023.)

Resoluutio kertoo, kuinka tarkkana ja yksityiskohtaisena kuva tuotetaan. Tähän esimerkkeinä voivat toimia sanat highly detailed, extremely detailed, 4k ja best quality. (Andrew, 2023a; Ma, 2023.)

Lisätietoihin voi sisällyttää kaikkea muuta kuvan kohdetta tai kuvanlaatua ehostavaa, kuten scifi, highly beautiful ja dramatic. Valokuvaussanastosta voi olla hyötyä, kun kuva halutaan tietystä perspektiivistä. Näihin lukeutuvat

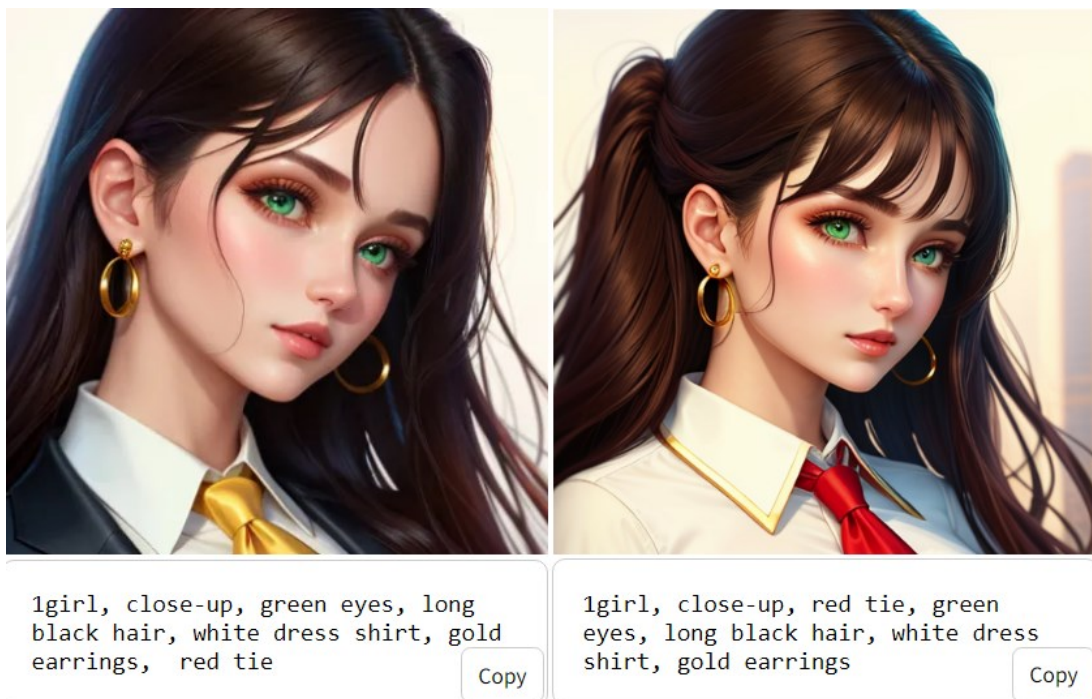
esimerkiksi sanat fisheye, close-up, medium shot ja full body shot. (Andrew, 2023a; anntartica, n.d.; Ma, 2023.)

Värillä saa muokattua kuvan värimaailmaa ja valaistuksella voidaan kuvailla valaistusta, joka kuvaan halutaan. Näihin sopivat sanat vibrant, vivid, vintage, cinematic lighting, gloomy, direct sunlight ja natural lighting. (Andrew, 2023a, 2023c; anntartica, n.d.; Ma, 2023.)

9.2 Sanajärjestys

Sanajärjestyksellä on myös merkitystä. Se asia, mitä käyttäjä haluaa eniten kuvassaan painottaa, tulee mainita ensimmäisenä. Ne asiat, jotka tulevat olemaan käyttäjälle vähiten tärkeitä, tulee jättää viimeisimmiksi.

Toisinaan, jos kuvaan vaaditaan monia erilaisia värejä, ne voivat mennä toisensa kanssa sekaisin. Tällöin on hyvä vaihtaa sanojen paikkaa, kuten alla olevassa esimerkkikuvassa 27.



Kuva 27. Kuvissa ovat täysin samat kehoitteet, mutta kehoitteiden paikkaa on vaihdettu. Todennäköisesti punainen solmio meni sekaisin kultaisten

korvakorujen kanssa, joten oli kannattavampaa vaihtaa kehotteen sijaintia eri paikkaan. (Ma, 2023).

9.3 Negatiivinen kehote

Negatiivinen kehote (negative prompt) on nimensä mukaisesti vastakohta tavalliselle kehotteelle. Negatiiviseen kehotteeseen tulee avainsanoina ne asiat, joita lopputuloksessa halutaan välttää, kuten epämuodostuneet sormet tai epäselvät kasvot.

Negatiivinen kehote ei kuitenkaan toimi siinä mielessä, että tekoälyä vain kättäisiin korjaamaan epämuodostuneet kädet. Kehote ennemminkin näyttää latentissa avaruudessa suunnan ja opastaa pois siitä osasta, jossa todennäköisesti esiintyy negatiivisena kehotteena mainittu asia, kuten tässä esimerkissä, epämuodostuneet kädet. (Ma, 2023)

Negatiivinen kehote on kuvageneraatiossa kätevä työkalu, koska sen avulla voi helposti poissulkea, vaikka värejä tai virheitä, joita tekoäly saattaa kuvaan tuottaa. Jos ei muita tekniikoita halua käyttää, on ainakin kätevää soveltaa negatiivista kehotetta.

9.4 () ja [] syntaksit

Tämä luku koskee vain AUTOMATIC1111:n käyttöliittymää. Sulku ja hakasulku merkeillä voidaan joko korottaa tai alentaa kehotteen merkitsevyyttä. Oletusarvoisesti painoarvo on aina 1. (Andrew, 2023a; Ma, 2023.)

Sulkumerkkejä käytetään tietyn kehotteen painoarvon lisäämiseksi, kuten esimerkiksi korostettu kehote `(brown eyes)` tarkoittaisi samaa kuin `(brown eyes:1.1)`. Painoarvoa voi lisätä joko lisäämällä useamman sulkumerkin kehotteen ympärille tai korottamalla vain desimaaliluvun määrää. `(brown eyes:1.4)` lisäisi ruskeiden silmien arvoa noin 40 prosentilla, kun `(brown eyes:0.6)` vähentäisi arvoa noin 40 prosentilla. Vastaavasti hakasulkeita

käytetään päinvastaiseen tarkoitukseen eli alentamaan kehotteen painoarvoa. (AI Demos, 2023; Andrew, 2023a; Ma, 2023.)

AUTOMATIC1111:n Stable Diffusion käyttöliittymässä voi myös kätevästi pikanäppäimillä lisätä kehotteen painostusta eli sulkumerkit kehotteen ympärille. Ensin maalataan teksti, sitten painetaan ctrl-näppäintä ja nuolinäppäimiä ylös tai alas.

Loppujen lopuksi, promptaukseen liittyy paljon pelkkää testailua, kokeilua ja harjoittelua. Vain harjoittelulla käyttäjä voi päästä tulokseen, johon voi lopulta olla itse tyytyväinen.

10 TEKIJÄNOIKEUDET JA EETTISYYS

Tässä luvussa käydään vielä läpi tekijänoikeuslakia, kerrotaan tekoälykuva-generaation eettisyydestä ja muistutetaan käyttäjän omasta vastuusta kuvien tuottamisessa.

10.1 Rikotaanko tekijänoikeuksia?

Tekijänoikeuslaissa (TOL 404/1961, 1 luku 1§) määritellään, että tekijänoikeus koskee henkilön kaikkia kirjallisia tai taiteellisia teoksia. Tämä antaa tekijälle oikeuden kontrolloida teoksensa käyttöä ja kopioimista.

Lain 4 § käsittelee teosten muuntelua ja muokkaamista. Jos joku teosta muokkaa, on hänellä tekijänoikeus tähän muokattuun teokseen, mutta hänellä ei ole oikeutta käyttää sitä tavalla, joka loukkaa alkuperäisen teoksen tekijänoikeutta. Kuitenkin, jos teosta on muokattu riittävästi ja on saatu aikaan uusi ja itsenäinen teos, ei tekijän tekijänoikeus enää riipu alkuperäisestä tekijänoikeudesta. (TOL 404/1961, 1 luku 4§.)

Pelkästään näiden perusteella voi päätellä, etteivät tällä hetkellä tekoälyllä tuotetut kuvat riko lakia. Tietysti, mitä enemmän generoidut kuvat muistuttavat olemassa olevia alkuperäisteoksia, sitä herkemmin se saattaa rikkoa teosten tekijänoikeuksia.

IltaSanomien artikkelissa Asianajaja Petteri Häkkäsen mukaan (Linnake, 2023) tekoälyllä tuotetuista kuvista voi jäädä tekijänoikeutta vähempi suoja käyttäjälle eli muut eivät välttämättä voi käyttää käyttäjän tekoälyllä tuotettua kuvaa kaupallisesti. Jos joku muu tuottaisi kuvan, joka olisi jotain alkuperäistä kuvaa vastaava, olisi kyseessä tahaton tekijänoikeusloukkaus ja siitä voi mahdollisesti joutua korvausvelvolliseksi. Kuitenkin Häkkäsen mukaan tämän kaltaisten tilanteiden tapahtuminen ei ole kovin todennäköistä eli syytä suureen huoleen ei ole olemassa. Tekoälyllä tuotettuja kuvia saa siis käyttää kaupalliseen toimintaan.

Tekoälykuvageneraatio on edelleen sen verran uusi ilmiö, ettei sitä varten ole vielä ennen kesää ehditty luoda uudempia tai tarkempia lakeja. Nyt viime kesänä, Euroopan parlamentilla on ollut kehitteillä säädöksiä liittyen tekoälyn turvallisuuteen ja läpinäkyvyyteen. Generatiivisten tekoälyjärjestelmien olisi noudatettava avoimuusvaatimuksia (on paljastettava, että sisältö on tekoälyn tuottamaa, mikä auttaa erottamaan niin sanotut deepfake kuvat oikeista) ja taatava turvatoimia laittoman sisällön tuottamista vastaan. (Euroopan parlamentti, 2023.)

10.2 Eettisyys

Itse Stable Diffusionia on mahdotonta käyttää eettisesti. Tämä johtuu siitä, että Stable Diffusionin perustuu LAION-5B datajoukkoon, joka sisältää 5,8 miljardia verkosta poimittua kuvaa. Koska kaikki kuvat ovat verkosta poimittuja, datan joukosta löytyy materiaalia äärioikeistojen teloituksista, äärimmäisestä väkivallasta, suostumuksettomasta pornografiasta, rasismista ja haitallisista yhteiskunnan stereotyyppioista. (Maiberg & Xiang, 2022.)

Taiteilija ja muusikko Holly Herndon kehitti sivuston nimeltä Have I Been Trained, joka antoi ihmisille mahdollisuuden tehdä hakuja avoimen lähdekoodin LAION-5B-tietokannasta. Sivuston tarkoituksena on auttaa taiteilijoita löytämään, sisältyykö heidän työnsä datajoukkoihin, joilla tekoälymallit ovat koulutettu. Ainakin teoriassa ihmiset voivat vaatia omien kuviensa poistamista datasta. (Maiberg & Xiang, 2022)

Hyvä esimerkki on digitaiteilija Greg Rutkowski, joka tuli laajasti tunnetuksi tekoälytaiteen yhteisössä. Useat käyttäjät generoivat kuvia matkimalla Gregin taiteellista tyyliä. Greg ei tosin pitänyt tästä ja tämän myötä hän kääntyi tekoälytaidetta vastaan. Lopulta Stability AI:n toimesta hänen sekä muiden digitaiteilijoiden nimiä poistettiin Stable Diffusionin datajoukosta. Kyseinen toimenpide ei kuitenkaan estänyt ihmisiä kehittämästä Gregin tyyliä mukailevaa mallia. Tämän mallin voi yhä löytää CivitAI:n verkkosivuilta. (Lanz, 2023.)

10.3 Käyttäjän oma vastuu

Sääntöjä ei ole tehty kuvien tuottamiseen. Tässä silti omia sekä verkosta poimiani vinkkejä, jotka voivat tehdä kuvien tuottamisesta reilumpaa jokaiselle.

Jos käyttäjällä on omaa graafista taitoa, voi hän etsiä inspiraatiota tekoälyllä, mutta luoda lopulta kuvan itse tai muokata generoitua kuvaa siten, että tekee siihen edes pieniä muutoksia, jättäen käytännössä oman kädenjälkensä työhön.

Vältä käyttämästä kehotteissa nykyajan taiteilijoiden nimiä tai sellaisten taiteilijoiden nimiä, joiden kuolemasta ei ole vielä paljonkaan aikaa. Tekijänoikeuslain mukaan tekijänoikeuden suoja-aika on 70 vuotta tekijän kuolemasta. Tuntemattoman tekijän teoksessa 70 vuoden suoja-aika alkaa siitä hetkestä, kun teos on julkistettu tai luotu. (TOL 404/1961, 4 luku 43 §; TOL 404/1961, 4 luku 44 §.) Älä siis myöskään käytä kehotteita, joissa viitataan johonkin taiteilijoiden suosimaan sivustoon. Verkossa voi helposti törmätä muiden käyttävän kehotteita, kuten ”trending in artstation” tai ”trending in devianart”.

Näiden lisäksi, ilmiselvästi, älä käytä ja generoi kuvia ihmisistä, jotka eivät ole antaneet lupaa kuviensa käyttöön.

Jos mieltä kuitenkin vaivaa kuvien tuottaminen ja valmiiden mallien opetusdatan eettisyys, on myös erittäin mahdollista kouluttaa tekoäly malli vain omilla valokuvilla ja käyttää tällaista mallia kuvien tuottamiseen. Toinen mahdollisuus on käyttää mallia, kuten CommonCanvas, joka on opetettu Creative Commons -lisensioituilla kuvilla.

11 LOPUKSI

Tämän työn aikana tutustuttiin yleisesti siihen, kuinka tekoälykuvageneraattorit toimivat ja pystyvät tuottamaan pelkästä tekstistä kuvia. Työssä esiteltiin Stable Diffusion ja asennettiin tietokoneelle paikallisesti AUTOMATIC1111 nimisen Github käyttäjän verkkokäyttöliittymä. Käyttöliittymän käyttöön lisättiin oma personoitu kuvamalli, joka opetettiin Googlen Colabissa DreamBooth -mallilla. Kaiken lisäksi työssä käytiin tarkemmin läpi käyttöliittymän toiminnot ja yleiset kehotetekniikat, joiden avulla pääsee alkuun kuvien tuottamisessa. Lopuksi pohdittiin tekoälytaiteen tekijänoikeuksia ja eettisyyttä. Selvisi, että tällä hetkellä kuvien tuottaminen tekoälyllä ei riko lakia. Ainakin niin kauan, kun kuvat eivät ole lähes samankaltaisia kenenkään toisen teoksen kanssa. Tekoälyllä generoituja kuvia saa käyttää kaupallisesti, mutta oikeus kuvaan voi olla normaalia heikompi.

Opinnäytetyötä toteuttaessa ymmärsin käytännössä, että jo hallussani oleva tieto on ollut vasta pintaraapaisu kaikkeen. Aihe on niin laaja, että siinä olisi voinut vielä syvällisemminkin perehtyä mihin tahansa työssä läpikäytyyn osuuteen. Tässä työssä käsittelin aihetta vasta yleisellä tasolla.

Opin, että erittäin kauniiden kuvien toteuttaminen ei ole niin yksinkertaista, miltä se voi vaikuttaa. Tulokset voivat vaihdella riippuen siitä, mitä mallia käyttää ja millaisen kehotteen antaa. Esimerkiksi pelkällä Stable Diffusionin 1.5 versiolla ei kovin erikoisia tuotoksia kykene luomaan, ellei osaa käyttää juuri oikeanlaisia kehoitteita. Tästä syystä, yksi itsensä kehittämisen vinkkinäni on tarkastella muiden kehoitteita ja kuvia, sillä niistä voi saada omaan käyttöön hyviä ideoita sekä tekniikoita.

Tällä hetkellä olen melko myönteinen tekoälyn käyttöön kuvageneraatiossa ja taiteessa. Kuvien tuottaminen tekoälyllä tarjoaa inspiraatiota ja auttaa uusien näkökulmien löytämisessä. Ihmiset, joilla ei ole taitoa tai on fyysisistä rajoitteista johtuva kykenemättömyys tuottaa taidetta, voivat tehdä sen nyt vapaasti tyylillä kuin tyylillä. Kun viime vuoden syksyllä kuulin ensimmäisen kerran AI-kuvageneraatiosta, innostuin, sillä nyt käsissäni oli mahdollisuus tuottaa kuvia ystävistäni, vaikka syntymäpäiväkorteiksi.

On kuitenkin myös tärkeää huomioida kuvageneraation huonotkin puolet. Tekoälykuvageneraatio tarjoaa mahdollisuuden käyttää toisten taiteilijoiden töitä ja taidetyyliä ilman heidän lupaansa. Taiteeseen on aina kuitenkin liittynyt muiden matkiminen ja kopiointi, joten koen aiheen hankalaksi kiistelyn aiheeksi. Ei ole oikeaa eikä väärää vastausta. Eettisestä näkökulmasta tuntuu epärealistalta, että toisten taidetyyliä voi matkia tuosta vain. Joidenkin tekoälykuvageneraattoreiden käyttöön vaaditaan maksua, ja kuvat, joita tekoälymallin opetuksessa on käytetty, voivat kuulua jollekin toiselle. Tietysti, rahalla maksetaan kulut, jotka menevät yhtiöiden palvelimilla toimivaan kuvagenerointiin. Tilanne on hankala. Kuvageneraatio voi lisäksi olla väärissä käsissä vahingoittavaa, sillä sitä voidaan helposti väärin käyttää. Tekoälyllä luotujen videoiden ja kuvien tuottaminen ja levittäminen on nyt helppoa. Julkisuuden henkilöitä voidaan asettaa valheellisiin tilanteisiin. Maaliskuussa verkossa levisi kuvia Donald Trumpin kaoottiselta vaikuttavasta pidätyksestä, mutta selvisikin, että kuvat olivatkin vain tekoälyllä tuotettuja. Jouduin itsekin erikseen etsimään, olivatko kuvat totta vai huijausta.

Tulevaisuudessa huijausten määrää voisi heikentää Euroopan parlamentin kehitteillä oleva lainsäädäntö. Vielä on kuitenkin epävarmaa, kuinka hyvin lainsäädäntöä pystytään edes noudattamaan. Hyvin mahdollisesti tällaisellekin laille löytyy oma kiertotiensä.

Mielestäni tästä opinnäytetyön aihealueesta voi vielä tehdä jatkotutkimuksia. Kuvien generointi ei rajoitu vain Stable Diffusioniin, sillä sen lisäksi on olemassa muitakin, maksullisia ja maksuttomia, perehtymisen arvoisia generaatoreita, kuten DALL-E 2 ja Midjourney. Vaihtoehtoisesti AUTOMATIC1111 käyttöliittymän lisäksi löytyy muitakin käyttöliittymiä, kuten ComfyUI ja InvokeAI. Työn kirjoittamishetkellä johdannossa mainitut text-to-speech ja text-to-video tekniikatkin kehittyvät yhä nopeampaa tahtia, joten nekin tulevat olemaan tutkimuskohteina erittäin arvokkaita.

LÄHTEET

AI Demos. (3.11.2023). What are (parentheses) and [brackets] in Stable Diffusion. Haettu 22.10.2023 osoitteesta <https://www.aidemos.info/what-are-parentheses-and-brackets-in-stable-diffusion/>

AI Network. (12.4.2022). DALL·E 2: meaning, limitations, and solutions. Haettu 12.3.2023 osoitteesta <https://medium.com/ai-network/dall-e-2-meaning-limitations-and-solutions-a988c87ddeae>

All Your Tech AI. (3.3.2023). Create AI Images Of Yourself for Free: Dream-Booth & Stable Diffusion [video]. Youtube. https://www.youtube.com/watch?v=q7wfFMIcvsU&t=137s&ab_channel=AllYourTechAI

All Your Tech AI. (18.4.2023). How To Install Stable Diffusion Models: InvokeAI [video]. Youtube. https://www.youtube.com/watch?v=8NhSXMCApBg&t=53s&ab_channel=AllYourTechAI

Amir. (15.3.2023). Web UIs for Stable Diffusion. Haettu 23.9.2023 osoitteesta <https://medium.com/aiartdev/web-uis-for-stable-diffusion-e965ae030fcf>

Andrew. (2.3.2023a). How does Stable Diffusion work? Haettu 15.3.2023 osoitteesta <https://stable-diffusion-art.com/how-stable-diffusion-work/>

Andrew. (17.8.2023b). Stable Diffusion prompt: a definitive guide. Haettu 11.9.2023 osoitteesta https://stable-diffusion-art.com/prompt-guide/#Keyword_weight

Andrew. (28.8.2023c). Stable Diffusion WebUI AUTOMATIC1111: A Beginner's Guide. Haettu 10.10.2023 osoitteesta <https://stable-diffusion-art.com/automatic1111/#Seed>

Andrew. (25.9.2023d). How to come up with good prompts for Stable Diffusion. Haettu 22.10.2023 osoitteesta https://stable-diffusion-art.com/how-to-come-up-with-good-prompts-for-ai-image-generation/#Some_good_keywords_for_you

anntartica. The Ultimate Stable Diffusion Prompt Guide. PromptHero. Haettu 10.9.2023 osoitteesta <https://prompthero.com/stable-diffusion-prompt-guide>

AssemblyAI. (15.4.2022). How does DALL-E 2 actually work? [video]. Youtube. https://www.youtube.com/watch?v=F1X4fHzF4mQ&ab_channel=AssemblyAI

AUTOMATIC1111. (n.d.-a) AUTOMATIC1111/stable-diffusion-webui: Stable Diffusion web UI. Haettu 14.10.2023 osoitteesta <https://github.com/AUTOMATIC1111/stable-diffusion-webui>

AUTOMATIC1111. (n.d.-b) Install and Run on NVidia GPUs. Haettu 15.10.2023 osoitteesta <https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Install-and-Run-on-NVidia-GPUs>

AUTOMATIC1111. (n.d.-c) Troubleshooting · AUTOMATIC1111/stable-diffusion-webui Wiki · GitHub. Haettu 14.10.2023 osoitteesta <https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Troubleshooting>

Chen, J. (21.9.2022). What Is a Neural Network? Investopedia. Haettu 3.12.2022 osoitteesta <https://www.investopedia.com/terms/n/neural-network.asp>

CivitAI. (n.d.) What the heck is Civitai? CivitAI. Haettu 10.10.2023 osoitteesta <https://civitai.com/content/guides/what-is-civitai?ref=404media.co>

Remarkable-Llama616. (13.1.2023). You don't need to give the write access token. It will work perfectly fine with a read only one. The tokens are meant for accessing hugging face. So the read only token will allow you download models like SD1.5 from the source. And then the write tokens is meant for operations where you take your custom model and then write it back to a huggingface depository. If you have no plans on writing back then there's no need for write access. [Kommentti keskustelupalstan kirjoitukseen When using Hugging Face access token in DreamBooth Stable Diffusion, what data do we share? : r/StableDiffusion]. Reddit. Haettu 20.10.2023 osoitteesta <https://www.reddit.com/r/StableDiffusion/comments/10an0av/comment/j45cq1n/>

Cunliffe, J. (17.10.2022a). Create Art From Your Face With AI For Free Part 2 | No Google Colab | No Code | Automatic1111 Guide [video]. Youtube. https://www.youtube.com/watch?v=P1dfwViVOIU&t=69s&ab_channel=JAMESCUNLIFFE

Cunliffe, J. (7.11.2022b). Create Art From Your Face For Free Part 3 | Dreambooth Tutorial | Automatic1111 | Stable Diffusion [video]. Youtube. https://www.youtube.com/watch?v=MT8HhBSMaWM&t=124s&ab_channel=JAMESCUNLIFFE

EdXD. (5.10.2022). How to Use DreamBooth to Fine-Tune Stable Diffusion (Colab). Haettu 17.10.2023 osoitteesta <https://bytexd.com/how-to-use-dreambooth-to-fine-tune-stable-diffusion-colab/>

Ekman, J. (9.6.2023). Beginners Guide to AUTOMATIC1111 WebUI - Tips & Tricks. Haettu 10.10.2023 osoitteesta <https://okuha.com/guide-to-automatic1111-webui/>

Euroopan parlamentti. (14.6.2023). MEPs ready to negotiate first-ever rules for safe and transparent AI. Euroopan parlamentti. Haettu 23.10.2023 osoitteesta <https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai>

Foong, N. W. (15.11.2022). How to Fine-tune Stable Diffusion using Dreambooth. Haettu 20.10.2023 osoitteesta <https://towardsdatascience.com/how-to-fine-tune-stable-diffusion-using-dreambooth-dfa6694524ae>

Frankenfield, J. (6.7.2022). Artificial Intelligence: What It Is and How It Is Used. Investopedia. Haettu 25.11.2022 osoitteesta <https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp>

Google. Colaboratory. (n.d.). Google. Haettu 3.10.2023 osoitteesta <https://research.google.com/colaboratory/faq.html>

Hugging Face. (n.d.) DreamBooth. Haettu 3.10.2023 osoitteesta <https://huggingface.co/docs/diffusers/training/dreambooth>

IBM. (3.6.2020). What is Artificial Intelligence (AI)? IBM. Haettu 26.11.2022 osoitteesta https://www.ibm.com/cloud/learn/what-is-artificial-intelligence#toc-deep-learn-md_Q_Of3

Kafke. (19.11.2022). Noob's Guide to Using Automatic1111's WebUI : r/StableDiffusion. Reddit. Haettu 21.10.2023 osoitteesta https://www.reddit.com/r/StableDiffusion/comments/yz2tbo/noobs_guide_to_using_automatic1111s_webui/

Kamph, S. (28.5.2023). How to Install Stable Diffusion - automatic1111 [video]. Youtube. https://www.youtube.com/watch?v=kqXpAKVQDNU&list=PLhwn-HtyeJqp wz4gdyPV4EwOrb1sexb4KB&index=23&ab_channel=Sebastian-Kamph

Klementti, A. (24.2.2022). Tekoälyllä kohti ketterää digisalamointia. Roblogi. <https://blogit.metropolia.fi/robologi/2022/02/24/tekoalylla-kohti-ketteraa-digisalamointia/>

Lanz, J. A. (30.6.2023). Greg Rutkowski Was Removed From Stable Diffusion, But AI Artists Brought Him Back. Haettu 15.10.2023 osoitteesta <https://decrypt.co/150575/greg-rutkowski-removed-from-stable-diffusion-but-brought-back-by-ai-artists>

Lewis, N. (24.10.2022). How to Run Stable Diffusion on Your PC to Generate AI Images. How-To Geek. Haettu 20.3.2023 osoitteesta <https://www.howtogeek.com/830179/how-to-run-stable-diffusion-on-your-pc-to-generate-ai-images/>

Linnake, T. (17.6.2023). Saanko käyttää tekoälyllä tekemääni tekstiä tai kuvaa? Luvassa on hirvittävä sotku – näin vastaavat asiantuntijat. Iltasanomat. Haettu 11.11.2023 osoitteesta <https://www.is.fi/digitoday/art-2000009637849.html>

Logunova, I. (17.11.2022). Stable Diffusion: Generating Images Out of Thin Air. Haettu 14.3.2023 osoitteesta <https://serokell.io/blog/stable-diffusion>

Ma, Y. (14.9.2023). Stable Diffusion Prompt Guide for Beginners. Haettu 11.10.2023 osoitteesta <https://aituts.com/stable-diffusion-prompts/>

Maiberg, E. & Xiang, C. (21.9.2022). ISIS Executions and Non-Consensual Porn Are Powering AI Art. Vice. Haettu 23.10.2023 osoitteesta <https://www.vice.com/en/article/93ad75/isis-executions-and-non-consensual-porn-are-powering-ai-art>

Merilehto, A. (2018). Tekoäly : matkaopas johtajalle. Alma Talent. Haettu 26.11.2022 osoitteesta [https://bisneskirjasto-almatalent-fi.lil-lukka.samk.fi/teos/GADBDXDTEB#/kohta:TEKO\(\(c4\)LY\(\(20\)/piste:b0](https://bisneskirjasto-almatalent-fi.lil-lukka.samk.fi/teos/GADBDXDTEB#/kohta:TEKO((c4)LY((20)/piste:b0)

Monge, J. (7.6.2023). How to Write the Best Stable Diffusion Prompts in 2023. Haettu 13.9.2023 osoitteesta <https://hackr.io/blog/stable-diffusion-prompts>

Render Realm. (2023). Picture Yourself with Stable Diffusion [video]. Youtube. Haettu 17.10.2023 osoitteesta https://www.youtube.com/watch?v=ksG_ZjEKuLA&t=55s&ab_channel=RenderRealm

Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. (2021). High-Resolution Image Synthesis with Latent Diffusion Models. Teoksessa Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Vsk. 2022-June, s. 10674–10685). IEEE Computer Society. <https://doi.org/10.48550/arxiv.2112.10752>

SAS. (n.d.) Artificial Intelligence (AI): What it is and why it matters. SAS. Haettu 26.11.2022 osoitteesta https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html

SECourses. (16.12.2022). Transform Your Selfie into a Stunning AI Avatar with Stable Diffusion - Better than Lensa for Free [video]. Youtube. Haettu 16.10.2023 osoitteesta https://www.youtube.com/watch?v=mnCY8uM7E50&t=472s&ab_channel=SECourses

Spektor, I. (31.8.2022). From DALL·E to Stable Diffusion: how do text-to-image generation models work? Haettu 14.3.2023 osoitteesta <https://tryo-labs.com/blog/2022/08/31/from-dalle-to-stable-diffusion>

Stability AI. (10.8.2022). Stable Diffusion launch announcement. Haettu 19.9.2023 osoitteesta <https://stability.ai/blog/stable-diffusion-announcement>

Stability AI. (26.7.2023). Announcing SDXL 1.0. Haettu 22.10.2023 osoitteesta <https://stability.ai/blog/stable-diffusion-sdxl-1-announcement>

Syntho. (n.d.) Mitä on synteettinen data? Haettu 14.3.2023 osoitteesta <https://www.syntho.ai/fi/what-is-synthetic-data/>

Tekijänoikeuslaki 404/1961. Haettu 7.11.2023 osoitteesta <https://www.finlex.fi/fi/laki/ajantasa/1961/19610404>

Tiu, E. (9.2.2020). Understanding Latent Space in Machine Learning. Haettu 15.3.2023 osoitteesta <https://towardsdatascience.com/understanding-latent-space-in-machine-learning-de5a7c687d8d>

Underfitted. (14.10.2022). Generating Images From Text. Stable Diffusion, Explained [video]. Youtube. Haettu 21.2.2023 osoitteesta https://www.youtube.com/watch?v=2o0x1hJdcVc&list=PLhwn-HtyeJqp wz4gdyPV4EwOrb1sexb4KB&index=20&ab_channel=Underfitted

Vahdat, A. & Kreis, K. (26.4.2022). Improving Diffusion Models as an Alternative To GANs, Part 1. NVIDIA Technical Blog. Haettu 15.3.2023 osoitteesta <https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

Vox. (2022). The text-to-image revolution, explained [video]. Youtube. Haettu 15.1.2023 osoitteesta https://www.youtube.com/watch?v=SVcsD-DABEkM&ab_channel=Vox

W3Schools. Python PIP. W3Schools. Haettu 17.10.2023 osoitteesta https://www.w3schools.com/python/python_pip.asp

Wayner, P. (31.5.2022). What is an embedding for AI? Haettu 13.3.2023 osoitteesta <https://venturebeat.com/ai/what-is-an-embedding-for-ai/>