

Kalle Karppinen

**FOTOREALISTINEN TUOTEKUVASTAMINEN
LATENTEILLA DIFFUUSIOMALLEILLA**



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2023

TIIVISTELMÄ

Karppinen, Kalle

Fotorealistinen tuotekuvastaminen latenteilla diffuusiomalleilla

Jyväskylä: Jyväskylän yliopisto, 2023, 89 s.

Tietojärjestelmätiede, kandidaatintyö

Ohjaajat: Mehtälä, Saana; Seppänen, Ville

Tämä kandidaatintyö tutki latenteilla diffuusiomalleilla (LDM) muodostettujen kuvien soveltuvuutta fotorealistiseen tuotekuvastamiseen ja ihmisten näkemyksiä LDM:llä generoiduista tuotekuvista. Lisäksi tutkimuksessa pohdittiin synteettiseen kuvangenerointiin liittyviä eettisiä näkökulmia UNESCO:n ja EC-COLA-viitekehyksien perusteella. LDM on koneoppimismalli, mikä sisältää pakattua tietoa kuvan keskeisistä piirteistä, jota voidaan myöhemmin käyttää uusien kuvien muodostamisessa. Fotorealismilla tarkoitetaan kuvan pyrkimystä olla oikean valokuvan näköinen. Työssä kerättiin tietoa tuotekuvastamisesta erityisesti markkinointiviestinnällisiä kanavia varten, ja luotiin kirjallisuuskatsauksen muodossa kuvauksia kirjoitushetkellä laadukkaimpia tuloksia tuottavista tekniikoista ja toimintamalleista. Tuotetun sisällön laadullisen onnistumisen ja tekniikoiden käyttämisen eettisten seikkojen tarkastelu oli myös olennainen osa tutkimusta. Kirjallisuuskatsauksen lisäksi tutkimuksessa kerättiin dataa mielipidekyselyn muodossa, jolla selvitettiin ihmisten toiveita synteettisen sisällön käyttämisestä markkinointiviestinnällisenä välineenä. Kyselytutkimuksella pyrittiin sitomaan tutkimusaiheen teoriaa todellisuuteen, ja todistamaan kuvangeneroimisen toimivuutta ja käyttökelpoisuutta. Kyselyä ja tätä tutkielmaa varten luodut kuvat valmistettiin kuvaamalla reaali maailman kohteesta neljäntoista kuvan kuvatietokanta, joka siirto-opetettiin olemassa olevan latentin diffuusiomallin Stable Diffusion v.2.1 non-EMA:n alkuperäisten arvojen yhteyteen hyödyntäen suurille kielimalleille kehitettyä LoRA (Low Ranking Adaptation of Large Language Models) menetelmää. Kuvattavaksi kohteeksi valikoitui symmetrinen pinnatuolimalli, joka kuvattiin neljässä eri värissä. Kandidaatintyön kyselytutkimus toteutettiin Google Forms mielipidekyselynä, jossa vastaajat saivat kertoa mielipiteitään synteettisen sisällön käyttämisestä markkinointiviestinnän välineenä, ja arvioida kuvien fotorealistista uskottavuutta Likert-asteikolla 1–5 ja vapaavalintaisten lisäkysymysten muodossa. Kyselytutkimustulokset osoittivat tekniikoiden omaavan suuren potentiaalin fotorealistisessa sisällöntuottamisessa, ja niiden käyttämisen olevan jossain tilanteissa markkinointiviestinnällisesti kannattavaa. Tutkimustulosten perusteella kuvasynteesillä voidaan saavuttaa jo ainakin suhteellinen fotorealismitaso. Toisaalta ne myös kertovat ihmisten mielipiteistä siitä, miten tekoälytekniikoiden käyttäminen tulisi tehdä avoimesti ja aihepiirin eettisiä аспекteja miettien, jottei kuluttajia johdettaisi harhaan.

Avainsanat: Kuvangeneroiminen, latentit diffuusiomallit, fotorealismi, generatiiviset koneoppimismallit, kuvamanipulaatio, markkinointi, tekoälyn etiikka.

ABSTRACT

Karppinen, Kalle

Photorealistic product imaging using latent diffusion models

Jyväskylä: University of Jyväskylä, 2023, 89 pp.

Information Systems, Bachelor's Thesis

Supervisors: Mehtälä, Saana; Seppänen, Ville

This bachelor's thesis investigated the suitability of images generated by Latent Diffusion Models (LDM) for photorealistic product representation and people's perceptions of LDM-generated product images. Additionally, the study explored ethical perspectives related to synthetic image generation based on the UNESCO and ECCOLA frameworks. LDM is a machine learning model that encapsulates compressed information about the key features of an image, which can later be used to generate new images. Photorealism refers to the image's attempt to resemble a real photograph. The thesis collected information on product representation, especially for marketing communication channels, and created literature reviews describing the techniques and methodologies that produced the highest quality results at the time of writing. Examining the qualitative success of the generated content and the ethical considerations of using the techniques was also an essential part of the research. In addition to the literature review, data was gathered through an opinion survey to explore people's desires regarding the use of synthetic content as a marketing communication tool. The survey aimed to bridge the gap between theoretical aspects of the research topic and reality, as well as to validate the effectiveness and usability of image generation. Images for the survey and this thesis were created by capturing a real-world object, a symmetric upholstered chair model, in fourteen different image variations. The image dataset was then transfer-trained into the existing Stable Diffusion v.2.1 non-EMA latent diffusion model using the LoRA (Low Ranking Adaptation of Large Language Models) method developed for large-scale language models. The bachelor's thesis survey was conducted using Google Forms, where respondents could express their opinions on using synthetic content as a marketing communication tool and evaluate the photorealistic credibility of the images on a Likert scale of 1-5, as well as through additional open-ended questions. The survey results indicated that the techniques have significant potential in photorealistic content generation, and their use can be beneficial in certain marketing communication contexts. Based on the research findings, relative photorealism levels can already be achieved through image synthesis. However, the results also highlight people's opinions on how the use of AI technologies should be conducted transparently and with consideration for ethical aspects, to avoid misleading consumers.

Keywords: Image generation, latent diffusion models, photorealism, generative machine learning models, image manipulation, marketing, ethics of artificial intelligence.

KIITOKSET

Erityiskiitokset työn ohjaajalle Saana Mehtälälle, joka oli suureksi avuksi ja tukena kandidaatintyöni tekemisessä. Hän antoi työstä arvokasta palautetta, neuvoja ja kannustusta, ja auttoi hiomaan työtäni koko tutkimusprojektin ajan huolellisesti ja suoraviivaisesti.

LISENSSIT

Tutkimuksen diffuusiomalli on siirto-opetettu LoRA-menetelmällä käyttäen Stable Diffusion v2.1 -mallia, joka on julkaistu [Creative ML OpenRAIL-M](#) -lisenssillä. Tämä mahdollistaa kaupallisen ja ei-kaupallisen käytön. Kuvien keräämiseen saatu suostumus Maskun Kalustetalo Oy:n Jyväskylän Seppälän toimipisteestä. Maskun Kalustetalo Oy on kuvangeneroimiseen käytetyn tuotteen yksinmyyntioikeuden haltija. Tutkimuskuvat on lisensoitu [CC BY-NC-ND](#) -lisenssillä.

KUVIOT

KUVIO 1	Latentilla diffuusiomallilla generoituja tuotekuvia. Kuvat generoitu valokuvasta muodostetun Canny-reunakartan ja tekstikehotteen perusteella. Kuvien leveys ja korkeus 1024 pikseliä (px) per kuva. 12
KUVIO 2	Negatiiviset upotukset päällekkäin. Kehote: <i>"A pink spoke chair, in a dark forest at night"</i> . 1024px x 1024px / kuva..... 15
KUVIO 3	Esimerkkikuva koulutustietokannasta. Kuvaus: <i>"A white Pinja spoke chair, in a ware house with concrete floor, in front of a garage door"</i> . 768px x 768px..... 17
KUVIO 4	Esimerkkikuva koulutustietokannasta. Kuvaus: <i>"A black Pinja spoke chair, in front of an oaken wooden wall, light grey floor"</i> . 768px x 768px. 17
KUVIO 5	Koulutuksen yhteydessä generoidut esimerkkikuvat (<i>sanity sample</i>). Koherenssi kasvaa koulutusaskelten myötä. 768px x 768px. 21
KUVIO 6	Pelkän tekstikehotteen perusteella generoituja kuvia (<i>Text-to-Image</i>). Kehote: <i>"Red Pinja spoke chair, on a beach, with high cliffs and stormy sea in the background"</i> . 768px x 768px..... 22
KUVIO 7	Tekstikehotteen ja taulukon ensimmäisen kuvan perusteella generoituja kuvia (<i>Image-to-Image</i>). Kehote: <i>"Red Pinja spoke chair, on a beach, with high cliffs and stormy sea in the background"</i> . 768px x 768px. Denoising strength: 0.75. 22
KUVIO 8	Canny-reunakartan ja tekstikehotteen perusteella generoituja kuvia (<i>Text-to-Image/ControlNet</i>). Kehote: <i>"Red Pinja spoke chair, on a beach, with high cliffs and stormy sea in the background"</i> . 1024px x 1024px / kuva. 23
KUVIO 9	Canny-reunakartan, tekstikehotteen, ja taulukon ensimmäisen kuvan perusteella generoituja kuvia (<i>Image-to-Image/ControlNet</i>). Kehote: <i>"Red Pinja spoke chair, on a beach, with high cliffs and stormy sea in the background"</i> . 1024px x 1024px / kuva. Denoising strength: 0.75. 23
KUVIO 10	Tuotteen taustan poistaminen ja Canny-reunakartan luominen. Rajatun kuvan koko oli 1024px x 1024px (tekohetkellä ControlNet-lisäosan maksimaalinen kehyskoko). 24
KUVIO 11	Tekstikehotteen ja Canny-reunakartan perusteella generoitu kuva. Kehote: <i>"A white Pinja spoke chair, in front of grass plains and distant mountains, during golden hour, sunset"</i> . 1024px x 1024px. 25
KUVIO 12	Mielipiteet tuotekuvastamisesta. 35
KUVIO 13	Mielipiteet synteettisesti generoidusta sisällöstä. 36
KUVIO 14	<i>Image-to-Image</i> kuvien uskottavuus uusilla tuotesävyillä. 37
KUVIO 15	<i>Text-to-Image</i> kuvien uskottavuus olemassa olevilla tuotesävyillä ja tausta-asetelmilla. 37

KUVIO 16	<i>Text-to-Image</i> kuvien uskottavuus keksityn tuotesävyyn perusteella. 38
KUVIO 17	<i>Image-to-Image</i> kuvien uskottavuus olemassa olevaan tuotesävyyn perustuen. 38
KUVIO 18	Superresoluutio- ja Inpainting-kuvien uskottavuus. 39
KUVIO 19	Näytteenottoaskelia: 0, 1 & 5. Kehote: " <i>Green Pinja spoke chair, during sunset, golden hour, snow on the ground, winter forest in the background, clouds in the sky</i> ". 1024px x 1024px / kuva. Ensimmäisessä kuvassa on kohinaa, mistä kuvan generoimisprosessi alkaa. Näytteenottoaskelten kasvaessa kuvan yksityiskohtaisuus ja täsmällisyys kehotteen kanssa paranevat. 58
KUVIO 20	Näytteenottoaskelia: 10, 20 & 30. Kehote: " <i>Green Pinja spoke chair, during sunset, golden hour, snow on the ground, winter forest in the background, clouds in the sky</i> ". 1024px x 1024px / kuva. 58
KUVIO 21	Näytteenottoaskelia: 40, 50 & 60. Kehote: " <i>Green Pinja spoke chair, during sunset, golden hour, snow on the ground, winter forest in the background, clouds in the sky</i> ". 1024px x 1024px / kuva. 58
KUVIO 22	Generoitu tuotokuva 60 näytteenottoaskeleen jälkeen. Kehote: " <i>Green Pinja spoke chair, during sunset, golden hour, snow on the ground, winter forest in the background, clouds in the sky</i> ". 1024px x 1024px. CFG scale: 4. 59
KUVIO 23	Generoitu tuotokuva 60. näytteenottoaskeleen jälkeen. CFG scale: 1, 4, 7 & 14. Kehote: " <i>Green Pinja spoke chair, during sunset, golden hour, snow on the ground, winter forest in the background, clouds in the sky</i> ". 1024px x 1024px / kuva. CFG scalen kasvaessa kuvan värikylläisyys nousee, ja se on tyylieltyemmän näköinen. Se voi toimia taiteellisissa kuvissa, mutta fotorealismissa värityksen on syytä olla hillittyä. 60
KUVIO 24	Näytteenottoaskelten 30 & 60 erotus. Kuvien väliset poikkeamat näkyvät vaalealla. Erot yksityiskohtaisuudessa ovat hienoja mutta havaittavia kirkkaalla näyttöpäätteellä. 61
KUVIO 25	Generoidun kuvan yksityiskohtaisuus 30 & 60 näytteenottoaskeleen jälkeen. 61
KUVIO 26	Skaalattu kuva. Kehote: " <i>Brown, peanut, wooden, dark oak colored Pinja spoke chair, white background, concrete floor</i> ". 256px x 256px. 62
KUVIO 27	Superresoluutio (1/2). Kehote: " <i>Brown, peanut, wooden, dark oak colored Pinja spoke chair, white background, concrete floor</i> ". Upscaler 1: SwinIR 4x. Upscaler 2: ESRGAN 4x (50 % intensiteetillä). 768px x 1024px (skaalattu ja rajattu kuva, alkuperäinen 256px x 256px). 63
KUVIO 28	Superresoluutio (2/2). Kehote: " <i>Brown, peanut, wooden, dark oak colored Pinja spoke chair, white background, concrete floor</i> ". Upscaler: ESRGAN 4x. 3072px x 4096px (skaalattu ja rajattu kuva, alkuperäinen 1024px x 1024px). 64

KUVIO 29	Fourier-muunnos (DFT:n vaihekuva) ESRGAN 2x superresoluutiosta ja siitä syntyvistä artefakteista generoidussa kuvassa. Artefaktit symmetrisiä viivoja vaaka- ja pystyakseleilla.	65
KUVIO 30	Fourier-muunnos (DFT:n vaihekuva) ESRGAN 4x superresoluutiosta ja siitä syntyviä artefakteja oikeassa valokuvassa.	66
KUVIO 31	Fourier-muunnos ESRGAN 6x superresoluutio generoidussa kuvassa. Lisätty digitaalista kohinaa, ja rajattu 2048px x 2048px resoluutioon käyttäen Bilinear- interpolointia. Kuvassa yhä havaittavissa viivat vaaka- ja pystyakselilla, mutta ne eivät ole yhtä symmetrisiä, ja koneellisesti tuotetun näköisiä. Toisaalta viivat vaaka- ja pystyakselilla viestivät yhä kuvan koneellisesta luonteesta.	67
KUVIO 32	Digitaalinen kohina havainnoituna DFT:n suuruuskuvassa. Kohina ilmenee symmetrisinä viivoina, jotka kulkevat kuvassa pystysuunnassa.	68
KUVIO 33	Digitaalisen kohinan piilottaminen havainnoituna DFT:n suuruuskuvana. Kuvaan tehty 6x ESRGAN superresoluutio, johon lisätty digitaalista kohinaa, ja joka on skaalattu 2048px x 2048px resoluutioon hyödyntäen Bilinear- interpolointia.	69

TAULUKOT

TAULUKKO 1	ECCOLA-teemat ja aiheita (mukailten Vakkuri ym., 2021). .	47
------------	---	----

SISÄLLYS

TIIVISTELMÄ

ABSTRACT

KIITOKSET & LISENSIT

KUVIOT & TAULUKOT

1	JOHDANTO.....	10
2	FOTOREALISTINEN KUVANGENEROIMINEN.....	13
2.1	Kuvangeneroiminen LDM:lla	13
2.2	LDM:n ja sen upotusten valitseminen	15
2.3	Kuvasynteesin ohjelmistotyökalut	15
2.4	Koulutettavan kuvatietokannan käsittely	17
2.5	LDM:n koulutusmenetelmät	18
2.6	LDM:n LoRA-kouluttaminen.....	19
2.7	LDM:n käyttäminen	21
2.8	LDM:lla generoimisen aliprosessit ja parametrit	26
2.9	Kuvan resoluutio ja sen muunnokset	27
2.10	In- & Outpainting	28
2.11	Fotorealistinen kuvangeneroiminen suhteutettuna aiempiin fotorealistisen kuvastamisen menetelmiin.....	29
2.12	Fotorealistisesti generoidun kuvamateriaalin havaitseminen	30
3	KYSELYTUTKIMUS SYNTEETTISESTÄ TUOTEKUVASTAMISESTA.....	31
3.1	Kyselytutkimuksen kuvaus.....	31
3.2	Kyselytutkimuksen rajoitukset.....	32
3.3	Kyselytutkimustulokset	33
3.4	Kyselytutkimuksen yhteenveto	39
4	FOTOREALISTISEN KUVANGENEROIMISEN EETTISYYS.....	41
4.1	Fotorealistinen tuotekuvastamisen eettisyys markkinoinnissa	41
4.2	UNESCO:n suositus tekoälyn eettisyydestä suhteessa fotorealistiseen tuotekuvastamiseen.....	42
4.3	Fotorealististen tuotekuvien generoiminen & ECCOLA-menetelmä	47
5	POHDINTA	49
6	YHTEENVETO	52
	LÄHTEET	54
	LIITE 1 NÄYTTEENOTTOASKELEET VAIHEITTAIN	58
	LIITE 2 GENEROITU KUVA & CFG SCALEN VAIKUTUS.....	59

LIITE 3	NÄYTTEENOTTOASKELTEN VERTAILU	61
LIITE 4	GENEROITU KUVA SKAALATTUNA ILMAN SUPER- RESOLUUTIOTA.....	62
LIITE 5	PIENEN KUVAN SUPERRESOLUUTIO	63
LIITE 6	SUUREN KUVAN SUPERRESOLUUTIO	64
LIITE 7	ARTEFAKTIT GENEROIDUN KUVAN DFT-VAIHEKUVASSA	65
LIITE 8	ARTEFAKTIT OIKEAN VALOKUVAN DFT-VAIHEKUVASSA	66
LIITE 9	ARTEFAKTIT INTERPOLOIDUSSA DFT-VAIHEKUVASSA	67
LIITE 10	DIGITAALINEN KOHINA JA DFT-SUURUUSKUVA.....	68
LIITE 11	DIGITAALISEN KOHINAN PIILOTTAMINEN JA DFT- SUURUUSKUVA	69
LIITE 12	KYSELYTUTKIMUKSEN TULOKSET	70

1 JOHDANTO

Tämän kandidaatintyön tarkoituksena on tutkia latenttien diffuusiomallien (*latent diffusion models*, LDM) soveltumista tuotekuvien fotorealistiseen (*photo-realistic*) generoimiseen. Aihe on tärkeä ja ajankohtainen, sillä kuvasynteesissä on vasta hiljattain tapahtunut murros, minkä seurauksesta kuvangeneroimisesta on tullut näin suoraviivaista fotorealistiseksi luokiteltavalla tarkkuustasolla. Fotorealismilla tarkoitetaan kuvan pyrkimystä näyttää mahdollisimman paljon valokuvan kaltaiselta, generoinnilla puolestaan sitä, että kuva luodaan automaattisesti ilman ihmisen väliintuloa, ja latentti viittaa generoimiseen käytetyn mallin datasisällön piilevään muotoon. Aiheen tutkimuksesta tekee keskeistä se, että aiempaa kirjallisuutta ei juurikaan löydy liittyen synteettiseen tuotekuvastamiseen, sillä aiemmin tuotekuvastaminen on ollut mahdollista valokuvaamalla, piirtämällä tai 3D-mallintamalla. Tuotekuvat ovat keskeinen osa markkinointiviestintää, ja niiden laatu ja realismisuus voivat vaikuttaa kuluttajien mielikuviin ja ostopäätöksiin (Honkala, 2009). LDM:t ovat uusi ja lupaava tekniikka, joka mahdollistaa tuotekuvien luomisen jopa ilman fyysistä tuotetta tai valokuvausta, mikä voi tarjota sekä säästöjä että uusia mahdollisuuksia markkinoijille. Toisaalta niiden käyttäminen voi myös aiheuttaa haasteita ja riskejä kuluttajille ja yhteiskunnalle, jos niitä hyödynnetään vihamielisesti tai se toteutetaan ilman läpinäkyvyyttä (Finlayson ym. 2019).

LDM:t perustuvat kohinanpoistoon, syväoppimiseen ja Markovin ketjuihin (Rombach, ym. 2022). Niillä saavutettavaa laatu voi olla erittäin realistisia ja jopa visuaalisesti erottamatonta valokuvista. LDM:lla voidaan tuottaa korkealaatuisia ja monipuolisia kuvia erilaisista aiheista ja tyyleistä (Rombach, ym. 2022; Nichol, ym. 2021; Dhariwal & Nichol, 2021). Rombach ym. (2022) määrittävät LDM:en toimivan muuntamalla kuvan tai tekstin ensin piileväksi, eli latentiksi esitykseksi. Latentti esitys on pienempi ja tiiviimpi kuin mitä alkuperäinen data, eli tieto pakataan alempaan ulottuvuuteen. Latentti esitys voidaan palauttaa takaisin kuvaksi käyttäen dekooderia, eli käänteistä muunnosta. Lisäksi ControlNet toiminnallisuudella kuvan mittasuhteet saadaan lukittua, ja kuvia voidaan generoida myös oikeisiin valokuviin perustuen (ks. kuvio 1) tuotteen mittasuhteet lukiten.

Tässä kandidaatintyössä tutkitaan miten LDM:a voidaan hyödyntää tuotekuvastamisessa ja -markkinoinnissa, tarkemmin millaisia synteesi tuloksia niillä voidaan saavuttaa, ja mitä mieltä ihmiset ovat niiden käyttämisestä. Työ on toteutettu kirjallisuuskatsauksella ja laadullisella kyselytutkimuksella. Kyselytutkimuksessa kerättiin ihmisten mielipiteitä generoitujen kuvien fotorealistisuudesta, ja tämän lisäksi heidän ajatuksiaan synteettisten kuvien käyttämisestä markkinointiviestinnässä. Kirjallisuuskatsauksessa lähteiden valintaa on ajanut etenkin laadullisen tutkimuksen tutkimuskuvien muodostamiseen käytetyt tekniikat ja niihin liittyvä taustakirjallisuus. Koska työssä käytetyt tekniikat ovat todella uusia, ja ne ovat myös kehittyneet suurilta osin tutkimusprosessin ollessa vielä käynnissä, lähdekirjallisuus on muodostettu pitkälti [arXiv.org](https://arxiv.org)-sivuston julkaisuista. Sivusto tarjoaa kehittäjille mahdollisuuden julkaista ennakkojulkaisuja, minkä ansiosta se sisältää tietoa myös ajankohtaisimmista tekniikoista niiden ollessa yhä kehitysvaiheessa. Tutkimusaiheen teknisen luonteen takia työssä johtopäätöksiä on muodostettu paljon myös järjestelmätestien perusteella, eli testamalla eri tekniikoita, ja vertaamalla niillä saavutettavia tuloksia keskenään.

Kandidaatintyön kyselytutkimuksen tulokset osoittivat, että latentit diffuusiomallit (LDM) ovat lupaava menetelmä tuotekuvien fotorealistiseen generointiin. Kirjallisuuskatsaus osoitti, että LDM-tekniikat ovat kehittyneet nopeasti, ja niillä voidaan luoda realistisia tuotekuvia. Kyselytutkimuksessa saadut tulokset puolestaan vahvistivat tämän havainnon, kun osallistujat pitivät generoituja kuvia fotorealistisina ja näkivät niiden potentiaalin markkinointiviestinnässä. Tämä viittaa siihen, että LDM-menetelmät voivat tarjota uusia mahdollisuuksia tuotekuvien luomiseen ja markkinointiviestinnän tehostamiseen.

Tämä kandidaatintyön jatko koostuu viidestä osiosta: fotorealistinen kuvangenerointi, kyselytutkimus synteettisestä tuotekuvastamisesta, fotorealistisen kuvangeneroimisen eettisyys, pohdinta ja yhteenveto. Ensimmäisessä osiossa esitellään fotorealistisen kuvangeneroinnin perusteet ja erilaiset lähestymistavat. Osiossa käsitellään myös LDM:n ja sen upotusten valintaa, kuvasynteessin ohjelmistotyökaluja, koulutettavan kuvatietokannan käsittelyä, LDM:n koulutusmenetelmiä sekä LDM:n käyttämistä kuvan generoimiseen. Lisäksi osiossa tarkastellaan generoimisen aliprosesseja ja niihin liittyviä parametreja, kuvan resoluutiota ja sen muunnoksia sekä Inpainting- ja Outpainting-tekniikoita. Toisessa osiossa esitellään toteutettu kyselytutkimus, joka keskittyy syntetisoitujen tuotekuvien arviointiin. Osiossa kuvataan kyselytutkimuksen toteutus ja sen tarkoitus. Osion pääpaino on kyselytutkimuksen tulosten esittelyssä ja niiden yhteenvedossa. Kolmannessa osiossa tarkastelu siirretään kuvangeneroimisen eettisyyteen, ja fotorealistista tuotekuvastamista suhteutetaan markkinointiin, EC-COLA-viitekehyksiin ja UNESCO:n (2021) suositukseen tekoälyn eettisyydestä. Kappaleessa määritetään tekijöitä, joita yritysten tulisi huomioida implementoidessaan fotorealistista kuvangeneroimista tuotekuvastamiseen. Neljännessä osiossa pohditaan kuvangeneroimisen tekniikoita, ja uusia kehityssuuntia niiden käyttämisessä. Siinä pohditaan myös työn tuloksia. Viimeisessä osiossa tiivistetään lopputyön keskeiset tulokset ja johtopäätökset. Yhteenvetoon sisältyy myös lyhyt katsaus työn merkitykseen ja mahdollisiin jatkotutkimuskohteisiin. Osi-

ossa arvioidaan tutkimuksen tuloksia ja tarkastellaan niiden merkitystä. Kappalessa esitellään mahdollisia jatkotutkimusaiheita. Työn luonteen takia siinä on myös paljon kuvioita ja liitetiedostoja, missä erilaisten parametrien vaikutuksia havainnoidaan kuvataulukoilla, sekä tuodaan ilmi kysymyskohtaiset tulokset kyselytutkimuksesta. Tämän kandidaatintyön tutkimuskysymykset olivat:

1. Miten ja millaisella laadulla LDM:a voidaan käyttää tuotekuvien fotorealistiseen generoimiseen?
2. Millaisia mielipiteitä ja asenteita ihmisillä on LDM:lla generoiduista tuotekuvista ja niiden käytöstä markkinointiviestinnässä?
3. Mitä eettisiä аспекteja fotorealistisen markkinointimateriaalin generoimisessa tulisi huomioida?



KUVIO 1 Latentilla diffuusiomallilla generoituja tuotekuvia. Kuvat generoitu valokuvasta muodostetun Canny-reunakartan ja tekstikehötteen perusteella. Kuvien leveys ja korkeus 1024 pikseliä (px) per kuva.

2 FOTOREALISTINEN KUVANGENEROIMINEN

2.1 Kuvangeneroiminen LDM:lla

LDM:t ovat siirto- ja generatiivisin koneoppimismenetelmin muodostettavia todennäköisyysmalleja, jotka saavuttavat laadukkaat synteetitulosensa jakamalla kuvanmuodostusprosessin vaiheittaisiksi askeleiksi, jossa kuvaa käsitellään esikoulutettujen kohinanvaimentamiseen perustuvien autoenkoodereiden latenttitallassa (Ho, Jain & Abbeel, 2020; Rombach ym., 2022). Autoenkooderi tarkoittaa neuroverkkoa, jota käytetään tehokkaaseen tiedon koodaukseen ja purkamiseen. Se koostuu kahdesta osasta: enkooderista ja dekodeerista. Latenttien diffuusiomallien asiayhteydessä enkooderi ottaa syötteenä sille annetun alkuperäisen datan, ja muuntaa sen pienempään talletustilaan mahtuvaksi latentiksi esitykseksi. Dekodeeri puolestaan ottaa syötteenä tämän latentin esityksen sisältämät arvot, ja pyrkii syntetisoimaan sen perusteella alkuperäistä muistuttavaa dataa (Ho, Jain & Abbeel, 2020; Rombach ym., 2022; Bank & Giryes, 2020). Latentti esitys viittaa tietojenkäsittelyssä ja koneoppimisessa datan tiivistettyyn muotoon, jossa pyrkimyksenä on säilyttää alkuperäisen datan keskeisimmät piirteet ja merkitys.

Latenttien diffuusiomallien oppiminen perustuu variaatiopäätelmämenetelmään (*variational inference*), joka on tapa arvioida datan todennäköisyyttä (Ho, Jain & Abbeel, 2020). Menetelmän avulla voidaan laskea arvo, joka kertoo, kuinka hyvin koulutettu malli sopii sille annettuun dataan. Tekniikka perustuu tietokonenäköjärjestelmiin, ja käytännössä siinä kuvasta havaitaan muotoja, joita suhteutetaan toisiinsa, jolloin voidaan määrittää todennäköisyys sille mitä kuva oikein sisältää. Tämä todennäköisyys perustuu johonkin olemassa olevaan kuvatietokantaan. Prosessissa muodostettua arvoa kutsutaan evidenssin alarajaksi (*evidence lower bound*). Variaatiopäätelmämenetelmän tarkoituksena on löytää sellainen variaatiojakauma (*variational distribution*), joka maksimoi evidenssin alarajan (Bishop, 2006).

Diffuusiopohjainen kuvangeneroiminen soveltuu niin täysin uusien kuvien muodostamiseen kuin olemassa olevien tyylittelyyn ja värjäämiseen tarvittaessa

hyvinkin kontrolloidulla tavalla. Diffuusioprosessi on stokastinen prosessi, jossa kuva häivytetään asteittain kohinaksi, joka palautetaan alkuperäiseen muotoonsa käyttämällä mallin oppimaa pisteytysfunktiota. Pisteytysfunktiota voidaan käyttää myös luomaan uusia kuvia samasta latentista vektorista tai yhdistelemään eri latentteja vektoreita uusiksi kuviksi (Rombach ym., 2022; Saharia ym., 2022; Zhang & Agrawala, 2023; Bar-Tal ym., 2023). Tuotekuvia tehdessä tämä voisi tarkoittaa vaikkapa kuvien generoimista materiaaleilla, joilla tuotetta ei ole koskaan kuvattu, mutta mitkä löytyvät koulutustietokannasta muiden tuotteiden kuvista. Tämä voi olla otollista etenkin yksilöitävissä tuotteissa, missä asiakkailta saattaa olla valittavissa satoja vaihtoehtoisia verhoilumateriaaleja, sävyjä tai niiden yhdistelmiä. Latentit diffuusiomallit mahdollistavat tuotekuvien tekemisen jopa ilman oikeaa tuotetta tai kameraa.

Kuvangeneroinnin avulla korkealaatuisia tuotekuvia voidaan tuottaa rajattomasti jopa yhden tuotekuvausprosessin perusteella. Tämä poikkeaa perinteisistä kuvantamistekniikoista, missä tuotteesta tulisi aina ottaa joko uusi kuva tai muokata olemassa olevaa, jos sen ominaisuuksia kuten taustaa tai väriä halutaan muokata. Generoidut tuotekuvat voivat myös pohjautua oikeisiin valokuviin, jolloin niiden mittasuhteet ja yksityiskohtaisuus tarvittaessa säilyvät, ja muutokset tapahtuvat näennäisesti vain tuotteen sävyyn ja taustaan (kuten taulukossa 1). Latentit diffuusiomallit ovat viimeaikaisten kehitysaskeliensa myötä asemoituneet yhdeksi tehokkaimmista kuvanmuodostusprosesseista. Niiden avulla on mahdollista muodostaa jopa fotorealistisesti täysin uusia kuvia tai tehdä osittaisia muutoksia olemassa oleviin (Rombach ym., 2022; Nichol ym., 2022; Karras ym., 2022; Saharia ym., 2022).

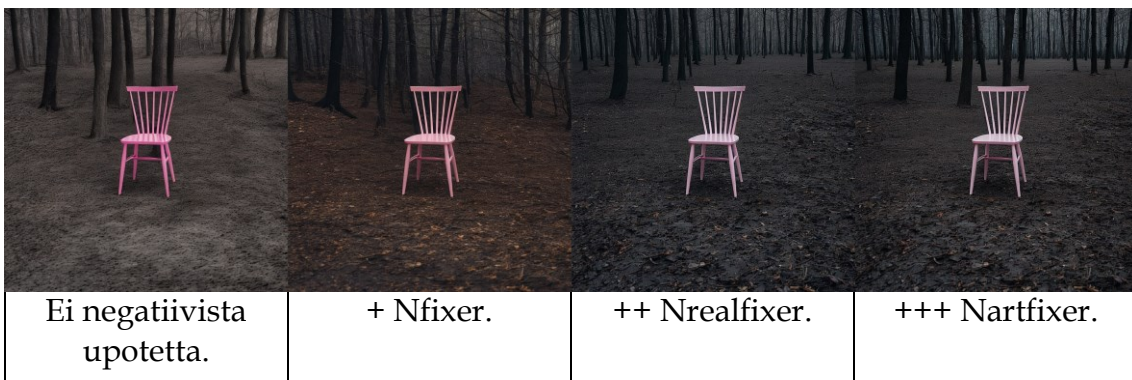
Tekniikan ei tarvitse myöskään rajoittua vain kaksiulotteisten kuvien generoimiseen, vaan sitä voidaan soveltaa esimerkiksi 3D-mallien (Anciukevičius ym., 2022; Karnewar ym., 2023) ja niiden tekstuurien tekemiseen (Richardson ym., 2023; Chen ym., 2023). 3D-malleja voidaan puolestaan käyttää uusien kuvien generoimiseen muun muassa niistä saatavien reuna- ja syvyyskarttojen avulla (Zhang & Agrawala, 2023). Tässä tutkimuksessa on kuitenkin keskitytty vain kaksiulotteisiin tuotekuviiin.

Latenttien diffuusiomallien rakenne ja muotoilu tarjoavat suoraviivaisen ohjausmekanismin kuvantamisprosessin hallitsemiseksi. Diffuusiomallien ansiosta kuvien generoiminen voi tapahtua täysin tekstikehoteisesti tai toisten kuvien ohjaamana (Rombach ym., 2022; Karras ym., 2022; Zhang & Agrawala, 2023). Siirto-oppimisen mahdollistavan tekniikkansa ansiosta kokonaisen latentin diffuusiomallin kouluttamisen sijaan olemassa olevaa mallia voidaan jatkokehittää haluttuun suuntaan hienosäätämällä sen parametrejä tai luomalla kokonaan sen ulkopuolisia painoja, näin vähentäen huomattavasti laskennallisten resurssien käyttämistä (Hu ym., 2021). Tämä on tutkielmassa keskeistä, sillä sen ansiosta valmiille diffuusiomallille on mahdollista kouluttaa kokonaan uusia konsepteja säilyttäen kuitenkin mallin alkuperäinen yksityiskohtaisuus ja kyvykkyydet.

2.2 LDM:n ja sen upotusten valitseminen

Latentin diffuusiomallin valinta tehdään tarvepohjaisesti. Erilaiset mallit soveltuvat erilaisiin tehtäviin ja niiden kirjo laaja ja kasvava. Kyselytutkimuksen tekemisen aikana parhaita tuloksia fotorealismien suhteen tarjosi Stability AI:n Stable Diffusion v2.1 LDM. Uudemmat mallit, kuten Illuminati Diffusion v1.1 (nykyisin Illuminutty¹), ovat kuitenkin haastaneet Stable Diffusionin asemaa fotorealistisesti generoitavassa laadussa. Uudemmat mallit hyödyntävät muun muassa Noise Offset -tekniikkaa, millä voidaan parantaa kuvien dynaamista aluetta ja valaistusta (Guttenberg, 2023). Toiminto on tarpeellista, sillä LDM:en meluaikatauluissa ja näytteenottoalgoritmeissa perinteisin metodein koulutettuna viimeistä aika-askelta ei pakoteta olemaan nolla signaali-kohinasuhteessa, mikä aiheuttaisi ristiriidan koulutuksen ja päättelyn välillä. Kyseinen virhe rajoittaa LDM:n kykyä tuottaa kirkkaita tai tummia kuvia (ks. kuvio 2) (Lin, Liu, Li & Yang, 2023). Tämän tutkimuksen esimerkkikuvat sisältävät 30 % Illuminati Diffusionin painoista (*weights*) eli tutkimukseen koulutettu malli on yhdistetty kyseisen mallin kanssa. Painot ovat diffuusiomallin todennäköisyysarvoja, jotka määrittävät generoitavan kohteen piirteet. LDM:ja voidaan yhdistää keskenään, eli erikseen koulutettuja ja hienosäädettyjä konsepteja voidaan siirtää niiden välillä ja käyttää yhdessä. Illuminati Diffusion toimii parhaiten yhdessä kolmen negatiivisen upotuksen (*negative embedding*) eli painoarvoja muuttavan mallin kanssa (ks. kuvio 2):

- Nfixer - parantaa yleistä koherenssia, ja on aina suositeltavaa käyttää muiden upotusten yhteydessä.
- Nrealfixer - muuttaa värit ja valon todentuntoisemmaksi.
- Nartfixer - pyrkii kehittämään kuvaa taiteellisemmalla ilmaisutyyllillä ja parantamaan sen väritystä.



KUVIO 2 Negatiiviset upotukset päällekkäin. Kehote: "A pink spoke chair, in a dark forest at night". 1024px x 1024px / kuva.

¹ <https://civitai.com/models/36152>

2.3 Kuvasynteessin ohjelmistotyökalut

Tämä tutkimus on tehty vapaassa jakelussa olevien ja avoimesti kehitettävien, usein yhteisö- tai yksilövetoisesti valmistettujen ohjelmistotyökalujen avulla. Koska tutkimuksessa käytettävät mallit ja ohjelmistot ovat olleet kirjoitushetkellä vajaan vuoden jakelussa, on ymmärrettävää, ettei asemaansa vakiinnuttaneita täysin vakaasti toimivia ohjelmistoja ole vielä saatavilla. Ohjelmiston kehitys jatkuu kuitenkin yhä, ja siihen on ajan mittaan tullut lukuisia päivityksiä ja parannuksia.

Tutkimuksen yksi keskeisimmistä työkaluista oli itse LDM, tarkemmin Stable Diffusion versio 2.1 non-EMA. Se on Stability AI:n kehittämä v-prediction-malli, joka käyttää tekstikooderina OpenCLIP-ViT/H:ta, jonka ovat kehittäneet LMU Münchenin CompVis-ryhmä ja Runway Research. Malli julkaistiin joulukuussa vuonna 2022. Diffuusiomalleissa v-prediction tarkoittaa sitä, että malli ennustaa kuvan pikselien arvoja kohinan sijaan. Tämä poikkeaa perinteisistä noise-prediction -malleista, jossa malli ennustaa kohinan määrän ja suunnan. Koneoppimisessa painot (*weights*) tarkoittavat muuttujia, jotka määritetään syötedatalle ennusteen tuottamiseksi. Ne edustavat kunkin syötemuuttujan tärkeyttä, ja voivat olla niin positiivisia kuin negatiivisiakin. EMA-mallissa jokainen termi kerrotaan eksponentiaalisella painokertoimella (*exponential moving average*), eli lasketaan eksponentiaalisesti diskontattu summa (Yazıcı, Winkler & Piliouras, 2019). Tällaisessa mallissa painokerroin pienenee eksponentiaalisesti ajan myötä, mikä tarkoittaa, että uudemmat havainnot saavat suuremman painoarvon kuin vanhemmat havainnot. EMA-malleja käytetään yleensä jatkamaan koulutusta tai parantamaan mallin vakautta, kun taas non-EMA mallit taipuvat paremmin uusien asioiden ja konseptien kouluttamiseen, sillä niissä kaikki havainnot (koulutettavat tuotekuvat kuvauksineen) saavat saman painoarvon mallia muodostaessa.

Pelkkä LDM ei kuitenkaan riitä kuvangeneroimisessa, vaan sitä käyttäkkeeseen tarvitaan erillistä ohjelmistoa. Tutkimukseen on valittu tähän tarkoitukseen AUTOMATIC1111:n Stable Diffusion selainpohjainen käyttöliittymä², joka perustuu Gradio-kirjastoon. Siihen on sisäänrakennettu lukuisia ominaisuuksia, joiden avulla kuvien generoiminen on suoraviivaista, ja mahdollista toteuttaa uusimpien teknologioiden avulla. Käyttöliittymän avoimen luonteen ansiosta siihen voi asentaa muiden kehittäjien valmistamia lisäosia, millä käytettävyyttä voidaan parantaa entiseltään. AUTOMATIC1111:n käyttöliittymän avulla voidaan suorittaa kuvangeneroimisprosessi kuvatietokannan alkukäsittelystä aina uusien kuvien generoimiseen asti. Ohjelmiston järjestelmäsuositukset ovat: vähintään 10 gigatavua vapaata tilaa kiintolevyllä, NVIDIA GPU 6 GB RAM-muistilla, PC, jossa on Windows (11, 10, 8.1 tai 8) käyttöjärjestelmä, ja Python 3.10.6 tai uudempi. Suorituskykyinen laitteisto mahdollistaa nopeamman ja laadukkaamman generoimisen.

² <https://github.com/AUTOMATIC1111/stable-diffusion-webui>

2.4 Koulutettavan kuvatietokannan käsittely

Tutkimuksen koulutustietokanta muodostui Pohjanmaan Kaluste Oy:n Pinja ruokatuolista otetuista kuvista. Kuvien ottamiseen saatiin lupa Maskun Kalustetalo Oy:n Jyväskylän toimipisteen henkilökunnalta. Tuote valikoitui tutkimukseen sen symmetrisyyden, riittävän monipuolisen värivalikoiman ja selkeiden yksityiskohtiensa ansiosta. Pystysuunnassa menevät pinnat kertovat lukumäärässään, jos generoimistulos ei ole todellisuuden mukainen. Tarvittaessa tuloksen voisi myös tarkistaa tarkalla 3d-mallilla. LDM:lle koulutettava kuvatietokanta koostui neljästätoista (14) kuvasta, jotka olivat rajattu 768 (leveys) x 768 (korkeus) pikselin (px) spatiaaliseen resoluutioon käyttäen Bilinear interpolointia. Lineaarinen interpolointi on menetelmä, jolla voidaan arvioida kaksiulotteisen funktion arvoja tunnettujen pisteiden välillä piirtämällä suora viiva näiden pisteiden yhdistämiseksi. Kaikki viivan pisteet, jotka eivät ole alkuperäisiä pisteitä, voidaan pitää interpoloituneina arvoina. Bilinear-interpolointi on menetelmä kaksiulotteisen funktion arvon arvioimiseksi suorakulmion sisäpisteissä, kun funktion arvo tunnetaan suorakulmion neljässä kulmassa. Bilinear-interpolointi perustuu lineaariseen interpolointiin, mutta siinä interpolointi suoritetaan sekä kuvan vaaka- että pysty akselilla (Wicklin, 2020). Tuotekuvia kerättiin neljästä tuotteesta neljässä eri värissä. Koska kuvastettava tuote on symmetrinen, kuvista luotiin kopiot, jotka käännettiin peilikuviksi. Näin koulutettavia kuvia oli kaiken kaikkiaan 28 kappaletta (esimerkiksi ks. kuvat 3 & 4).

Kuville luotiin yksilölliset kuvaukset analysoimalla niiden sisältöä BLIP (Bootstrapping Language Image Pre-Trained for unified vision-language understanding and generation) kuva-analyysillä. BLIP-kuvaukset antavat hyvän pohjan kuvauksiin (Li, Li, Xiong & Hoi, 2022), mutta niiden sisältöä täytyy usein



KUVIO 3 Esimerkkikuva koulutustietokannasta. Kuvaus: *“A black Pinja spoke chair, in front of an oaken wooden wall, light grey floor”*. 768px x 768px.



KUVIO 4 Esimerkkikuva koulutustietokannasta. Kuvaus: *“A white Pinja spoke chair, in a warehouse with concrete floor, in front of a garage door”*. 768px x 768px.

korjata virheiden osalta, ja lisätä siihen yksilöiviä tietoja, kuten tuotteen väri ja nimi. Jos kuvaus on vaikkapa muotoa: *“Flower field with a chair”*, voisi sen korjata seuraavasti: *“A white Pinja spoke chair, outside during summer evening, with purple Lupinus polyphyllus growing in the background”*. Kuvaukset voi tehdä myös manuaalisesti, kunhan kuvauksen ja tekstikuvauksen nimet ovat samat (esimerkiksi kuva (1).png & kuva (1).txt). Kuvauksien yksilöiminen mahdollistaa jatkossa niiden käyttämistä samoilla termeillä, ja parantaa myös kouluttamistuloksia (Li, ym. 2022; Li, Li & Ho, 022; Ruiz ym., 2022) Niiden avulla diffuusiomalli voi hyödyntää aiempaa tietämystään sille koulutettujen käsitteiden osalta generoidessaan koulutuskuvia. Lisäksi kuvauksia voidaan käyttää myöhemmin kuvien generoimisen kehotteina. Lisäämällä pilkkuja kuvauslauseisiin, voidaan kuvien tunnistetta sekoittaa (*shuffle tags*) koulutuksen aikana, ja näin vaihtaa ensimmäisen pilkun jälkeisten lauseenosien paikkaa satunnaisesti, joka voi parantaa koulutustuloksia (Malinovsky, Sailanbayev & Richtárik, 2021). Tarkkojen kuvausten ansiosta diffuusiomallille ei tarvitse myöskään erillistä regularisointikuvatietokantaa. Regularisointi (*regularisation*) on tekniikka, jolla pyritään välttämään ylisovittamista, ja parantamaan mallin yleistämiskykyä, ja sitä kuinka monta erilaista tuotetta se voi esimerkiksi tuottaa yhdestä tuoteluokasta. Regularisoinnissa mallille annettaisiin lisäkuvia, jotka eivät sisällä koulutettavaa konseptia, mutta jotka ovat muuten samankaltaisia kuin koulutusdata. Esimerkiksi jos mallille koulutetaan konseptia *“Pinja pinnatuoli”*, regularisointikuvat voivat olla kuvia muista pinna- tai ruokatuolimalleista. Keskeistä on kuitenkin se, että kuvat generoidaan alkuperäisten latentin diffuusiomallin painojen avulla käyttäen kehotteena oikeaa tuoteluokkaa. Generoimalla regularisointikuvat saadaan luotua parempi käsitys mallin kyvykkyyksistä, mutta toisaalta se on tarpeetonta, jos mallin koulutustietokanta ei ole sisältänyt näihin liittyvää tietoa laadukkaiden tuloksien taakkaamiseksi.

2.5 LDM:n koulutusmenetelmät

Jos tuotekuvia halutaan generoida jostain oikeasta tuotteesta jollain tietyllä tyyllillä (kuten vaikkapa fotorealistisesti tai öljymaalauksen tapaan), latentti diffuusiomalli tulee hienosäätää tehtävää varten siirto-oppimismenetelmin. Kirjoitushetkellä LDM:en kouluttamiseen oli kehitetty kolme keskeistä tekniikkaa, joista jokainen soveltuu hieman erilaisiin tehtäviin, ja tarjoaa omat etunsa ja haittansa:

1. DreamBooth: Muuttaa latentin diffuusiomallin rakennetta. Luo tarkkoja ja yksityiskohtaisia tuloksia (Ruiz ym., 2022). Soveltuu vain kokonaisten mallien luomiseen, eli vaatii paljon talletustilaa, mikä voi vaikeuttaa kouluttamista.
2. LoRA (*Low Rank Adaptation of Large Language Models*): Muuttaa DreamBoothin tapaan latentin diffuusiomallin rakennetta, mutta sen sijaan että päivittää siinä olemassa olevia arvoja, se luo siihen uusia tasoja, jotka

sisältävät lisätietoa koulutettavan mallin painoista (Hu ym., 2021; Paul & Cuenca, 2023). Tekniikka on tehty alun perin suurten kielimallien kouluttamiseen (Hu ym., 2021). Kielimallien parametrimäärät ovat diffuusio-pohjaiseen kuvangeneroimiseen verrattuna valtavia. Koska LoRA-tekniikka on suunniteltu toimimaan suurten parametrimäärien kanssa, se on laskennallista tehokkaampi kuin DreamBooth. LoRA-painot ovat myös pienempiä tiedostokooltaan, ja ne voidaan pitää erillään alkuperäisestä mallista (Paul & Cuenca, 2023). Tähän tutkimukseen valikoitui LoRA-koulutusmenetelmä sen alhaisten järjestelmävaatimuksien, ja lupaavien testitulosten ansiosta. Malli toimii kuin upote, mutta sen vaikutusta voidaan määrittää hienosäätää määrittämällä sen arvo kehotteessa.

3. Textual Inversion: Tarjoaa pienimmän tiedostokoon, ja toimii eri mallien välillä (Gal ym., 2022). Ei suoraan muuta mallia, vaan päivittää upotteen (*text embedding*) käyttämiä sisäisiä parametrejä. Voidaan käyttää etenkin hienosäätämiseen ja tyylyttelyyn. Textual Inversion -mallit toimivat upotteen tapaan.

2.6 LDM:n LoRA-kouluttaminen

Tutkimuskyselyn LoRA-kouluttaminen suoritettiin AUTOMATIC1111:n käyttöliittymän DreamBooth-lisäosalla³. Tutkimuksessa lisäosan suurimmaksi rajoitteeksi osoittautui sen kyvyttömyys tuottaa pelkkiä LoRA-painoja. Muodostetut painot täytyi yhdistää jonkin olemassa olevan mallin kanssa. Koska diffuusiomallit eroavat toisistaan niiden käyttötarkoituksen takia, olisi laskennallisesti ja työmäärältään raskasta kouluttaa jokainen uusi konsepti uudestaan kaikille eri malleille. Tämän lisäksi lisäosa hidasti ainakin kirjoitushetkellä radikaalisti käyttöliittymän käynnistysaikaa. Tehokkaampana ratkaisuna toimia esimerkiksi kohya-ss:n sd-scripts⁴, minkä avulla mallin kouluttamisen voi suorittaa AUTOMATIC1111:n käyttöliittymän ulkopuolella. Menetelmän avulla voidaan tuottaa pelkkiä LoRA-painoja .safetensors-tiedostomuodossa, eikä niitä tarvitse yhdistää olemassa olevan mallin kanssa. LoRA:sta tekee käytännöllisen ja laskennallisesti tehokkaan etenkin sen malliriippumattomuus. LoRA-malleja voidaan säilyttää omissa kirjastoissaan, ja käyttää niitä vain tarvekohtaisesti. Pohjustavaa mallia voidaan kehittää tai se voidaan jopa vaihtaa kokonaan toiseen tyyliuntaan, eikä painojen käyttämisen tulisi siltikään aiheuttaa ongelmia. Toisin kuin upotteita, LoRA-malleja voidaan käyttää myös muuttamaan prosentuaalisesti generoimisen lopputulemaa. Sen sijaan että se olisi binaarisesti joko päällä tai ei päällä, sen painoja voidaan kutsua esimerkiksi puolikkaalla painoarvolla (painot x 0.5). Tällä voidaan hienosäätää generoimista, ja pyrkiä poistamaan turhia

³ <https://cobanov.dev/diffusion/dream-booth/>

⁴ <https://github.com/kohya-ss/sd-scripts>

yksityiskohtia generoitavasta kuvasta, mitä saattaa muodostua mallia kouluttaessa etenkin, jos kuvausympäristöissä ei ole riittävästi vaihtelevuutta.

Tutkielmassa tehtyjen testien perusteella vaikuttaisi siltä, että, onnistuneiden LoRA-painojen luominen vaatii monipuolisen ja tarkan kuvatietokannan. Kuvien on oltava hyvin tarkennettuja, rajattuja, ja eri kuvakulmissa ja ympäristöissä kuvattuja. Tämän lisäksi on keskeistä sisällyttää tietokantaan kuvakohtaiset tarkasti kuvaansa kuvailevat tekstikuvaukset. Tämä on olennaista etenkin, jos koulutusdatassa kuvia on otettu useista eri värisistä tuotteista, sillä muuten tuotteen värityksen muuttaminen saattaa olla haasteellista. Lopputulosta voidaan myös parantaa regularisoinnilla. Regularisoinnin tarpeellisuus on kuitenkin riippuvainen mallin kouluttamisen tekstikuvausten tarkkuudesta, sillä niiden avulla voidaan myös välttää mallin ylikouluttamista ja parantaa sen yleistämiskykyä. Tämän tutkimuksen mallien kouluttamisessa tekniikkaa ei käytetty ollenkaan.

LDM:en kouluttamiseen vaadittavat parametrit ja menetelmät vaihtelevat tarve- ja tapauskohtaisesti. Kyselytutkimusta varten LDM:n kouluttamiseen käytettiin seuraavia LoRA-parametreja:

- LoRA extended (suurempi tiedostokokoo, mutta parantaa koulutustuloksia).
- Koulutuskierrokset (*epochs*): 200.
 - Koulutuskierroksia oli suhteellisen paljon latentille diffuusiomallille, mutta lopullinen malli on yhdistetty Illuminati Diffusion v.1.1 -mallin kanssa suhteella 7/3, mikä vähensi ylisovittamisesta, eli liiallisesta koulutuksesta, syntyviä artefakteja säilyttäen kuitenkin keskeisimmät uudet LoRA-painot. Mallin kouluttamisessa ei käytetty myöskään ollenkaan regularisointikuvia.
- Koulutusresoluutio: 768px x 768px (käytössä laitteiston, optimoinnin, ja pohjustavan LDM:n rajoitusten takia).
- Mixed precision: fp16.
 - Viittaa 16- ja 32-bittisten liukulukutyypin käyttöön mallin kouluttamisessa. Nopeuttaa toimintaa ja käyttää vähemmän muistia (Li, Cheng & Lin, 2021).
- Memory attention: xformers (vähentää videomuistin käyttöä).
 - Tämän lisäksi kaatumisen ehkäisemiseksi AUTOMATIC1111:n käyttöliittymän käynnistymisparametreissa koulutuksen aikana --xformers ja --no-half, ja muulloin pelkästään -xformers.
- Sanity Sample kehote: *“A pink Pinja spoke chair, in a snowy forest”*.
- Määrittelee esimerkkikuvien kehotteen. Auttaa havainnoimaan milloin mallia on koulutettu tarpeeksi (ks. kuvio 5).
- Instance & sample -kehote: [filewords]. Lukee joko tiedostonimen tai aiemmin määritetyn kuvauksen, ja suorittaa kouluttamisen sen perusteella (*text embedding*).



KUVIO 5 Koulutuksen yhteydessä generoidut esimerkkikuvat (*sanity sample*). Koherenssi kasvaa koulutusaskelten myötä. 768px x 768px.

2.7 LDM:n käyttäminen

LDM:lla voidaan generoida kuvia useilla tavoilla. Generoiminen voidaan suorittaa puhtaasti tekstikehotteisesti (*Text-to-Image*, ks. kuvio 6), jonkin olemassa olevan kuvan pohjalta (*Image-to-Image*, ks. kuvio 7) tai generoimalla olemassa olevaan kuvaan osittaisia muutoksia Inpainting-menetelmällä (Rombach ym., 2022; Nichol ym., 2022; Karras ym., 2022; Yang ym., 2022). Jo pelkän tekstikehotteen perusteella voidaan muodostaa laadukkaita tuotoksia, mutta tällöin tulokset ovat usein arvaamattomampia ja kuvan värittäminen ja asetteluun voidaan vaikuttaa vain suuntaa antavasti (ks. kuvio 6). Toisaalta kehotteen avulla voidaan muodostaa useita toisistaan poikkeavia tuotekuvia ja valita näistä parhaimmat jatkokehitykseen tai käyttöön.

Tekstikehote on edellytyksenä kaikissa generoimistyylyissä ja siis sen on syytä olla riittävän spesifi kuvastamisen onnistumiseksi. Onnistuakseen kehotteen tulisi vastata jotain mallille koulutettujen konseptien yhdistelmää. Kehotteen voi antaa mallille positiivisena tai negatiivisena. Positiivisena kehotteena tulisi olla se, mitä kuvaan halutaan generoida, ja negatiivisena se, mitä tulisi generoimisessa välttää. Omassa työssäni negatiivisena kehotteena kutsuttiin negatiivisia upotuksia *nfixer*, *nrealfixer* ja *nartfixer*. Negatiivisena kehotteena voisi kuitenkin olla myös vaikkapa yleisesti termit *noise*, *low resolution* ja *bad quality*. Ensimmäisenä tarkoituksena olisi vähentää kuvassa olevaa kohinaa, toisen heikkoon resoluutioon yhdistettävissä olevia artefakteja, ja kolmannen yleistä huonoa laatua. On kuitenkin mallista riippuvaista mitä kehotteita siinä voidaan ja tarvitsee käyttää.

Pelkän tekstikehotteen lisäksi LDM:a voidaan käyttää jonkin olemassa olevan kuvan avulla. Tällä niin generoidun tuotoksen sävyttämiseen kuin asetteluunkin voidaan vaikuttaa tarkemmin. Tulokset ovat alkuperäistä kuvaa muistuttavia, mutta niiden asettelu ja mittasuhteet eivät ole kuitenkaan absoluuttisia kopioita siitä. Tämän ansiosta kyseistä generoimismenetelmää voidaan soveltaa alkuperäisen kuvan virheiden korjaamiseen (ks. kuvio 7). Kuvasta kuvaan generoidessa keskeisessä roolissa on parametri *Denoising strength*. Tämä määrittää sen kuinka paljon generoitu sisältö vastaa sille annettua kuvaa. Asetuksen ollessa pois käytöstä, eli 0, kuva on täsmälleen sama kuin mitä alkuperäinenkin. Kun

asetus taas nostetaan tasolle 1, kuva on täysin uniikki, ja se muistuttaa alkuperäistä enää keskeisiltä piirteiltään.



KUVIO 6 Pelkän tekstikehötteen perusteella generoituja kuvia (*Text-to-Image*). Kehote: “Red Pinja spoke chair, on a beach, with high cliffs and stormy sea in the background”. 768px x 768px.



KUVIO 7 Tekstikehötteen ja taulukon ensimmäisen kuvan perusteella generoituja kuvia (*Image-to-Image*). Kehote: “Red Pinja spoke chair, on a beach, with high cliffs and stormy sea in the background”. 768px x 768px. Denoising strength: 0.75.

Taulukon neljä kuvista käy ilmi, että pelkästään tekstistä kuvia generoidessa kuvien väliset mittasuhteet poikkeavat todellisuudesta, ja että kuvien välillä on paljon eroa. Lisäksi kuvista löytyy koherenssivirheitä. Kuvastettavan tuotteen yksityiskohdat, kuten pintojen määrä tai paksuus sisältävät artefakteja ja vaihtelua niiden paksuuksissa ja pituuksissa. Taulukossa viisi voidaan kuteinkin havaita, että vaikka generoiminen ei onnistu täydellisesti, mittasuhteiltaan onnistuneimpia kuvia voidaan hyödyntää uusien kuvien generoimisessa, minkä ansiosta kuvien mittasuhteet vaikuttavat paremmilta, mutta ne poikkeavat silti todellisuudesta. Alkuperäisessä kuvassa on virheitä pintojen muodoissa, mitkä sen perusteella generoiduista kuvista kuitenkin katoavat.

Puhtaasti teksti- ja kuvapohjaisesti voidaan saavuttaa suhteellisen hyviä tuloksia, mutta generoinnin mittasuhteiden ja yksityiskohtien täsmällisyyteen voidaan vaikuttaa vielä tarkemmin hyödyntämällä kuvasuhteiden normalisoivia ControlNet-toiminnallisuuksia (Zhang & Agrawala, 2023). Tämä onnistuu esimerkiksi AUTOMATIC1111:n verkkopohjaisen käyttöliittymän ControlNet⁵-lisäosalla. ControlNetin avulla voidaan generoida mallin alkuperäistä koulutusresoluutiota suurempia kuvia (kuten kuviossa 3) ilman vääristymiä (mitä havaittavissa kuvioissa 6 & 7) tai kuvannettavan kohteen monistumista.

Myöhemmässä tutkimusvaiheessa kuvien normalisointiin käytettiin ControlNet-mallia canny-sd21-safe. ControlNet tarjoaa useita mahdollisuuksia kuvastamista normalisoivien toiminnallisuuksien suhteen. Sen avulla voidaan

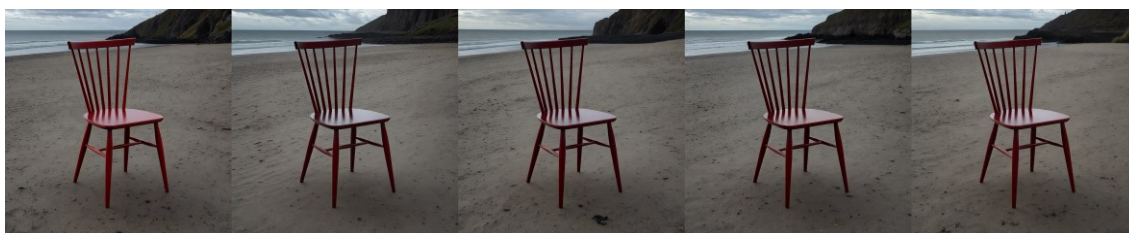
⁵ <https://github.com/Mikubill/sd-webui-controlnet>

generoida sisältöä karkeista piirroksista aina yksityiskohtaisiin syvyys- ja reunakarttoihin asti (reunakarttojen muodostaminen ks. kuvio 10). Yksityiskohtaisuudesta on hyötyä etenkin fotorealistisia tuotekuvia generoidessa. Canny-reunakartat sisältävät tietoa myös tuotteen heijastuksista, mikä syvyyskartoista uupuu, ja mahdollistavat sen ansiosta myös tarkkojen heijastusten generoimisen.

Reunakarttoja voidaan hyödyntää pelkän tekstikehotteen kanssa (taulukko 6) tai niiden apuna voidaan käyttää jotain tiettyä kuvaa (kuviossa 9). Uusia kuvia generoidessa taustoihin saadaan enemmän vaihtelevuutta pelkän tekstikehotteen avulla, ja mieluisen taustan löytyessä kuvan sävytystä ja yksityiskohtia voidaan pyrkiä korjaamaan generoimalla sen perusteella lisää kuvia.



KUVIO 8 Canny-reunakartan ja tekstikehotteen perusteella generoituja kuvia (*Text-to-Image/ControlNet*). Kehote: “Red Pinja spoke chair, on a beach, with high cliffs and stormy sea in the background”. 1024px x 1024px / kuva.



KUVIO 9 Canny-reunakartan, tekstikehotteen, ja taulukon ensimmäisen kuvan perusteella generoituja kuvia (*Image-to-Image/ControlNet*). Kehote: “Red Pinja spoke chair, on a beach, with high cliffs and stormy sea in the background”. 1024px x 1024px / kuva. Denoising strength: 0.75.

Kuviossa kahdeksan esitellään kuinka reunakartat normalisoivat generoitavan tuotteen sommittelun ja mitoituksen, ja niiden avulla voidaan luoda kuvia, joiden mitoitus on äärimmäisen tarkka, mutta joiden taustoissa on kuitenkin yhä paljon vaihtelevuutta. Normalisoitu kehotepohjainen generointi mahdollistaa siis samankaltaisten tuotekuvien generoimisen, mutta joissa on yhä enemmän vaihtelua keskenään. Kuviossa yhdeksän esitetään kuinka tuotekuvien mittasuhteet mittasuhteet ja myös taustan keskeiset ominaisuudet säilyvät, mutta kuvan värityksessä on hienoa vaihtelua. Verrattuna kuitenkin kuvion seitsemän kuviin, taustassa olevat eroavaisuudet ovat hienompia, vaikka käytetty Denoising strength olisikin sama.



KUVIO 10 Tuotteen taustan poistaminen ja Canny-reunakartan luominen. Rajatun kuvan koko oli 1024px x 1024px (tekoahella ControlNet-lisäosan maksimaalinen kehyskoko).

Kuvioissa kymmenen ja esitetään kuinka Canny-reunakartan voi muodostaa leikkaamalla kohde ensin taustastaan. Tällöin reunakartta ei sisällä ylimääräistä tietoa, mikä voisi häiritä generoimista. Toisaalta reunakarttaan voi lisätä myös manuaalisesti piirteitä, joiden avulla generoimista saadaan normalisoitua entiseltään. Jos tuote vaikkapa halutaan generoida sisätiloissa, voidaan reunakarttaan laittaa viiva kohtaan, johon halutaan huoneen lattian ja seinän sauman asettuvan. Tämä vähentää vääristymiä, mitkä ovat yleisiä etenkin suuremmissa generoimiskoissa. Tämän tutkimuksen esimerkkikuvissa käytetty reunakartta on tehty kuvasta, jota ei ollut alkuperäisessä koulutustietokannassa, eli perspektiivi on täysin uusi. Reunakarttaan hyödynnettävän kuvan ei siis tarvitse olla mallin tuntemasta kuvakulmasta, jotta se saavuttaisi laadukkaita synteetitulosia (ks. kuvio 11).



KUVIO 11 Tekstikehotteen ja Canny-reunakartan perusteella generoitu kuva. Kehote: "A white Pinja spoke chair, in front of grass plains and distant mountains, during golden hour, sunset". 1024px x 1024px.

2.8 LDM:lla generoimisen aliprosessit ja parametrit

Tässä luvussa käsitellään kandidaatintyössä tehtyjä havaintoja tuotekuvien fotorealistisesta generoimisesta, ja siihen tarvittavista prosesseista ja niiden parametreista. Erilaiset mallit vaativat kuitenkin erilaista käsittelyä, ja sen vuoksi parhaita tuloksia voi usein saavuttaa tutustumalla mallin lähdekirjallisuuteen tai muihin siihen tehtyihin havaintoihin. Tutkimuksen kuvangeneroimisessa käytetyt menetelmät olivat:

- Näytteenottomenetelmä (*sampling method*): UniPC^{6,7} (käytössä jos *Text-to-Image*) tai DDIM (käytössä jos *Image-to-Image*).
 - Näytteenoton tärkein tavoite on tehdä päätelmiä suuremmasta ryhmästä pienen ryhmän tiedon perusteella (Yang, ym. 2022). Näytteenottomenetelmän valinta vaikuttaa siihen, miten latentti tila tutkitaan, ja miten monipuolisia kuvia sen perusteella voidaan luoda. Sekä UniPC että DDIM näyttävät hyötyvän korkeammasta lukemasta.
 - UniPC on koulutukseton kehys, joka on suunniteltu malliriippumattomaksi ja nopeaksi näytteenottomenetelmäksi (Zhao, Rao, Zhou & Lu, 2023).
 - DDIM (Deep Density Inverse Mapping) voi kalibroida regressiotöiden epävarmuuden mallintamalla niiden ulostulon jakauman käänteisyyden tiheyttä. DDIM-mallilla voidaan auttaa parantamaan ennustusten luotettavuutta ja tarkkuutta (Song, Meng & Ermon, 2020).
 - Parhaita tuloksia tuottava näytteenottomenetelmä voi vaihdella mallien ja koulutusmenetelmien erojen takia.
- Näytteenoton askeleet (*sampling steps*): 60.
 - Vaikuttaa kuvan yksityiskohtaisuuteen (esitelty liitteessä 1). Mitä enemmän askelia, sitä yksityiskohtaisempi lopputulos (ks. liitteet 11 & 13).
- CFG (Classifier-Free Guidance) scale: 3.5–4.
 - Asetus Stable Diffusion -mallissa. Säättää kuinka paljon kuva näyttää tekstisyötteeltä, mutta korkeammat arvot lisäävät todennäköisyyttä etenkin väriarvojen liialliselle tyylyttelylle fotorealismiin tähdätessä (liite 12).
 - Käytetty CFG scale arvo on pienin mahdollinen, joka tuottaa halutun tekstisyötteen
- Denoising strength: 0.25–1.
 - Keskeinen parametri *Image-to-Image*-generoimisessa. Mitä pienempi arvo, sitä enemmän generoitava kuva vastaa alkuperäistä sille annettua kuvaa.

⁶ <https://unipc.ivg-research.xyz/>

⁷ <https://github.com/wl-zhao/UniPC>

2.9 Kuvan resoluutio ja sen muunnokset

Fotorealistisen tuotekuvastaminen edellyttää yksityiskohtaisuutta. Kuvien yksityiskohtaisuuden määrä on usein riippuvainen kuvan resoluutiosta. Kuvista on siis syytä generoida mahdollisimman suuria, jotta ne sisältävät myös mahdollisimman paljon myös yksityiskohtia. Kuvan lopullinen resoluutio ja tiedostomuoto tulisi kuitenkin määrittää myös julkaisualustan suositusten perusteella. Generoidun tuotekuvan resoluution ei kannata olla alustan maksimaalista tukevaa resoluutiota suurempi, sillä sovellusalusta saattavat pakata kuvaa pienemmäksi ja tuettuihin tiedostomuotoihin eri algoritmein, jotka mahdollisesti vaikuttavat negatiivisesti kuvan laatuun.

Generoidun kuvan spatiaalinen resoluutio vaikuttaa myös siihen kuinka suurella näytöllä sitä voidaan esittää ilman, että kuvan digitaalinen ja fyysinen pikselikoko eroavat merkittävästi. Jos digitaaliset pikselit ovat fyysisiä pikseleitä suurempia, kuvaa tulisi joko skaalata näyttöön sopivaksi (liite 4) tai näyttää sen natiiviresoluutiolla, jolloin se ei kuitenkaan täytä koko näyttötilaa. Jos taas digitaalinen resoluutio on suurempi kuin näytön resoluutio, kuva voi hyötyä resoluutioiden erosta johtuvasta Downsampling anti-aliasing -ilmiöstä, joka voi vähentää kuvan rosoisuutta ja lisätä sen yksityiskohtaisuutta ja laatua. Tämä perustuu kuvan interpolointiin ja on verrattavissa Supersampling-tekniikkaan. Downsampling anti-aliasing on tekniikka, jolla parannetaan kuvanlaatua vähentämällä aliasointia eli kuvan säröilyä. Aliasing syntyy, kun kuvaa pienennetään tai näytetään aliresoluutioisella näytöllä. Downsampling anti-aliasing toimii siten, että ennen kuvan pienentämistä siihen sovelletaan matalataajuussuodatinta (*low-pass filter*), joka tasoittaa kuvan teräviä reunoja ja vähentää korkeita taajuuksia.

Jos halutaan kasvattaa kuvan resoluutiota ilman yksityiskohtien menetystä, voidaan käyttää superresoluutiota. Superresoluutiossa kuvan kokoa suurennetaan generoimalla uutta tietoa pikseleiden väliin (ks. liitteet 5 & 6). Tämän tavoitteena on parantaa kuvien yksityiskohtaisuutta (Wang, Dong & Shan, 2021; Liang, Sun, Van & Timofte, 2021).

Lanczos-algoritmi on yksi esimerkki superresoluutiomenetelmästä. Se perustuu matemaattiseen funktioon nimeltä sinc-funktio. Lanczos-algoritmi on tyypillisesti hyvä muuttamaan kuvan kokoa, mutta se voi aiheuttaa joitakin ei-toivottuja pehmennyksiä yksityiskohdissa, ja voi olla vielä hieman epätasainen (Bashir, Wang, Khan & Niu, 2021). Lanczos-algoritmi on erinomainen yleinen skaalausalgoritmi, mutta se ei tuo esiin uusia elementtejä kuvissa. Sitä voidaan kuitenkin käyttää hyvin esim. kuvaresoluution kaksinkertaistamiseen, ja pohjana muilla kuvanparantamisalgoritmeille.

ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) on tässä tutkimuksessa keskeisessä roolissa ollut superresoluutiomenetelmä. Sen avulla voidaan pyrkiä tuottamaan fotorealistisia kuvia yksittäisen kuvan superresoluutiolla. ESRGAN perustuu generatiiviseen vastustusverkkoon (GAN), joka koostuu kahdesta osasta: generaattorista ja diskriminaattorista. Generaattori pyrkii luomaan korkearesoluutioisia kuvia alhaisen resoluution kuvista, kun taas

diskriminaattori pyrkii erottamaan aidot ja synteettiset kuvat. Generaattori ja diskriminaattori kilpailevat keskenään ja parantavat samalla kuvien laatua (Wang, Xie, Dong & Shan, 2021).

Lanczos- ja ESRGAN-superresoluutiomenetelmiä on mahdollista hyödyntää suoraan AUTOMATIC1111:n käyttöliittymästä käsin. ESRGAN ja Lanczos eroavat toisistaan sekä algoritmin että tuloksen suhteen. ESRGAN on monimutkaisempi ja tehokkaampi menetelmä, joka hyödyntää syväoppimista ja GAN-mallia. Lanczos säilyttää kuvien alkuperäisen sisällön paremmin kuin ESRGAN, mutta se ei lisää uusia yksityiskohtia tai piirteitä kuvissa (Bashir, Wang, Khan & Niu, 2021).

LDM:a voidaan myös hyödyntää superresoluutiossa. Latent upscale on prosessi, jossa LDM:a sovelletaan superresoluutioskaalaamiseen. Tutkimuksen yhteydessä sen tuotokset olivat kuitenkin arvaamattomia, ja etenkin kuvan keskeisten mittasuhteiden säilyttämisessä oli ongelmia. ControlNet v1.1 -versiossa lanseerattu Tile-upscale on prosessi, jolla generoitavan kuvan tuloksia voidaan pyrkiä normalisoimaan. Menetelmä käyttää esikäsittelijää, jolle syötetään pienikokoinen kuva, minkä perusteella se generoi uuden kuvan (Mikubill, 2023). Tutkimuskyselyä valmistellessa lisäosaa ei ollut vielä kehitetty, ja kirjoitushetkellä sitä ei voitu soveltaa Stable Diffusion v.2.1 LDM:lle.

2.10 In- & Outpainting

Generoidun kuvan sisältö tai rajaus ei ole aina toivotunlainen. Superresoluutio suurentaa kuvaa kokonaisuutena samassa suhteessa, mutta LDM:t mahdollistavat myös kuvien osittaisen muokkauksen. Esimerkiksi kuvan taustaa voidaan laajentaa sen alkuperäisten reunojen ulkopuolelle Outpainting-tekniikalla tai kuvan sisältöä voidaan muuttaa osittain Inpainting-toiminnoilla. Inpainting oli tärkeä erityisesti laadullisen kyselytutkimuksen tutkimuskuvien muodostamisessa, sillä ennen ControlNet-lisäosaa generoitu sisältö sisälsi todennäköisemmin jotain korjaamista vaativia ongelmia (ks. kuvio 6). Molemmat prosessit voidaan toteuttaa ohjaamalla latentin diffuusiomallin generointiprosessia erilaisilla ehdoilla, kuten tekstikehotteella ja rajauslaatikoilla (Rombach ym., 2022). Kyselytutkimustulosten perusteella tekniikat kuitenkin vaikuttivat heikentävästi kuvien kohe-reenssiin, joten niiden käyttäminen ei ole etenkään fotorealismiin tähdätessä aina suotavaa.

2.11 Fotorealistinen kuvangeneroiminen suhteutettuna aiempiin fotorealistisen kuvastamisen menetelmiin

Vaikka fotorealistinen kuvangeneroiminen latenteilla diffuusiomalleilla onkin suhteellisen uusi menetelmä, voidaan kuvangeneroimisen eettisyyttä verrata myös aiempiin tutkimuksiin, missä muita menetelmiä on käytetty fotorealistisen materiaalin tuottamiseen. Näitä metodeja ovat muun muassa kuvamanipulaatio ja 3D-mallit yhdistettynä johonkin realistiseen renderöintiohjelmistoon. Kuvangeneroimisen suurin eroavaisuus kyseisiin menetelmiin on sen suhteellinen yksinkertaisuus; metodilla voidaan tuottaa kokonaisvaltaisesti fotorealistisesti kuvia jopa pelkän kehotteen perusteella. Aiemmissä menetelmissä manuaalisen työn määrä on ollut olennaisemmassa merkityksessä, mutta toisaalta ne mahdollistavat kuvan yksityiskohtien hallitsemisen korkeammalla tasolla.

Wangin ja Hunin (2023) tutkimuksessa käsitellään kuvamanipulaatiota ja sen vaikutusta katsojan havaintoon ja tulkintaan kuvasta. Tutkimuksessa käsitellään myös tekniikan väärinkäyttömahdollisuuksia ja eettisiä ongelmia niissä saattaa olla. Tutkimuksen mukaan kuvamanipulaation vaikutus katsojan havaintoon ja tulkintaan kuvasta voidaan jakaa kolmeen osaan: faktapoikkeamaan, tilanteen uudelleenrakentamiseen ja luottamuskriisiin. Teoksessa käsitellään myös kuvamanipulaation mahdollistavia elementtejä, kuten suodattimia (*filters*), uudelleenjulkaisua ja arvojen muutosta (Wang & Hu, 2023). Teoksen perusteella voidaan todeta, että kuvamanipulaatioon ei siis aina edes manuaalista työtä. Joskus esimerkiksi puhelimen kameraohjelmisto saattaa ajaa muutoksia kuviin automaattisesti.

Holmgrenin ja Anderssonin (2012) tutkimuksessa käsitellään fotorealistista montaasia kuvankäsittelyn muotona, jossa pyritään luomaan kuvia, jotka näyttävät todellisilta, mutta jotka eivät ole peräisin yhdestä valokuvasta. Montaasissa käytetään erilaisia kuvankäsittelyohjelmia, kuten Photoshopia yhdistämään eri kuvia tai kuvaelementtejä yhdeksi kuvaksi. Toisin kuin mitä LDM:a käyttäessä montaasissa on tärkeää huomioida valonlähteet, varjot, perspektiivi ja värit, jotta lopputulos näyttää uskottavalta. Tutkielmassa annetaan esimerkkejä erilaisista fotorealistisista montaaseista, kuten maisemista, henkilökuvista ja surrealistisista kuvista. Kirjoittajat potivat myös sitä, miten fotorealistinen montaasi vaikuttaa katsojan havaintoon ja tulkintaan kuvasta. Teoksessa todetaan, että fotorealistinen montaasi voi olla sekä informatiivinen että manipulatiivinen, ja mahdollistaa uusien merkitysten ja tarinoiden luomisen kuvien avulla (Holmgren & Andersson, 2012).

Wangin ja Hunin (2023) tutkimuksessa ja Holmgrenin ja Anderssonin (2012) kandidaatintyössä käsitellään kuvamanipulaatioita, mutta samat eettiset ongelmat ovat myös mukana kokonaisvaltaisen kuvangeneroimisessa. Materiaalin erona on lähinnä sen tuottamistapa. Fotorealistisesti generoitu sisältö, joka ei ole läpinäkyvästi ja eettisesti tuotettu, saattaa harhauttaa tai vääristää katsojan käsitystä todellisuudesta. Siispä sen voisi sanoa olevan manipulatiivista, ellei kuvien luonnetta kerrota selkeästi kuvan yhteydessä.

2.12 Fotorealistisesti generoidun kuvamateriaalin havaitseminen

Toistaiseksi keinotekoisesti generoituja kuvia voidaan jossain tapauksissa havaita kaksiulotteisella diskreetillä Fourier-muunnoksella (*discrete fourier transform, DFT*) (Corvi ym., 2022). Fourier-muunnos antaa tietoa kuvan taajuussisällöstä hajottamalla kuvaa sen siniaaltojen summaksi. Tekniikalla voidaan paljastaa selkeimpiä generoimisartefakteja. Fourier-muunnosta käytettiin generoidun sisällön tutkimiseksi, mutta tutkimuksen nopean testiprosessin perusteella sillä saattoi havaita yksiselitteisesti ainoastaan superresoluutiolla muodostetut kuvat. Superresoluution myötä kuvissa on havaittavissa raitoja niin pysty kuin vaak akselilla (liite 7). Samoja artefakteja ilmenee myös normaalissa valokuvassa superresoluutioskaalauksen myötä (liite 8).

Superresoluution jälkeen sekä oikeaan valokuvaan että generoituihin kuviin lisättiin kohinaa sen vaikutuksen tutkimiseksi. Tutkimus perustui kohinan potentiaaliin vaikutuksiin kuvantamisartefaktien havaitsemisessa. Jos superresoluutiolla skaalattuun sisältöön lisätään ilman muita toimenpiteitä esimerkiksi Uniform-kohinaa, se näkyi kuvassa toistuvana ja symmetrisenä kuviointina (liite 10). Tämä johtuu siitä, ettei kohina ole täysin satunnaista. Kuvien todellista luonnetta pystyi kuitenkin piilottamaan hyödyntämällä interpolointia. Kuvat skaalattiin ensiksi 4x suuremmaksi 9960px x 9960px resoluutioon, ja niihin lisättiin 2 % värillistä Gaussian- ja 1 % monokromaattista Uniform-kohinaa. Tämän jälkeen kuvat skaalattiin takaisin aiempaan 2048px x 2048px resoluutioon käyttäen Bilinear-interpolointia. Skaalaamisen myötä kuvantamisartefaktit katosivat ainakin osittain, ja niiden hahmottaminen Fourier-muunnoksen avulla ei ollut enää yksiselitteistä (liite 9 & 11). Kohina lisää yksityiskohtaisuuden vaikutelmaa ja interpoloinnissa kuvapisteiden keskiarvoilla selkeimmät synteesiartefaktit saadaan piilotettua. Toisaalta kuvissa on yhä mukana selkeät viivat vaak- ja pysty akselilla, mitä oikeista valokuvista tai generoiduista kuvista ei voi havaita niin yksiselitteisesti.

Tuloksien perusteella voidaan päätellä, että uusia kuvasynteesin tunnistavia järjestelmiä on syytä kehittää disinformaation levittämisen ehkäisemiseksi. Voi olla, että koneoppimista vastaan tulee kehittää uusia kehittyneempiä koneoppimisjärjestelmiä, jotka koulutetaan tunnistamaan sisältöä suuremmalla todennäköisyydellä, tai että autenttisiin kuviin aletaan lähtökohtaisesti sisällyttämään esimerkiksi Zero-knowledge proof (nollatiedoston todistus), ja kuvat, jotka eivät tätä sisältä luokitellaan mahdollisesti tekoälyllä tuotetuksi. Zero-knowledge proof (nollatiedon todistus) on kryptografian käsite, joka mahdollistaa yhden osapuolen todistaa toiselle osapuolelle jonkin väitteen oikeellisuuden ilman, että samalla paljastettaisiin mitään ylimääräistä tietoa. Protokollan tarkoituksena on vakuuttaa verifioija ilman, että todistaja paljastaa mitään tietoa, joka voisi auttaa verifioijaa itse suorittamaan tehtävää. Zero-knowledge proof -menetelmä käytetään laajasti tietoturvaratkaisuissa, kuten salaisen tiedon todentamisessa tai pääsynvalvonnassa. Se voi siis mahdollistaa turvallisen tiedonvaihdon ilman, että osapuolten tarvitsee luottaa toisiinsa täysin.

3 KYSELYTUTKIMUS SYNTEETTISESTÄ TUOTEKUVASTAMISESTA

3.1 Kyselytutkimuksen kuvaus

Tutkielmassa kuvangeneroimisen laadulle haettiin validiteettia kyselytutkimuksen avulla. Kyselyssä kerättiin ensiksi ihmisten mielipiteitä synteettisesti generoitujen kuvien hyödyntämisestä markkinointiviestinnässä sekä heidän aiemmasta tietämyksestään ja kokemuksistaan LDM:sta. Kysymykset keskittyivät myös generoitujen kuvien hyödyntämisen eettisyyteen, ja ihmisten toiveisiin siitä, miten niiden käytöstä tulisi viestiä. Tämän jälkeen kyselyyn vastaajat tarkastelivat itsessään tekniikalla tuotettuja tuloksia, ja antoivat mielipiteensä siitä, kuinka fotorealistisesti tuotokset onnistuivat Likert asteikolla 1–5. Kysymysten yhteydessä oli myös vapaavalintaisia kysymyksiä, joissa vastaaja saattoi kertoa mielipiteensä onnistumiset ja epäkohdat. Vastausvaihtoehtoina olivat: kokonaisuus, tausta (eli tuotteen ympäristö), heijastukset, valaistus, varjot, väritys, yksityiskohdat ja mittasuhteet. Menetelmä mahdollisti monipuolisen tiedonkeruun, ja antoi mahdollisuuden kerätä tietoa sekä ihmisten yleisistä mielipiteistä, että heidän henkilökohtaisista kokemuksistaan aiheen suhteen. Tutkimus toteutettiin kyselytutkimuksena Google Formsin avulla erityisesti mobiilikäyttäjää ajatellen. Tämän kandidaatintyön kyselytutkimuksessa keskityttiin kaksiulotteisten tuotekuvien generoimiseen LDM:lla. Diffuusiopohjaiset kuvantamistekniikat toimivat tarvittaessa myös audion (Forsgren & Martiros, 2022), 3D-mallien (Anciukevičius ym., 2022; Karnewar ym., 2023) ja videon (Khachtatryan ym., 2023; Luo ym., 2023) generoimiseen. Tutkimuskyselyn jakaminen suoritettiin sovellusalueita Instagramia ja WhatsAppia hyödyntäen. Kyseiset markkinointialustat soveltuvat kuvapohjaiseen tuotemarkkinointiin erinomaisesti niiden visuaalisen luonteen ja interaktiivisuuden ansiosta, ja siispä valinta oli luonteva tuotekuvastamiseen. Otanta koostui tutkijan kontaktipiiristä. Tutkimukseen vastasi kaiken kaikkiaan 44 henkilöä. Tutkimukseen osallistuneista yli puolet (26 henkilöä) oli 24–

30-vuotiaita, vajaa neljännes (10) 18–24-vuotiaita, ja loput (8) 30–60-vuotiaita. Vastaajista noin 60 % oli miehiä, 38 % naisia ja 2 % muunsukupuolisia. Kyselytutkimuksessa kysyttiin vastaajilta myös heidän tämänhetkisestä korkeimmasta koulutustasostaan. Heistä 56,8 % oli suorittanut alemman korkeakoulututkinnon, 18,2 % ylioppilastutkinnon, 15,9 % ammatillisen perustutkinnon, 4,5 % lukion ja 4,5 % ylemmän korkeakoulututkinnon. Vastaajien tausta heidän opintoalajensa suhteen oli laaja. Heistä 29,5 % koulutus pohjautui liike- ja/ tai taloustieteisiin, 20,5 % humanistisiin tieteisiin, 18,2 % tekniikkaan ja/ tai tietojenkäsittelyyn, 11,4 % sosiaali-, terveys- ja hyvinvointialaan, ja loput erillisiin koulutushaaroihin kuten kasvatustieteisiin. Vastanneista 50 %:lla ei olla ei ollut mitään taustatietoja latentilla diffuusiomallilla kuvangeneroimisesta. 31,8 % vastasi omaavansa vain vähän taustatietoa aiheesta, 11,4 % jonkin verran, ja loput kohtalaisesti tai paljon. Henkilöistä 61,4 % ei ollut käyttänyt tai nähnyt käytettävän kuvangeneroimismenetelmiä, 22,7 % kertoi nähneensä niiden käyttöä muttei ollut itse kokeillut, 9,1 % oli kokeillut menetelmiä, ja loput olivat käyttäneet niitä jonkin verran tai paljon. Tutkimuskyselyn lopussa 68,2 % vastaajista kertoi mielipiteensä muuttuneen diffuusiopohjaisten kuvangeneroimismenetelmien hyödyntämisestä.

3.2 Kyselytutkimuksen rajoitukset

Teknologian nopea kehittyminen aiheutti ongelmia kyselytutkimuksen suunnittelussa, ja tulosten ajankohtaisena pitämisessä tutkielman tekemisen aikana. Koska kehitystä on tapahtunut tutkimuskyselyn valmistelun jälkeen muun muassa generoitavassa spatiaalisessa resoluutiassa, kuvien väri- ja valotoistossa sekä mittasuhteiden oikeellisuudessa, eivät siinä käytetyt kuvat enää demonstroi parasta tämänhetkisillä tekniikoilla saavutettavaa laatua. Kyselytutkimuksen materiaalit pohjautuivat pelkkään Stable Diffusion v.2.1 non-EMA LDM:n, sillä tutkimuskuvien valmisteluajalla ei ollut vielä saatavilla LDM:a joka olisi valmistettu Noise Offset -tekniikalla. Tämän takia tutkimus suunniteltiin keräämään myös yleisiä mielipiteitä synteettisesti generoidun sisällön käyttämisestä markkinointiviestinnän välineenä, eikä ainoastaan kuvilla saavutettavasta laadusta. Kuvan fotorealiseksi määrittäminen voi olla myös kontekstista riippuvaista, ja koska tässä tutkimuksessa samasta tuotteesta on muodostettu useita vaihtoehtoisia kuvia, katsoja pystyi kiinnittämään erityisesti huomiota niiden eroavaisuuksiin ja epäkohtiin. Kuvastamalla useampia tuotteita ja lisäämällä kysymysten määrää olisi tutkimustuloksista saanut yksityiskohtaisempaa tietoa.

Kyselytutkimuksen rajoitteena on myös vähäinen osallistujamäärä. Kyselyyn osallistui 44 henkilöä, eli sen otanta on suhteellisen pieni, mikä voi aiheuttaa vääristymiä tutkimustuloksissa. Tämän lisäksi tutkimuksen diffuusiomallin koulutukseen käytetty kuvamateriaali kerättiin älypuhelimella automaattisia asetuksia hyödyntäen, mikä vaikuttaa generoimisella saavutettavaan laatuun. Suuremmalla sensorilla ja laadukkaammalla optiikalla voitaisiin parantaa kuvien laatua ja yksityiskohtaisuutta. Koulutustietokannan koko oli myös pieni, vain 28 kuvaa, joista puolet oli käänteisiä kopioita. Suurempi vaihtelu taustojen ja valaistuksen

suhteen parantaa saavutettavissa olevaa laatua. Lisäksi yksi myöhemmin ilmenneistä epäkohdista kyselyssä oli se, ettei jokaisen kuvan yhteydessä ollut myös oikeita valokuvia, joihin tuotosta olisi voinut verrata.

Puuttuvia tekniikoita ovat myös aiemmin kuvatut koherenssiin, valaistukseen, väritykseen ja yleiseen fotorealismiin vaikuttavat negatiiviset upotukset, nixer, nrealfixer ja nartfixer. Kuvia generoidessa ControlNet-lisäosaa ei myöskään ollut saatavilla Stable Diffusion v2.1 pohjautuviin malleihin, eli generoitujen kuvien mittasuhteet eivät ole absoluuttisia, ja ne sisältävät paljon virheitä. Kyselytutkimuksen kuviksi valittiin parhaiten onnistuneet kuvat, ja kuvan generoimista toteutettiin niin tekstiin kuin kuviin perustuen. Myöhemmin ControlNet on vähentänyt tarvetta generoida useita kuvia, ja kuvat onnistuvat todennäköisemmin vähemmällä kokeilukerroilla suoraan haluttuun muotoon.

Tutkimuksen laadun arvioinnissa on tärkeää ottaa huomioon otannan luotettavuus ja validiteetti. Nämä käsitteet kuvaavat sitä, kuinka hyvin tutkimusmenetelmä, -tekniikka tai -testi mittaa sitä, mitä sen on tarkoitus mitata. Luotettavuus viittaa mittauksen johdonmukaisuuteen ja validiteetti mittauksen tarkkuuteen (Middleton, 2019). Tutkimuksessa käytettiin otantana tutkijan kontaktipiiriin kuuluvia henkilöitä, jotka vastasivat kyselytutkimukseen Instagramin ja WhatsAppin kautta. Otanta oli suhteellisen pieni (n=44) ja kysely kohdistettiin varsin rajalliselle vastaajajoukolle. Tämä voi heikentää otannan luotettavuutta ja validiteettia, sillä tulokset eivät välttämättä ole toistettavissa tai edustavia laajemmalle populaatiolle. Lisäksi kyselytutkimus on itsearviointimenetelmä, joka voi sisältää virheitä tai puutteita vastaajien muistin, rehellisyyden tai ymmärryksen vuoksi. Tutkimuksessa kysymysten muotoilu kehitettiin yhdessä ohjaajan avustuksella mahdollisimman selkeäksi ja helposti käytettäväksi, ja keskeistä terminologiaa avattiin kyselyn alussa väärinkäsitysten ehkäisemiseksi.

Tässä kandidaatintyössä käsiteltiin suhteellisen uutta aihetta, josta ei ole vielä paljon tutkimusta. Tutkimuksella pyrittiin avaamaan uusia näkökulmia ja mahdollisuuksia aihepiirissä. Kandidaatintyö on laajuudeltaan rajallinen, joka osaltaan vaikuttaa empiirisen menetelmän valintaan ja kattavuuteen sisällön osalta. Jatkotutkimuksen avulla tutkimusasetelmaa voidaan kehittää suuremmalla ja satunnaisemmalla otoksella, joka kattaisi laajemman kohderyhmän. Tämä voisi auttaa vahvistamaan tai kumoamaan tulokset ja antamaan niille enemmän painoarvoa. Lisäksi voitaisiin hyödyntää muita menetelmiä kuin kyselytutkimusta, jotka voivat syventää ymmärrystä ilmiöstä ja sen taustatekijöistä. Jatkotutkimuksessa voitaisiin myös tarkastella kriittisesti kyselyn kysymyksiä ja vastausvaihtoehtoja, ja systemaattisesti kehittää näitä osa-alueita esimerkiksi selkeyden, yksiselitteisyyden ja objektiivisuuden suhteen.

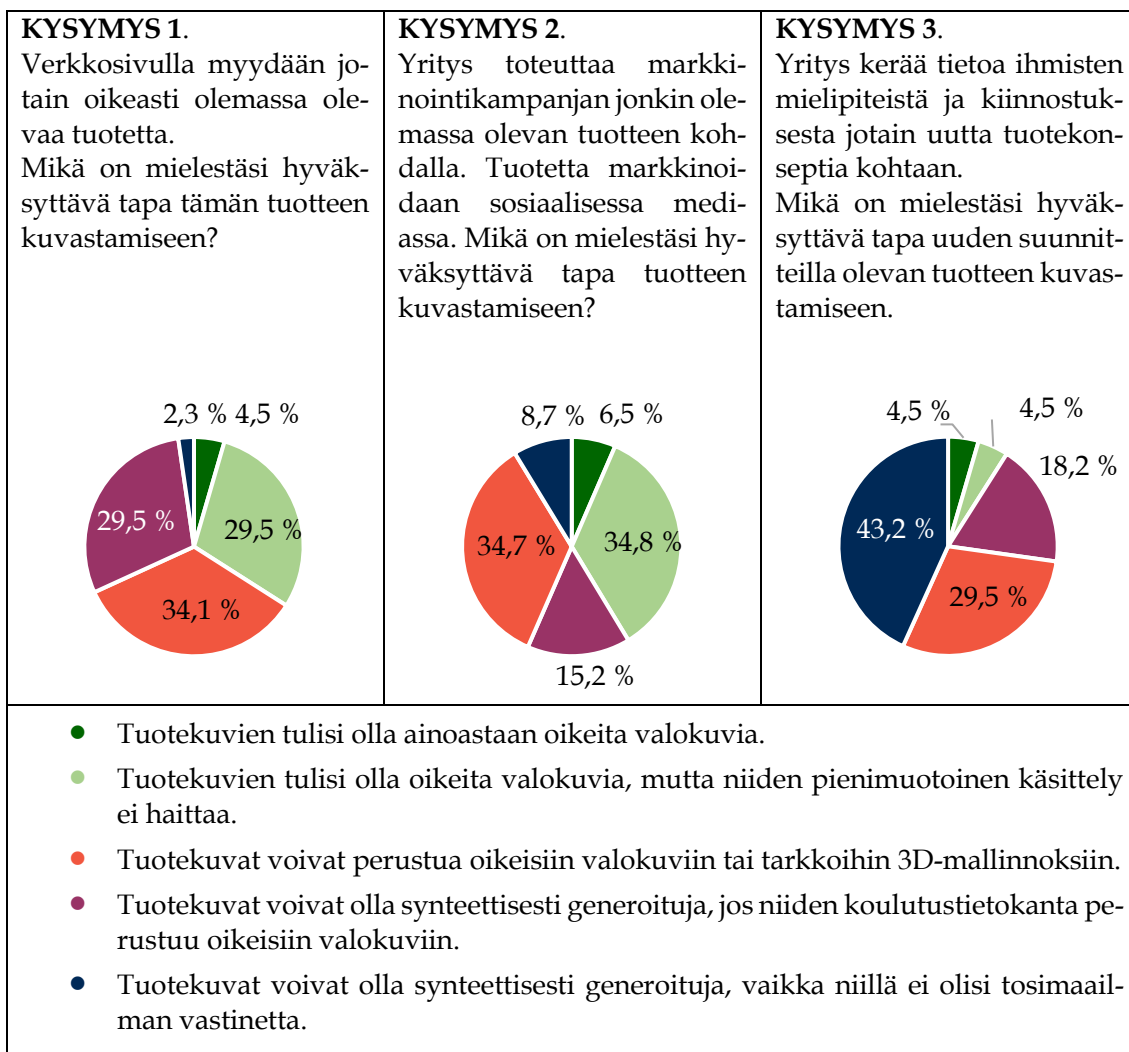
3.3 Kyselytutkimustulokset

Tämä alaluku sisältää yhteenvedon mielipidekyselystä (liite 12). Tutkimukseen osallistuneista valtaosa (15 vastaajaa) oli sitä mieltä, että tuotekuvien pitäisi olla

oikeita valokuvia tai tarkkoja 3D-mallinnoksia, jos niitä käytetään jonkin olemassa olevan tuotteen kaupitteluun (ks. kuvio 12, kysymys 1). Toiseksi eniten ääniä (13 vastaajaa) sai vaihtoehto, joka hyväksyi ainoastaan oikeat valokuvat, mutta salli niiden pienimuotoisen editoinnin. Kolmanneksi (12 vastaajaa) tuli vaihtoehto, jonka mukaan tuotekuvat voivat olla synteettisesti generoituja niiden koulutustietokannan perustuessa todellisiin valokuviin. Loput vastaajista äänestivät vaihtoehtoa, jossa sallittiin ainoastaan oikeat muokkaamattomat valokuvat (2 vastaajaa), ja vastaavasti toinen ääripää (2 vastaajaa) hyväksyi täysin synteettisesti kuvastetun sisällön, millä ei olisi mitään reaali maailman vastinetta. Valtaosa salli siis kuvien käsittelyn tavalla tai toisella tai hyväksyi 3D-mallin tuotekuvastamiseen, vaikka yritys käyttäisi kuvaa tuotteen myymiseen sen omalla tuotesivullaan.

Yrityksen markkinoidessa jotain olemassa olevaa tuotetta sosiaalisessa mediassa, mielipiteet jakautuivat tasaisesti, ja 16 vastaajaa siirtyi kannattamaan vaihtoehtoa, jonka mukaan tuotekuvat saisivat olla kevyesti muokattuja valokuvia (ks. kuvio 12, kysymys 2). Vastaavasti 16 vastaajaa äänesti vaihtoehtoa, jonka mukaan tuotekuvat voisivat olla synteettisesti generoituja, kunhan ne perustuisivat valokuviin. Vastaajista seitsemän hyväksyi valokuvat tai 3D-mallit tämänkaltaiseen markkinoimiseen, ja neljä hyväksyi niiden täyden syntetisoimisen. Yksi vastaajista hyväksyi ainoastaan oikeat käsittelemättömät valokuvat markkinoimiseen sosiaalisessa mediassa. Suhteellisen moni siis hyväksyi synteettisesti generoidun sisällön käyttämisen markkinointivälineenä. Lisäksi tarkkaakin 3D-mallia voidaan käsitellä latentilla diffuusiomallilla, ja siihen voidaan näin generoida esimerkiksi tausta tai siinä olevat materiaalit.

Jos yritys kerää tietoa ihmisten mielipiteistä ja kiinnostuksesta jonkin uuden tuotekonseptin osalta (ks. kuvio 12, kysymys 3), reilu enemmistö (19 vastaajaa) hyväksyi täysin synteettisesti luodut kuvat, vaikka niitä ei olisi koulutettu erikseen mihinkään tiettyyn reaali maailman tuotteeseen pohjautuen. 13 vastaajista hyväksyi toimintaan syntetisoidut kuvat, jotka pohjautuisivat oikeisiin valokuviin. Kahdeksan vastaajaa hyväksyi ainoastaan valokuvat tai tarkat 3D-mallit tuotekuvastamiseen. Äänistä loput jakautuivat tasaisesti oikeiden valokuvien ja pienimuotoisen editoinnin hyväksyvien (2 vastaajaa), sekä ainoastaan käsittelemättömiin valokuviin (2 vastaajaa) äänestäviin henkilöihin.

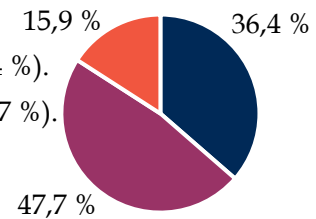


KUVIO 12 Mielipiteet tuotekuvastamisesta.

Kun kysyttiin, tulisiko markkinointimateriaalissa olla mainittuna sen sisällön synteettisestä luonteesta, enemmistö (21 vastaajaa) vastasi kyllä, jos sisällöllä pyritään fotorealismiin (ks. kuvio 13, kysymys 4). Toiseksi eniten ääniä sai vaihtoehto, joka kannatti merkintää myös vahvasti tyylyteltyjen kuvien kohdalla. Tyylytelty kuva voisi muistuttaa esimerkiksi öljymaalausta. Vastaajista seitsemän ei pitänyt erillistä mainintaa tarpeellisena. Tutkimuksessa kysyttiin myös millainen maininnan pitäisi olla, johon enemmistö (28 vastaajaa) äänesti maininnaksi kuvaan tai kuvaukseen sisällytettyä tekstiä. 12 ääntä sai puolestaan vaihtoehto, joka kannatti tuotekuviin selkeän vesileiman sisällyttämistä (ks. kuvio 14, kysymys 5).

KYSYMYS 4. Tulisiko markkinointimateriaaleissa olla mainittuna, mikäli sen sisältö on synteettisesti generoitua?

- Kyllä, myös vahvasti tyylitellyissä kuvissa (16 ääntä, 36,4 %).
- Kyllä, mutta vain tähdätessä fotorealismiin (21 ääntä, 47,7 %).
- Ei ole tarvetta mainita (7 ääntä, 15,9 %).



KYSYMYS 5. Jos vastasit aiempaan kysymykseen kyllä, millainen maininnan olisi syytä olla?



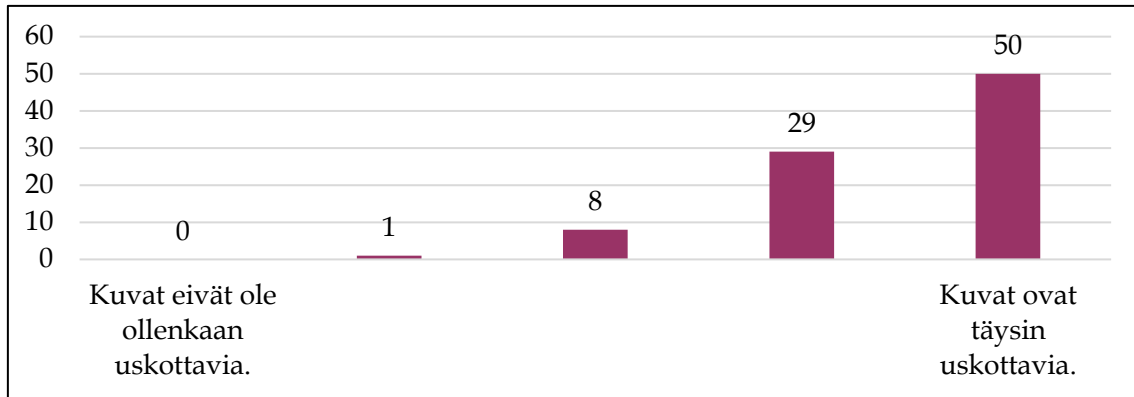
28 vastaajaa.

KUVIO 13 Mielenpitoet synteettisesti generoidusta sisällöstä.

Kun kyselytutkimuksessa siirryttiin tarkastelemaan synteettisesti generoitua sisältöä, tulokset olivat suhteellisen positiivisia kuvamateriaalin valmisteluhetken tekniikan rajoituksista huolimatta. Sen sijaan että vastaajilta olisi kysytty suoraan onko tuotettu sisältö heidän mielestään fotorealista vai ei, he saivat antaa mielipiteensä Likert-asteikolla numeroiden 1–5 väliltä. Näissä kysymyksissä nro. 1 viittasi fotorealismien epäonnistumiseen ja nro. 5 puolestaan sen täydelliseen onnistumiseen. Tutkimuskuvat generoitiin viittä eri kehote- ja generoimistekniikkaa hyödyntäen. Näitä olivat:

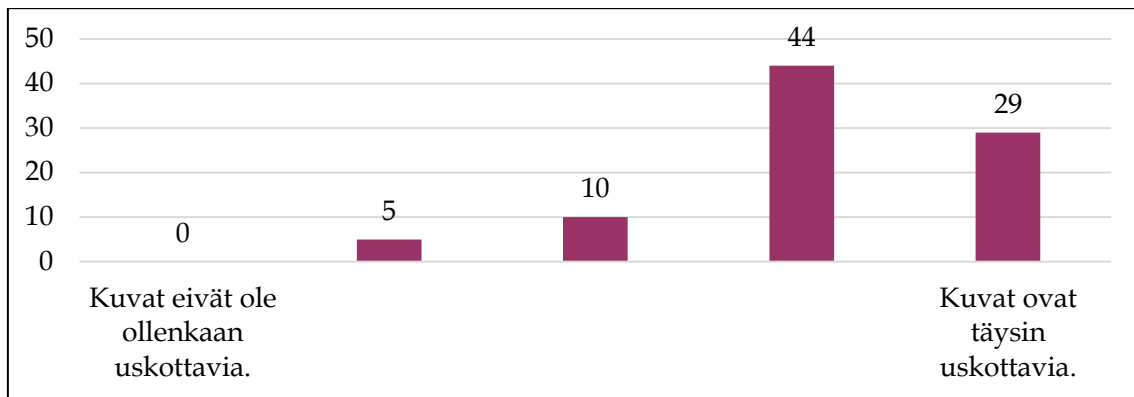
1. Tekstipohjainen generoiminen olemassa olevan tuotesävyyn perusteella (kehote vastaava kuin mitä koulutustietokannassa käytetty).
2. Tekstipohjainen generoiminen uuden tuotesävyyn perusteella (kehotteessa väri, jota ei koulutustietokannassa).
3. Tekstipohjainen -generoiminen, superresoluutio ja Inpainting.
4. Kuvapohjainen generoiminen olemassa olevan tuotesävyyn perusteella (kehotteessa ja kuvassa tuotesävy, joka löytyy myös koulutustietokannasta).
5. Kuvapohjainen generoiminen keskityn tuotesävyyn perusteella (kehotteessa ja kuvassa väritys, jota ei ole itse määritetty koulutustietokantaan, vaan se kuuluu mallin alkuperäisiin kyvykkyyksiin).

Parhaat arviot (ks. kuvio 14) saivat *Image-to-Image*-tekniikalla generoidut kuvat, vaikka ne generoitiin keksityillä tuotesävyillä (liite 12, yhteenveto kysymysten 10 & 11 tuloksista). Näissä kuvissa pohjustava kuva auttoi kuvien mittasuhteissa ja sävytyksessä, ja ne olivat myös suhteellisen korkealla resoluutiolla (960 x 960 px) generoituja.



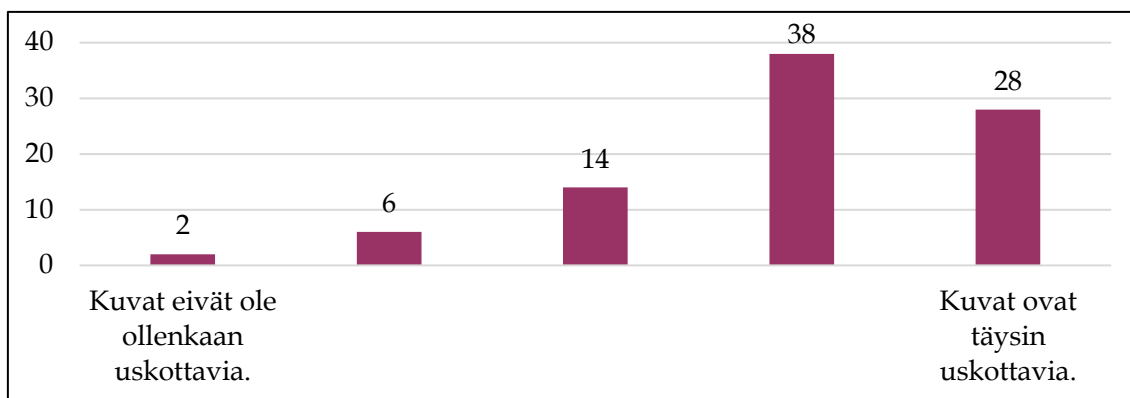
KUVIO 14 *Image-to-Image* kuvien uskottavuus uusilla tuotesävyillä.

Toiseksi parhaaksi menetelmäksi (ks. kuvio 15) valikoitui *Text-to-Image*-generoiminen olemassa oleviin tuotesävyihin ja tausta-asetelmiin perustuen (liite 12, yhteenveto kysymysten 6 & 7 tuloksista). Näissä kuvissa generoiminen suoritettiin samoilla kehoitteilla kuin mitä koulutuksessa käytettiin kuvauksina.



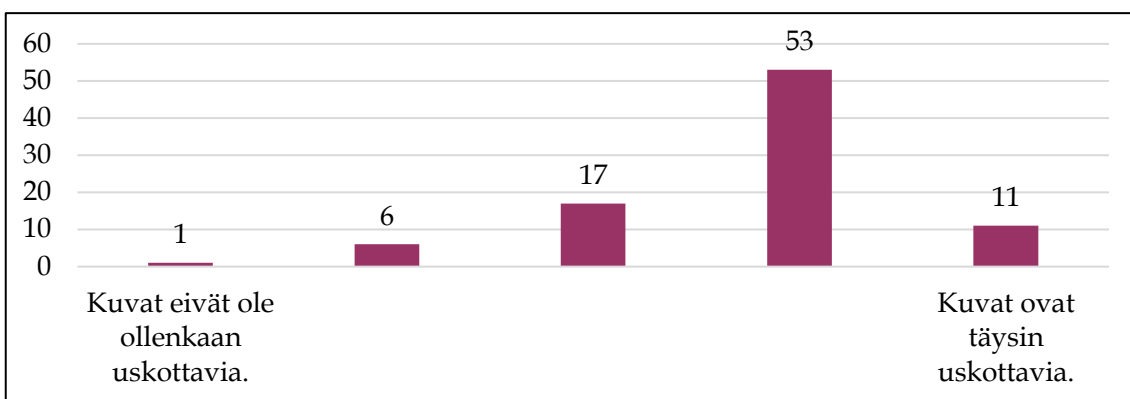
KUVIO 15 *Text-to-Image* kuvien uskottavuus olemassa olevilla tuotesävyillä ja tausta-asetelmilla.

Kolmanneksi parhaaksi tekniikaksi äänestettiin kuvat (ks. kuvio 16), jotka generoitiin *Text-to-Image*-tekniikalla keksittyjen tuotesävyjen perusteella (liite 12, yhteenveto kysymysten 12 & 13 tuloksista). Kuvien tausta ja sävytys poikkesi koulutustietokannasta). Näissä tuloksissa generoiminen suoritettiin koulutustietokannan mukaisella 768 x 768 px resoluutiolla, sillä suuremmissa kokoluokissa kuvien mittasuhteet epäonnistuivat, eivätkä ne olleet julkaisukelpoisia. Tämä luultavasti johtui siitä, ettei mittasuhteita normalisoitu generoitavaan kuvaan.



KUVIO 16 *Text-to-Image* kuvien uskottavuus keksityn tuotesävyyn perusteella.

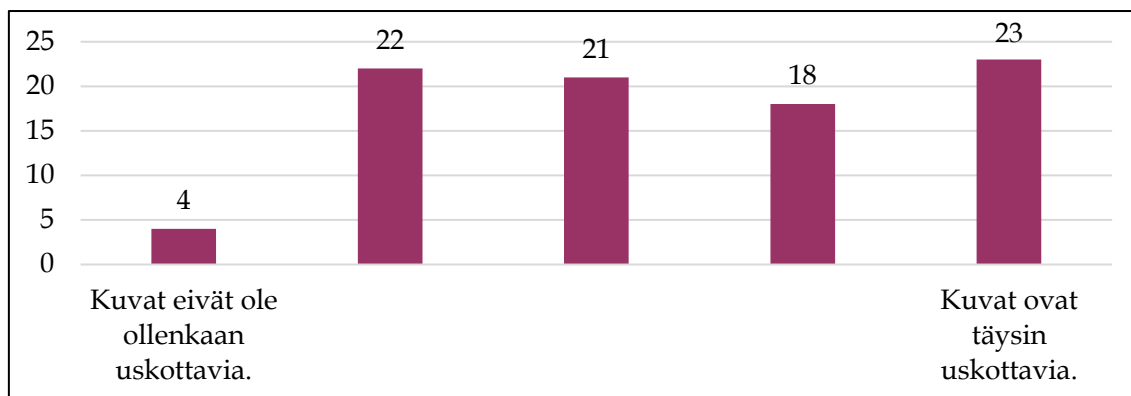
Neljäntenä (ks. kuvio 17) olivat *Image-to-Image*-tekniikalla tuotetut kuvat olemassa oleviin tuotesävyihin perustuen (liite 12, yhteenveto kysymysten 8 & 9 tuloksista). Toisaalta kuvien realismi arvioitiin vain suhteellisen fotorealistiseksi täydellisen onnistumisen sijaan. Näissä kuvissa yksityiskohdat, kuten puun syyt tuottivat generoimisessa ongelmia. Lisäksi kuvissa generoitua kuvaa verrattiin suoraan koulutustietokannan materiaaliin, jolloin kuvien epäkohdat olivat selkeämmin havaittavissa.



KUVIO 17 *Image-to-Image* kuvien uskottavuus olemassa olevaan tuotesävyyn perustuen.

Heikoimmat tulokset (ks. kuvio 18) saavuttivat superresoluutiolla ja Inpainting-tekniikalla valmistetut kuvat (liite 12, yhteenveto kysymyksistä 14 & 15). Niissä hajauma vastauksissa oli selkeämpi heikomman onnistumisen kannalta. Kuvien generoimisessa käytettiin *Text-to-Image* menetelmää pohjien generoimisessa. Tämän jälkeen niille suoritettiin superresoluutio. Tämän jälkeen kuvista rajattiin 768 x 768 px alue, jonka päälle tuote generoitiin Inpainting-menetelmällä. Ongelmia on etenkin kuvien mittasuhteissa ja yleisessä koherenssissa. Kuvien sävytys ja valotus on työlästä saada yhtäläiseksi, etenkin jos kuvaan generoidaan kokonaisia elementtejä lisää. Epäkohtana tässä menetelmässä oli, että kuvan olisi voinut generoida uudestaan *Image-to-Image*-metodilla, jolloin epäkohtia olisi saattanut voida korjata. Kuvan resoluutio oli kuitenkin sen verran laaja, ettei näytönohjain videomuisti riittänyt prosessin suorittamiseksi. Myöhemmin ControlNet on

esimerkiksi Tile-upscale-, Canny-reunakartta- ja Lineart -menetelmillään sallinut paljon suurempien resoluutioiden generoimisen ilman erillistä tarvetta Inpainting toiminnolle.



KUVIO 18 Superresoluutio- ja Inpainting-kuvien uskottavuus.

Näiden kysymysten lisäksi vastaajilta kysyttiin vapaavalintaisien kysymysten muodossa mikä kuvissa onnistui ja mikä niissä epäonnistui (liite 12, vapaavalintaiset kysymykset). Valtaosa vastaajista luokitteli kuvien kokonaisuuden onnistuneeksi, kun taas useimmissa kuvissa yksityiskohdat katsottiin epäonnistuneiksi. Mielipiteet vaihtelivat tekniikan mukaan, ja niiden perusteella on muodostettu johtopäätöksiä tekniikoista kaavioiden 1–5 esittelyihin. Superresoluutiokuvissa kuvien taustat luokiteltiin heikommin menestyneiksi. Useimmat luokittelivat kuvien mittasuhteet onnistuneiksi, vaikka generoimishetkellä ControlNet-lisäosa ei ollut normalisoimassa niitä. ControlNetin kehittämisen myötä superresoluution tarve on nykyisin vähäisempi.

3.4 Kyselytutkimuksen yhteenveto

Ihmisten mielipiteet ja asenteet LDM:lla generoiduista tuotekuvista ja niiden käytöstä markkinointiviestinnässä vaihtelevat tilanteen ja tekniikan mukaan. Tutkimustulosten mukaan useimmat ihmiset hyväksyvät tuotekuvat, jos ne ovat tarkkoja 3D-malleja tai oikeita valokuvia, joita on korkeintaan muokattu kevyesti. Moni oli myös sitä mieltä, että kuville ei saisi tehdä mitään muutoksia. Synteettiset kuvat, jotka on generoitu valokuvien pohjalta tai ilman reaali maailman vastinetta saivat vähemmän hyväksyntää etenkin tilanteissa, jossa niillä kuvastettaisiin jotain reaali maailman tuotetta. Toisaalta markkinointitilanteissa, missä jotain uutta tuotetta kehitellään, suuri osa vastaajista hyväksyi myös täysin synteettisesti generoidut kuvat. Lisäksi jos kuvaan sallitaan pienimuotoinen muokkaaminen, voisi se tarkoittaa myös sitä, että olemassa olevaan kuvaan voidaan tehdä jotain korjauksia LDM:lla.

Kyselyyn vastanneet henkilöt haluavat myös, että kuvien synteettinen luonne mainitaan, etenkin jos kuvat pyrkivät fotorealismiin. Kuvien laadun

suhteen ihmiset arvioivat latenteilla diffuusiomalleilla generoidut kuvat suhteellisen fotorealistisiksi. Kasvava ymmärrys ja uudet tekniikat, kuten Noise Offset ja ControlNet ovat myös parantaneet huomattavasti generoimisella saavutettavaa laatua. Tutkimus osoittaa, että LDM:lla on mahdollista tuottaa korkealaatuisia, jopa miltei fotorealistisia ja monipuolisia tuotekuvia eri tarkoituksiin, mutta niiden käyttöön liittyy yhä eettisiä haasteita. Eettisesti olisi tärkeää ottaa huomioon kuluttajien oikeus saada tietää kuvien alkuperä ja mahdolliset manipulaatiot, sekä välttää harhaanjohtavaa tai petollista markkinointia. Kyselyn tulokset kuitenkin viittaavat tekniikoiden suureen potentiaaliin etenkin, jos niiden käyttämisestä ollaan avoimia, ja se suoritetaan rehellisesti ja totuutta vääristämättä.

4 FOTOREALISTISEN KUVANGENEROIMISEN EETTISYYS

Uusien tekniikoiden kehittyessä raja todellisten valokuvien ja keinotekoisesti generoidun sisällön välillä hälvenee. Synteettistä sisältöä on yhä vaikeampi erottaa autenttisesta, mikä voi avata useita eri mahdollisuuksia enemmän tai vähemmän hyväksyttävään toimintaan. Varjopuolenaan kuvangeneroiminen mahdollistaa todentuntoisen disinformaation tuottamisen. Vaikka kandidaatintyössä keskityttiinkin tuotekuvastamiseen LDM:lla, teknologian relevanssi kuvangeneroimisessa voi muuttua. Riippumatta teknologiasta, fotorealistisen kuvangeneroimisen eettisyys on keskeinen tarkasteltava tekijä aihetta käsiteltäessä. Tässä kappaleessa keskitytään fotorealistisen kuvangeneroimisen eettisiin näkökulmiin, ja siihen miten eettisyyttä voidaan huomioida tuotteita markkinoidessa.

4.1 Fotorealistinen tuotekuvastamisen eettisyys markkinoinnissa

Fotorealistinen tuotekuvastaminen latentilla diffuusiomallilla on potentiaalinen sovellus markkinoinnissa, sillä se voi tarjota asiakkaille tarkan ja houkuttelevan kuvan tuotteesta tai palvelusta. Esimerkiksi Google Shopping -palvelussa on käytetty latenttia diffuusiomallia virtuaaliseen sovitteluun, jossa asiakkaat voivat kokeilla erilaisia meikkejä ja aurinkolaseja omalla kasvokuvallaan ja nähdä, miltä ne näyttävät eri valaistuksissa ja kulmissa (Kemelmacher-Shlizerman, 2023; Rincon, 2023). Generoidut tuotekuvat ovat yksilöllisimpiä, ja voivat näin vastata paremmin asiakkaan tarpeisiin.

Fotorealistinen tuotekuvastaminen LDM:llä ei kuitenkaan ole täysin ongelmaton markkinoinnin näkökulmasta. Kuten Sipolainen (2021) on tutkinut, markkinointi voi olla harhaanjohtavaa tai erehdyttävää, jos se antaa väärän tai liioitellun kuvan tuotteesta tai sen ominaisuuksista. Tämä voi johtaa markkinointirikokseen tai petokseen, jotka ovat rangaistavia tekoja Suomen laissa (Sipolainen, 2021). Fotorealistinen tuotekuvastaminen LDM:llä voi olla harhaanjohtavaa,

jos se esimerkiksi piilottaa tai vähättelee tuotteen virheitä tai puutteita tai jos se ei kerro selvästi, että kyseessä on keinotekoisesti luotu kuva eikä valokuva.

Lisäksi fotorealistinen tuotekuvastaminen latentilla diffuusiomallilla herättää eettisiä kysymyksiä keinotekoisesta älyn hyödyntämisestä markkinoinnissa. Hermann (2022) on pohtinut, että keinotekoinen äly voi edistää yhteiskunnallista hyvää, jos sitä käytetään vastuullisesti ja läpinäkyvästi. Hän esittää neljä eettistä periaatetta keinotekoisesta älyn markkinointisovelluksille: arvokkuus luomisen kaikille sidosryhmille, yksityisyyden ja autonomian kunnioittaminen, oikeudenmukaisuus ja syrjimättömyys sekä kestävyys ja vastuullisuus (Hermann 2022). Fotorealistinen tuotekuvastaminen latentilla diffuusiomallilla voi rikkoa näitä periaatteita, jos se esimerkiksi loukkaa kuluttajien tai muiden osapuolten oikeuksia, etuja tai arvoja tai jos se aiheuttaa ympäristö- tai sosiaalisia haittoja.

Näin ollen fotorealistinen tuotekuvastaminen latentilla diffuusiomallilla on innovatiivinen ja lupaava menetelmä markkinoinnissa, mutta se vaatii myös huolellista harkintaa ja sääntelyä. Markkinoijien tulisi noudattaa hyvää markkinointitapaa ja kuluttajasuojalakeja sekä eettisiä periaatteita keinotekoisesta älyn käytössä. Kuluttajien tulisi olla tietoisia fotorealistisen tuotekuvastamisen mahdollisuuksista ja riskeistä sekä käyttää kriittistä ajattelua ja medialukutaitoa. Fotorealistinen tuotekuvastaminen latentilla diffuusiomallilla voi olla hyödyllinen työkalu markkinoinnissa, mutta se ei saa korvata rehellisyyttä, luottamusta ja vastuullisuutta.

4.2 UNESCO:n suositus tekoälyn eettisyydestä suhteessa fotorealistiseen tuotekuvastamiseen

UNESCO eli Yhdistyneiden kansakuntien kasvatus-, tiede- ja kulttuurijärjestö on YK:n erityisjärjestö, jonka tavoitteena on edistää rauhaa ja yhteistyötä koulutuksen, tieteen, kulttuurin ja viestinnän avulla. UNESCO:n jäsenvaltioita ovat kaikki YK:n jäsenvaltiot sekä muutamat muut valtiot, jotka ovat hakeneet ja saaneet jäsenyyden UNESCO:n yleiskokouksen äänestyksellä. Tammikuussa 2020 UNESCO:n jäsenvaltioita oli 193 ja assosioituneita jäsenvaltioita 113. Assosioituneet jäsenvaltiot ovat sellaisia alueita tai alueiden ryhmiä, jotka eivät ole kansainvälisesti tunnustettuja itsenäisiä valtioita⁸.

UNESCO:n suositus tekoälyn etiikasta (2021) oli ensimmäinen laatuaan maailmassa. Suositus hyväksyttiin UNESCO:n yleiskokouksessa marraskuussa 2021 yksimielisesti kaikkien 193 jäsenvaltion toimesta. Suosituksen tavoitteena on tarjota eettinen viitekehys tekoälyn kehittämiselle ja käytölle, joka perustuu ihmisoikeuksiin, ihmisarvoon ja yhteiseen hyvään. Suositus käsittelee tekoälyn eettisiä kysymyksiä ja haasteita, kuten vinoumia, syrjintää, ihmisoikeusloukkauksia ja ilmastohaittoja, sekä esittää arvoja, periaatteita ja toiminta-alueita, joiden avulla voidaan edistää vastuullista ja ihmisoikeuksia kunnioittavaa tekoälyn

⁸ <https://en.unesco.org/countries/member-states>

kehittämistä ja käyttöä eri aloilla. UNESCO:n suositus korostaa myös tekoälyn mahdollisuuksia ja riskejä yhteiskunnalle, ympäristölle ja ihmisille sekä tarvetta lisätä yhteistyötä ja solidaarisuutta eri maiden ja toimijoiden välillä tekoälyn globaalien hallinnan ja oikeudenmukaisen jakamisen varmistamiseksi. Suositus perustuu kansainväliseen oikeuteen, erityisesti ihmisoikeuslakiin, sekä UNESCO:n aiempiin asiakirjoihin, kuten YK:n yleismaailmalliseen ihmisoikeuksien julistukseen (*Universal Declaration of Human Rights*) (1948) ja Yleismaailmalliseen bioetiikan ja ihmisoikeuksien julistukseen (*Universal Declaration on Bioethics and Human Rights*) (2005), jotka muodostavat suosituksen normatiivisen perustan. Suositus on laadittu myös globaalilla lähestymistavalla, joka ottaa huomioon eri maiden ja alueiden erilaiset olosuhteet, kulttuurit ja arvot sekä pyrkii ohjaamaan tekoälyteknologioita kohti yhteisiä tavoitteita ja yhteistä hyvää. Teoksen sisältöä voidaan suhteuttaa hyvin kyselytutkimuksen havaintoihin muun muassa ihmisten toiveista toiminnan avoimuudesta. UNESCO:n (2021) tekstissä määritellään tekoälyn käyttämisen avoimuudesta seuraavasti:

- Avoimuus: tekoälyn toimintaperiaatteet, algoritmit, tietolähteet ja vaikutukset ovat ymmärrettäviä ja selitettävissä kaikille asianosaisille.
- Avoimuus edellyttää, että tekoälyn kehittämiseen ja käyttöön osallistuvat tahot ovat tunnistettavissa ja vastuussa tekojensa seurauksista.
- Avoimuus on välttämätöntä luottamuksen rakentamiseksi tekoölyyn sekä sen eettisen arvioinnin ja valvonnan mahdollistamiseksi.
- Avoimuus edistää myös tekoälyn moniarvoisuutta, moninaisuutta ja osallisuutta sekä ehkäisee tekoälyn väärinkäyttöä ja väärinkäsityksiä.

YK:n ihmisoikeuksien yleismaailmallinen julistus (1948) sekä UNESCO:n suositus tekoälyn etiikasta (2021) tukevat kuvangeneroimista etenkin, jos sen käyttämistä voidaan suhteuttaa taiteeseen ja koulutukseen. YK:n (1948) julistuksen 27. artiklassa todetaan kaksi ihmisoikeutta:

1. ”Jokaisella on oikeus vapaasti osallistua yhteiskunnan sivistyselämään, nauttia taiteista sekä päästä osalliseksi tieteen edistyksen mukanaan tuomista eduista”.
2. ”Jokaisella on oikeus niiden henkisten ja aineellisten etujen suojaamiseen, jotka johtuvat hänen luomastaan tieteellisestä, kirjallisesta tai taiteellisesta tuotannosta”.

UNESCO:n (2021) suositus määrittelee neljä avainarvoa, mitkä luovat perustan tekoälyä käyttävien järjestelmien eettisyyteen. Nämä arvot ovat luotu ihmisiä, yksilöitä, yhteisöjä ja ympäristöä ajatellen:

1. Ihmisoikeudet ja ihmisarvo: Kattaa ihmisoikeuksien ja perusvapauksien sekä ihmisarvon kunnioittamisen, suojelun ja edistämisen.
2. Rauhanomainen yhteiselo: Oikeudenmukaiset ja toisiinsa liittyvät yhteiskunnat.
3. Monimuotoisuuden ja osallisuuden varmistaminen.

4. Ympäristön ja ekosysteemin kestävyys.

”Jäsenvaltioiden tulisi edistää tekoälyn opetusta ja digitaalista koulutusta taiteilijoille ja luovan alan ammattilaisille, jotta he voivat arvioida tekoälyteknologioiden soveltuvuutta omassa ammatissaan sekä osallistua sopivien tekoälyteknologioiden suunnitteluun ja käyttöönottoon. Tekoälyteknologioita käytetään luomaan, tuottamaan, jakamaan, lähettämään ja kuluttamaan monenlaisia kulttuurisia tuotteita ja palveluita, ja samalla otetaan huomioon kulttuuriperinnön, monimuotoisuuden ja taiteellisen vapauden säilyttämisen tärkeys.” (UNESCO, 2021, artikla 96)

UNESCO:n (2021) suosituksessa määritellään myös kymmenen ydinperiaatetta ihmisoikeuskeskeiselle lähestymistavalle tekoälyn etiikan tarkasteluun, mitä voidaan soveltaa myös fotorealistisessa tuotekuvastamisessa.

1. Suhteellisuus ja vahingoittamattomuus: Tekoälyjärjestelmien käyttö ei saa ylittää sitä, mikä on tarpeen oikeutetun tavoitteen saavuttamiseksi. Riskinarviointia tulisi käyttää tällaisesta käytöstä mahdollisesti aiheutuvien haittojen ehkäisemiseksi (UNESCO, 2021). Fotorealistisen tuotekuvastamisen käyttö tulisi olla suhteellista sen tavoitteeseen, eli järjestelmän tulisi voida luoda uskottavia ja tarkkoja kuvia tuotteista ilman fyysistä mallia tai valokuvausta. Tämä voi olla tarpeen esimerkiksi verkko-kaupassa, jossa asiakkaat haluavat nähdä tuotteet ennen ostamista. Fotorealistisen tuotekuvastamisen mahdollisesti aiheuttamat haitat tulisi arvioida ja ehkäistä, esimerkiksi harhaanjohtavien tai loukkaavien kuvien luominen, asiakkaiden pettymys tai luottamuksen menetys tai datan väärinkäyttö tai väärinkäsitys.
2. Turvallisuus: Tekoälytoimijoiden tulisi välttää ei-toivottuja haittoja (turvallisuusriskit) ja hyökkäyshaavoittuvuuksia (turvariskit) ja niihin on puututtava (UNESCO, 2021). Fotorealistisen tuotekuvastamisen tulisi olla turvallista sekä teknisesti että eettisesti. Teknisesti se tarkoittaa, että fotorealistisen tuotekuvastamisen menetelmät ja algoritmit ovat luotettavia, tehokkaita ja virheettömiä, eivätkä sisällä haavoittuvuuksia tai riskejä, jotka voisivat johtaa ei-toivottuihin tai vaarallisiin tuloksiin. Eettisesti se tarkoittaa, että fotorealistisen tuotekuvastamisen käyttö on vastuullista, rehellistä ja läpinäkyvää, eikä aiheuta haittaa ihmisille, ympäristölle tai yhteiskunnalle.
3. Oikeus yksityisyyteen ja tietosuojaan: Yksityisyyttä on suojeltava ja edistettävä tekoälyn koko elinkaaren ajan. Järjestelmille on myös luotava asianmukaiset tietosuojakehykset (UNESCO, 2021). Fotorealistisen tuotekuvastamisen tulisi kunnioittaa ja edistää yksityisyyttä ja tietosuojaa koko sen elinkaaren ajan. Tämä tarkoittaa, että fotorealistisen tuotekuvastamisen datan kerääminen, käyttö ja jakaminen on tehty asianmukaisesti ja laillisesti, eikä loukkaa kenenkään henkilökohtaisia tai immateriaalisia oikeuksia. Lisäksi fotorealistisen tuotekuvastamisen datan säilyttämiseen, poistamiseen ja muokkaamiseen on oltava selkeät ja turvalliset käytännöt. Jos koulutusdataa kerätään asiakkailta, se on syytä

säilöä esimerkiksi LoRA-painojen muodossa, jotta sen poistaminen generoimistietokannasta on tarvittaessa mahdollista.

4. Useita sidosryhmiä koskeva ja mukautuva hallinto ja yhteistyö: Tietojen käytössä on kunnioitettava kansainvälistä oikeutta ja kansallista suvereniteettia. Lisäksi eri sidosryhmien osallistuminen on välttämätöntä tekoälyn hallintaan liittyvien osallistavien lähestymistapojen kannalta (UNESCO, 2021). Fotorealistisen tuotekuvastamisen hallintoihin ja yhteistyöhön tulisi osallistua useita sidosryhmiä, kuten tuottajia, asiakkaita, viranomaisia, tutkijoita ja kansalaisyhteiskuntaa. Tämä edellyttää avointa ja osallistavaa vuoropuhelua sekä kansainvälistä että kansallista tasolla tekoälyn eettisistä kysymyksistä ja niiden ratkaisemisesta. Lisäksi fotorealistisen tuotekuvastamisen datan käytössä tulisi kunnioittaa kansainvälistä oikeutta ja kansallista suvereniteettia sekä ottaa huomioon eri kulttuuriset ja sosiaaliset arvot ja normit.
5. Vastuullisuus ja järjestelmävastuu: Tekoälyjärjestelmien tulee olla tarkastettavissa ja jäljitettävissä. Valvonta-, vaikutustenarviointi-, auditointi- ja muita asianmukaisia tarkastusmekanismeja tulisi käyttää, jotta vältetään ristiriidat ihmisoikeusnormien kanssa, ja ehkäistään ympäristön hyvinvointiin kohdistuvat uhat (UNESCO, 2021). Fotorealistisen tuotekuvastamisen tulisi olla vastuullista ja järjestelmävastaavaa, eli sen tulisi osoittaa, miten se noudattaa eettisiä periaatteita ja käytäntöjä sekä miten se vastaa mahdollisista haitoista tai virheistä. Tämä edellyttää, että fotorealistisen tuotekuvastamisen menetelmät, algoritmit ja tulokset ovat tarkastettavissa ja jäljitettävissä, eli niiden alkuperä, prosessi ja validointi ovat selkeästi dokumentoituja ja saatavilla.
6. Läpinäkyvyys ja selitettävyys: Tekoälyjärjestelmien eettinen (T&E) käyttöönotto riippuu niiden läpinäkyvyydestä ja selitettävyydestä. T&E:n tason tulisi olla kontekstiin sopiva, koska T&E:n ja muiden periaatteiden, kuten yksityisyyden ja turvallisuuden, välillä voi olla jännitteitä (UNESCO, 2021). Fotorealistisen tuotekuvastamisen tulisi olla läpinäkyvää ja selitettävää, eli sen tulisi ilmoittaa selkeästi ja ymmärrettävästi sen menetelmistä, tavoitteista ja rajoituksista sekä niiden vaikutuksista. Tämä tarkoittaa, että fotorealistisen tuotekuvastamisen käyttö on merkitty asianmukaisesti, esimerkiksi ilmoittamalla, että kyseessä on generoitu kuva, ja että asiakkaat saavat riittävästi tietoa siitä, miten kuva on luotu ja miten se vastaa todellista tuotetta. Lisäksi fotorealistisen tuotekuvastamisen tulisi tarjota selityksiä sen toiminnasta ja tuloksista, eli sen tulisi kertoa, miksi se on luonut tietyn kuvan, ja mitkä ovat sen taustalla olevat logiikat ja kriteerit.
7. Ihmisen valvonta ja päättäväisyys: Jäsenvaltioiden olisi varmistettava, että tekoälyjärjestelmät eivät syrjäytä perimmäistä ihmisen vastuuta ja vastuuvollisuutta (UNESCO, 2021). Fotorealistisen tuotekuvastamisen tulisi olla ihmisen valvonnan ja päättäväisyyden alaisena, eli sen tulisi antaa ihmisille mahdollisuus ohjata, valvoa ja keskeyttää sen toimintaa sekä tehdä lopulliset päätökset sen käytöstä. Tämä tarkoittaa, että

fotorealistisen tuotekuvastamisen käyttö ei saa syrjäyttää perimmäistä ihmisen vastuuta ja vastuuvollisuutta eikä vaarantaa ihmisten itsemääräämisoikeutta tai autonomiaa. Lisäksi fotorealistisen tuotekuvastamisen tulisi kunnioittaa ihmisten arvoja, mieltymyksiä ja valintoja. Näissä järjestelmissä olisi myös valmistajan vastuulla tarkistaa tilattavien tuotteiden oikeellisuus ennen niiden valmistamista. Jos järjestelmä generoi jotain tuotteen yksityiskohtia virheellisesti, niistä tulisi huomauttaa sekä asiakasta että kuvantamisjärjestelmän kehittäjätahoa, jotta ongelmat saataisiin ratkaistu mahdollisimman tehokkaasti.

8. Kestävyys: Tekoälyteknologioita tulisi arvioida suhteessa niiden vaikutuksiin kestävyys suhteen, joka ymmärretään joukkona jatkuvasti kehittyviä tavoitteita, mukaan lukien YK:n kestävä kehityksen tavoitteissa asetetut tavoitteet (UNESCO, 2021). Fotorealistisen tuotekuvastamisen tulisi edistää kestävyttä sekä ympäristön että yhteiskunnan kannalta. Tämä tarkoittaa, että fotorealistisen tuotekuvastamisen vaikutuksia kestävyteen tulisi arvioida suhteessa niiden hyötyihin ja haittoihin sekä lyhyellä että pitkällä aikavälillä. Lisäksi fotorealistisen tuotekuvastamisen tulisi noudattaa kestävä kehityksen periaatteita ja tavoitteita, kuten YK:n kestävä kehityksen tavoitteissa asetetut tavoitteet, sekä minimoida sen aiheuttamat ympäristövaikutukset ja -riskit.
9. Tietoisuus ja lukutaito: Tekoälyn ja datan yleistä ymmärrystä tulisi edistää avoimen ja saavutettavan koulutuksen, kansalaistoiminnan, digitaalisten taitojen ja tekoälyn etiikkakoulutuksen sekä media- ja tietolukutaidon avulla (UNESCO, 2021). Fotorealistisen tuotekuvastamisen tulisi lisätä tietoisuutta ja lukutaitoa tekoälystä ja datasta sekä niiden mahdollisuuksista ja haasteista. Tuotekuvastamisen tulisi tarjota avointa ja saavutettavaa koulutusta, kansalaistoimintaa, digitaalisia taitoja ja tekoälyn etiikkakoulutusta sekä media- ja tietolukutaitoa kaikille sidosryhmille, jotka ovat mukana tai vaikuttaneita sen käytöstä. Tämä voi auttaa lisäämään ymmärrystä siitä, miten fotorealistinen tuotekuvastaminen toimii, mitä hyötyjä ja riskejä siihen liittyy, ja miten sitä voidaan käyttää vastuullisesti ja kestävästi. Lisäksi fotorealistisen tuotekuvastamisen tulisi edistää kriittistä ajattelua ja eettistä pohdintaa sen eettisistä kysymyksistä ja vaikutuksista.
10. Oikeudenmukaisuus ja syrjimättömyys: Tekoälytoimijoiden tulisi edistää sosiaalista oikeudenmukaisuutta ja syrjimättömyyttä ja sen käyttöön on omaksuttava kattava lähestymistapa varmistukseksi, että tekoälyn edut ovat kaikkien saatavilla (UNESCO, 2021). Fotorealistisen tuotekuvastamisen tulisi edistää sosiaalista oikeudenmukaisuutta, eli sen tulisi pyrkiä vähentämään eriarvoisuutta ja epätasa-arvoa sekä edistämään yhteiskunnallista hyvinvointia ja osallisuutta. Fotorealistinen tuotekuvastaminen ei myöskään saa syrjäyttää tai syrjiä ketään sen käytön tai vaikutusten perusteella. Fotorealistisen tuotekuvastamisen tulisi edistää oikeudenmukaisuutta, eli sen tulisi noudattaa oikeudenmukaisia ja eettisiä sääntöjä ja normeja sekä kunnioittaa kaikkien oikeuksia ja velvol-

lisuuksia. Tämä tarkoittaa, että fotorealistisen tuotekuvastamisen tulisi olla reilua ja tasapuolista sekä asiakkaiden että tuottajien kannalta, eikä se saa aiheuttaa haittaa tai vahinkoa kummallekaan osapuolelle. Fotorealistisen tuotekuvastamisen tulisi edistää syrjimättömyyttä, eli sen tulisi ottaa huomioon ja kunnioittaa ihmisten monimuotoisuutta ja erilaisuutta sekä välttää ennakkoluuloja, stereotyyppioita tai stigmatisointia. Tämä tarkoittaa, että fotorealistisen tuotekuvastamisen tulisi olla inkluusiivista ja sensitiivistä eri asiakasryhmien tarpeille, mieltymyksille ja arvoille, eikä se saa loukata tai alentaa ketään sen kuvien tai viestien perusteella.

4.3 Fotorealististen tuotekuvien generoiminen & ECCOLA-menetelmä

Organisaatio, joka haluaa käyttää LDM:ää markkinointimateriaalin tuottamiseen, joutuu pohtimaan tekniikan eettisiä näkökohtia. Yksi tapa tähän on käyttää ECCOLA-kortteja, jotka on suunniteltu tarjoamaan erilaisia näkökulmia eettisten kysymysten herättämiseksi tekoälyapplikaation kehitysprosessin aikana. ECCOLA-kortit sisältävät kahdeksan teemaa ja useita erillisiä aiheita, jotka sopivat käsiteltävään teemaan (ks. taulukko 1). Kehittäjät voivat hyödyntää ECCOLA-korttien kysymyksiä toteuttaakseen eettisiä näkökulmia huomioivia toimenpiteitä ohjelmistokehityksessä (Vakkuri ym., 2021).

TAULUKKO 1 ECCOLA-teemat ja aiheita (mukaillen Vakkuri ym., 2021).

Teema	Aihe
Tietosuoja	Henkilötiedot, yksityisyys, salassapito
Turvallisuus	Vahingot, vastuu, luottamus
Oikeudenmukaisuus	Syrjintä, tasa-arvo, monimuotoisuus
Hyvinvointi	Hyödyt, haitat, vaikutukset
Autonomia	Valinnanvapaus, manipulointi, riippuvuus
Läpinäkyvyys	Selitykset, perustelut, todisteet
Vastuullisuus	Eettiset periaatteet, arvot, normit
Osallistuminen	Osallisuus, vuorovaikutus, yhteistyö

LDM:llä fotorealistista markkinointimateriaalia tuottaessa voisi soveltaa esimerkiksi seuraavia kysymyksiä ECCOLA-kortteihin pohjautuen. On kuitenkin yrityksen vastuulla tuottaa omat vastauksensa kysymyksiin, ja määrittää muut järjestelmälle relevantit kysymykset ennen sen implementoimista:

- Tietosuojaja: Henkilötiedot.
 - Miten varmistetaan, että tuotekuvastossa ei esiinny henkilötietoja ilman asianomaisten suostumusta?
 - Miten käsitellään henkilötietoja sisältäviä syötteitä tai tuloksia?
- Turvallisuus: Luottamus.
 - Miten varmistetaan, että tuotekuvasto ei sisällä harhaanjohtavia tai epätarkkoja kuvia?
 - Miten osoitetaan asiakkaille ja muille sidosryhmille, että tuotekuvasto on luotettava ja laadukas?
- Oikeudenmukaisuus: Syrjintä.
 - Miten varmistetaan, että tuotekuvasto ei sisällä syrjiviä tai loukkaavia kuvia?
 - Miten huomioidaan erilaiset asiakasryhmät ja heidän tarpeensa ja mieltymyksensä?
- Hyvinvointi: Vaikutukset.
 - Miten arvioidaan tuotekuvaston vaikutuksia asiakkaiden ja muiden sidosryhmien hyvinvointiin?
 - Miten minimoidaan kuvangeneroimisen mahdolliset negatiiviset vaikutukset ja maksimoidaan mahdolliset positiiviset vaikutukset?
- Autonomia: Manipulointi.
 - Miten varmistetaan, että tuotekuvasto ei manipuloi asiakkaita tai muita sidosryhmiä ostamaan tai tekemään jotain, mitä he eivät halua tai tarvitse?
 - Miten kunnioitetaan asiakkaiden ja muiden sidosryhmien itsemääräämisoikeutta ja valinnanvapautta?
- Läpinäkyvyys: Selitykset.
 - Miten kerrotaan asiakkaille ja muille sidosryhmille, että tuotekuvasto on luotu LDM:llä?
 - Miten selitetään, miten LDM toimii ja millaisia syötteitä ja ohjausmekanismeja se käyttää?
- Vastuullisuus: Eettiset periaatteet.
 - Miten noudatetaan eettisiä periaatteita, jotka koskevat LDM:n käyttöä tuotekuvaston luomiseen?
 - Miten valitaan sopivat eettiset periaatteet eri tilanteisiin ja konteksteihin?
 - Miten taataan, että kuvangeneroiminen tapahtuu ekologisesti ja minimoiden sen ympäristövaikutuksia?
- Osallistuminen: Osallisuus.
 - Miten otetaan huomioon asiakkaiden ja muiden sidosryhmien mielipiteet ja palautteet tuotekuvastosta?
 - Miten mahdollistetaan heidän osallistumisensa tuotekuvaston kehittämiseen ja parantamiseen?

5 POHDINTA

Kuvangeneroimisen teknologiat kehittyvät jatkuvasti, ja niiden avulla voidaan luoda yhä realistisempia ja tarkempia kuvia. On mahdollista, että lähitulevaisuudessa saavutetaan absoluuttinen fotorealismi, eli kyky tuottaa kuvia, joita ei voi erottaa oikeista valokuvista. Tämä edellyttää kuitenkin paitsi generoitavien kuvien resoluution kasvattamista, myös niiden monipuolisuuden ja luovuuden lisäämistä. Lisäksi kuvangeneroimisen teknologioiden käytettävyys ja saatavuus vaikuttavat siihen, miten laajasti niitä voidaan hyödyntää eri aloilla ja tarkoituksissa. Tässä tutkimuksessa käytetyt ohjelmistotyökalut ovat uusia ja yksittäisten kehittäjien ylläpitämiä, joten niiden toimivuus ja standardoiminen voivat vielä muuttua paljolti. Voisi myös olettaa, että jatkossa diffuusiopohjaiset kuvangeneroimismenetelmät integroituvat paremmin tunnettuihin ja laajasti käytettyihin kuvankäsittelyohjelmistoihin. Tällä hetkellä Adobe Photoshopissa⁹ ja GIMP¹⁰:issä toiminnallisuutta voidaan hyödyntää lisäosien avustuksella, ja hiljattain Adobe julkaisi myös oman Generative Fill -toimintonsa¹¹, missä kuviin voidaan lisätä sisältöä generatiivisin menetelmin.

Yksi uusimmista potentiaalisesti etenkin generoimisen suoraviivaisuuteen vaikuttavista tekniikoista on ekspressiivinen tekstistä kuvaan generoiminen. Ekspressiivisyydellä tarkoitetaan, että tekstielementteihin voi itsessään määrittää generoimisen parametrejä eri fonttiasetuksia säätämällä (Ge, Park & Huang, 2023). Sen sijaan että kehoitteessa sanottaisiin: *"A red Pinja spoke chair in a green meadow dotted with burgundy flowers"*, voisi kehoite olla muotoa: *"**A Pinja spoke chair** in a **meadow** dotted with **flowers**"*. Tuolin fontti olisi punainen, niityn vihreä, ja kukkien viininpunainen, mikä indikoisi haluttua sävyä sen sijaan että se määritettäisiin kirjoitetussa muodossa. Kirjasintyyppi voi vastaavasti määrittää jotain generoimisen parametria. Tällöin generoitu kuva saattaisi vastata paremmin haluttua lopputulosta ilman, että tuotteeseen määritetty väri vaikuttaisi esimerkiksi kuvan taustaan tai muihin siinä esiintyviin elementteihin.

⁹ <https://github.com/AbdullahAlfaraj/Auto-Photoshop-StableDiffusion-Plugin>

¹⁰ <https://github.com/hamsolo474/stable-diffusion-gimp-plugin>

¹¹ <https://www.adobe.com/products/photoshop/generative-fill.html>

Kaupallisessa mielessä keskeisiä käyttökohteita kuvangeneroimisella voisi olla asiakkaan tai myyjän käytettävissä oleva järjestelmä, jossa voidaan valita ennalta koulutetusta tuotekatalogista haluttu tuote, ja sitten määrittää muut generoimisen parametrit. Tällaisella järjestelmällä voisi kuvastaa loputtomasti vaihtoehtoisia tuotteita, mistä voisi olla hyötyä etenkin räätälöitävien huonekalujen myymisessä. Huonekaluvalmistajilla voi olla useita vaihtoehtoisia kangastoimitajia, joista jokaisen valikoimassa saattaa olla satoja kankaita erilaisilla väreillä. Osassa tuotteista voidaan mahdollisesti myös yhdistää useita eri kankaita halutun lopputuloksen saavuttamiseksi. Tällaisissa tapauksissa on taloudellisesti kannattamatonta kuvata kaikkia erilaisia sävy- ja kangasyhdistelmiä, ja siispä kuvausprosessi suoritetaan yleensä vain muutamalla keskeisellä sävytyksellä. Ilman tuotekuvaa ostopäätös on satunnaisten olemassa olevien tuotekuvien, asiakkaan mielikuvituksen sekä myymälä- ja kangasmallien varassa. Siispä tilattavan tuotteen lopputuleman hahmottaminen voi olla haasteellista. Latenteilla diffuusiomalleilla on mahdollista tuottaa loputtomasti erilaisia variaatioita kuvastettavasta tuotteesta, ja näin generoida toivottuja kuvia aina tapauskohtaisesti. Halutessaan tekniikkaa voisi soveltaa myös oikeiden valokuvien kanssa, ja näin generoida tuote ympäristöön, minne sitä oikeasti ollaan hankkimassa. Toiminnallisuuksien vakaa implementoiminen vaatisi kuitenkin laajaa sovelluskehitystä, ja sen saavuttaminen tämänhetkisillä työkaluilla voi olla vaikeaa. Latenteja diffuusiomalleja voidaan soveltaa lukuisiin muihinkin tarkoituksiin, ja niistä voi olla apua kaikessa kuvallisen sisällön tuottamisessa.

Kuvangeneroimisen teknologioiden käyttö markkinointiviestinnässä edellyttää kuluttajien hyväksyntää ja luottamusta. Kuluttajien mielipiteet ja asenteet LDM:lla generoiduista tuotekuvista ja niiden käytöstä markkinointiviestinnässä vaihtelevat tilanteen ja tekniikan mukaan. Tämän tutkimuksen kyselytutkimuksen tulokset osoittavat, että useimmat ihmiset hyväksyvät tuotekuvat, jos ne ovat tarkkoja 3D-malleja tai oikeita valokuvia, joita on korkeintaan muokattu kevyesti. Jotkut vastaajista olivat kuitenkin sitä mieltä, että tuotekuvien tulisi olla mahdollisimman autenttisia eikä niitä saisi muokata millään tavalla. Synteettiset kuvat, jotka on generoitu valokuvien pohjalta tai ilman reaali maailman vastinetta, saivat vähemmän hyväksyntää etenkin tilanteissa, jossa niillä kuvastettaisiin jotain reaali maailman tuotetta. Toisaalta markkinointitilanteissa, joissa jotain uutta tuotetta kehitellään, suuri osa vastaajista hyväksyi myös täysin synteettisesti generoidut kuvat.

Kuluttajien hyväksyntään vaikuttaa myös se, miten avoimesti kuvangeneroimisen teknologioiden käyttö kerrotaan heille. Kyselyyn vastanneet henkilöt halusivat, että kuvien synteettinen luonne mainitaan, etenkin jos kuvat pyrkivät fotorealismiin. Tämä voi lisätä kuluttajien tietoisuutta ja arvostusta kuvangeneroimisen teknologioiden mahdollisuuksista ja rajoituksista. Lisäksi se voi vähentää mahdollisia pettymyksiä tai epäluottamusta, jos kuvat eivät vastaa todellisuutta tai odotuksia. On kuitenkin yhä myyntiorganisaation vastuulla varmistaa, ettei generoitu sisältö poikkea todellisesta. Kuluttajien hyväksyntään liittyy myös synteettisen generoimisen eettisyys. Kuvangeneroimisen teknologioiden käyttö markkinointiviestinnässä voi aiheuttaa eettisiä ongelmia, jos niillä

tuotetaan harhaanjohtavia, väärennettyjä tai haitallisia kuvia. Eettisesti olisi tärkeää ottaa huomioon kuluttajien oikeus saada tietää kuvien alkuperä ja mahdolliset manipulaatiot, sekä välttää harhaanjohtavaa tai petollista markkinointia. Kyselyn tulokset kuitenkin viittaavat tekniikoiden suureen potentiaaliin etenkin, jos niiden käyttöön ollaan avoimia ja se suoritetaan rehellisesti ja totuutta vääristämättä.

6 YHTEENVETO

Tämä kandidaatintyö tutki LDM:n soveltuvuutta tuotekuvien fotorealistiseen generoimiseen ja ihmisten näkemyksiä LDM:lla generoiduista tuotekuvista. Lisäksi tutkimuksessa pohdittiin synteettiseen kuvangenerointiin liittyviä eettisiä näkökulmia UNESCO:n ja ECCOLA-viitekehyksien perusteella. Tutkielmassa hyödynnettiin avoimesti ja yhteisövetoisesti kehitettyä AUTOMATIC1111:n verkkopohjaista käyttöliittymää ja sen lisäosia tutkimusmateriaalin muodostamiseksi. Työssä tutkimuskohde kuvattiin ja siitä luotiin kuvatietokanta, joka siirto-opetettiin Stable Diffusion v.2.1 non-EMA LDM:n arvojen yhteyteen hyödyntäen suurille kielimalleille kehitettyä LoRA (Low Ranking Adaptation of Large Language Models) menetelmää. Tekniikalla saavutettavan laadun todistamiseksi tehtiin kyselytutkimus, jonka tarkoituksena oli kerätä laadullista tietoa ihmisten mielipiteistä kuvangeneroimismenetelmien käytön hyväksyttävyydestä ja niiden tuottamasta laadusta.

Tutkimuksessa selvitettiin, miten LDM:tä voidaan käyttää tuotekuvien fotorealistiseen generoimiseen ja millaisia mielipiteitä ja asenteita ihmisillä on LDM:lla generoiduista tuotekuvista. Tutkimuksen tulokset osoittivat, että LDM:llä voidaan generoida fotorealistisia tuotekuvia, mutta ihmisten mielipiteet ja asenteet vaihtelivat riippuen tuotteen tyypistä ja kuvan laadusta. Kyselytutkimuksen perusteella voidaan kuitenkin todeta, että tekniikka on potentiaalinen menetelmä tulevaisuuden markkinointiviestinnässä. Tekniikalla voidaan myös saavuttaa ainakin suhteellinen fotorealismien taso, ja uudet kehitykset tutkimuskuvien generoimisen jälkeen ovat parantaneet tuloksia entisestään. Tutkimuksen rajoituksia olivat kyselytutkimuksen pieni otanta, kuvamateriaalin laatu ja koulutustietokannan koko sekä LDM:n kehittyminen tutkimuksen aikana.

Tutkimuskyselyssä kerätyt tulokset viittaavat tekniikoiden toimivuuteen, mutta ne myös kertoivat tutkimushetkellä ajankohtaisista kuvasynteesin ongelmista. Kyselyn tulosten perusteella voidaan vetää johtopäätöksiä siitä, että kuvangeneroiminen voi olla hyväksyttävää useissa markkinointiviestinnällisissä tarkoituksissa, kuten vaikkapa uuden tuotteen kehittelyvaiheessa ja mielipiteiden keräämisessä. Tämä voisi konkretisoida myös järjestelmänä, jossa asiakas pääsee itse määrittämään generoimisen parametrit. Ainakin teoriassa tällaisella

järjestelmällä voitaisiin reaaliajassa valita halutut tuoteomaisuudet, kuten vaikkapa tuotteen verhoilu ja sävy, ja siinä generoiminen voitaisiin suorittaa oikeiden mallistossa olevien tuotekuvien perusteella. Kyselytutkimuksen perusteella avoimuus LDM:en käyttämisen suhteen vaikuttaisi olevan avaintekijänä sen hyväksyttävyydessä etenkin tuotekuvien pyrkiessä fotorealismiin. On keskeistä, ettei kuluttajia johdeta harhaan epärealistisilla malleilla, jotka vaikkapa muuttavat tuotteen todellisia ominaisuuksia, kuten sen verhoilua tai sävytystä todellisuudesta poikkeavaksi.

Tutkimus osoitti, että LDM pystyy luomaan korkealaatuisia ja uskottavia kuvia ja että ihmisten mielipiteet ja asenteet LDM:lla generoituihin tuotekuviin ovat pääosin positiivisia tai neutraaleja. Tutkimus tarjoaa uusia näkökulmia ja mahdollisuuksia markkinointiviestinnän sisällöntuotantoon hyödyntämällä syväoppimisen mallien potentiaalia. Jatkotutkimus voisi tarkastella uusilla tekniikoilla saavutettavaa laatua, sitä kuinka generoimisprosessia voidaan nopeuttaa, suuremmilla kuvatietokannoilla saavutettavia tuloksia ja esimerkiksi resoluutioltaan yksityiskohtaisemman sisällön tuottamista. Lisäksi tutkimus videon fotorealistisesta generoimisesta avaisi uusia ovia markkinointiviestinnälliseen sisällöntuottamiseen. Erityisen tärkeää olisi myös kehittää järjestelmiä synteettisen sisällön tunnistamiseksi, jotta tekniikalla tehtäviä väärinkäytöksiä voitaisiin ehkäistä. Vaikka tutkielmassa synteettisen sisällön tunnistamista tarkasteltiin vain pintapuolisesti, voidaan generoimisartefaktien piilottamisen todeta olevan ainakin osittain mahdollista superresoluution, satunnaisen kohinan ja interpoloinnin avulla.

LÄHTEET

- Anciukevičius, T., Xu, Z., Fisher, M., Henderson, P., Bilen, H., Mitra, N. J. & Guerrero, P. (2022). RenderDiffusion: Image Diffusion for 3D Reconstruction, Inpainting and Generation. *arXiv preprint arXiv:2211.09869*.
- Bank, D., Koenigstein, N. & Giryes, R. (2020). Autoencoders. *arXiv preprint arXiv:2003.05991*.
- Bar-Tal, O., Yariv, L., Lipman, Y. & Dekel, T. (2023). MultiDiffusion: Fusing Diffusion Paths for Controlled Image Generation. *arXiv preprint arXiv:2302.08113*, 2.
- Bashir, S. M. A., Wang, Y., Khan, M. & Niu, Y. (2021). A comprehensive review of deep learning-based single image super-resolution. *PeerJ Computer Science*, 7, e621.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Haettu 20.3.2023 osoitteesta <https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>
- Chen, D. Z., Siddiqui, Y., Lee, H. Y., Tulyakov, S. & Nießner, M. (2023). Text2Tex: Text-driven Texture Synthesis via Diffusion Models. *arXiv preprint arXiv:2303.11396*.
- Corvi, R., Cozzolino, D., Zingarini, G., Poggi, G., Nagano, K. & Verdoliva, L. (2022). On the detection of synthetic images generated by diffusion models. *arXiv preprint arXiv:2211.00680*.
- Cuenca, P. & Paul, S. (26.1.2023), *Using LoRA for Efficient Stable Diffusion Fine-Tuning*. Haettu 5.4.2023 osoitteesta <https://huggingface.co/blog/lora>
- Dhariwal, P. & Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34, 8780-8794.
- Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L. & Kohane, I. S. (2019). *Adversarial attacks on medical machine learning*. *Science*, 363 (6433), 1287-1289. <https://doi.org/10.1126/science.aaw4399>
- Forsgren, S. & Martiros, H. (2022), *Riffusion – Stable diffusion for real-time music generation*. Haettu 4.4.2023 osoitteesta <https://www.riffusion.com/about>
- Gal, R., Alaluf, Y., Atzmon, Y., Patashnik, O., Bermano, A. H., Chechik, G. & Cohen-Or, D. (2022). An image is worth one word: Personalizing Text-to-Image generation using textual inversion. *arXiv preprint arXiv:2208.01618*.
- Ge, S., Park, T., Zhu, J. Y. & Huang, J. B. (2023). Expressive Text-to-Image Generation with Rich Text. *arXiv preprint arXiv:2304.06720*.
- Guttenberg, N. (2023). *Diffusion With Offset Noise*. Cross Labs. Haettu 5.4.2023 osoitteesta <https://www.crosslabs.org/blog/diffusion-with-offset-noise>

- Hermann, E. (2022). Leveraging artificial intelligence in marketing for social good-An ethical perspective. *Journal of Business Ethics*, 179(1), 43-61. <https://doi.org/10.1007/s10551-021-04843-y>.
- Ho, J., Jain, A. & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 6840-6851.
- Holmgren, H. & Andersson, V. (2012). Fotorealism i montage.
- Honkala, K. (2009). *Ulkoisen markkinointiviestinnän projekti: Tapaustutkimus: Kaatumisen ehkäisyprojekti*. Turun yliopisto. <https://urn.fi/URN:NBN:fi-fe2015091813302>
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2021). Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.
- Karnewar, A., Vedaldi, A., Novotny, D. & Mitra, N. (2023). HOLODIFFUSION: Training a 3D Diffusion Model using 2D Images. *arXiv preprint arXiv:2303.16509*.
- Karras, T., Aittala, M., Aila, T. & Laine, S. (2022). Elucidating the design space of diffusion-based generative models. *arXiv preprint arXiv:2206.00364*.
- Kemelmacher-Shlizerman, I. (2023). *How AI makes virtual try-on more realistic*. The Keyword. Haettu 20.6.2023 osoitteesta <https://blog.google/products/shopping/virtual-try-on-google-generative-ai/>.
- Khachatryan, L., Movsisyan, A., Tadevosyan, V., Henschel, R., Wang, Z., Navasardyan, S. & Shi, H. (2023). Text2Video-Zero: Text-to-Image Diffusion Models are Zero-Shot Video Generators. *arXiv preprint arXiv:2303.13439*.
- Li, B., Cheng, S. & Lin, J. (2021). tcFFT: Accelerating half-precision FFT through tensor cores. *arXiv preprint arXiv:2104.11471*.
- Li, J., Li, D., Xiong, C. & Hoi, S. (2022, June). Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International Conference on Machine Learning* (pp. 12888-12900). PMLR.
- Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L. & Timofte, R. (2021). Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1833-1844).
- Lin, S., Liu, B., Li, J. & Yang, X. (2023). Common Diffusion Noise Schedules and Sample Steps are Flawed. *arXiv preprint arXiv:2305.08891*.
- Luo, Z., Chen, D., Zhang, Y., Huang, Y., Wang, L., Shen, Y., ... & Tan, T. (2023). Decomposed Diffusion Models for High-Quality Video Generation. *arXiv preprint arXiv:2303.08320*.

- Malinovsky, G., Sailanbayev, A. & Richtárik, P. (2021). Random reshuffling with variance reduction: New analysis and better rates. *arXiv preprint arXiv:2104.09342*.
- Middleton, F. (2019). *Reliability vs. validity in research | Difference, types and examples*. Scribbr. Haettu 9.6.2023 osoitteesta <https://www.scribbr.com/methodology/reliability-vs-validity/>.
- Mikubill. (2023). New Model: The finished tile model is released. *GitHub discussion*. GitHub. Haettu 14.5.2023 osoitteesta <https://github.com/Mikubill/sd-webui-controlnet/discussions/1142>.
- Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I. and Chen, M. (2021). Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*.
- Richardson, E., Metzger, G., Alaluf, Y., Giryes, R. & Cohen-Or, D. (2023). Texture: Text-guided texturing of 3D shapes. *arXiv preprint arXiv:2302.01721*.
- Rincon, L. (2023). *Virtually try on clothes with a new AI shopping feature*. The Keyword. Haettu 20.6.2023 osoitteesta <https://blog.google/products/shopping/ai-virtual-try-on-google-shopping/>.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10684-10695).
- Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M. & Aberman, K. (2022). DreamBooth: Fine tuning Text-to-Image diffusion models for subject-driven generation. *arXiv preprint arXiv:2208.12242*.
- Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., ... & Norouzi, M. (2022, July). Palette: Image-to-Image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings* (pp. 1-10).
- Sipolainen, J. (2021). *Harhaanjohtavuudesta erehdyttämiseen – tutkimus markkinointirikoksen ja petoksen välisestä rajanvedosta*. Lapin yliopisto. <https://lauda.ulapland.fi/handle/10024/64895>.
- Song, J., Meng, C. & Ermon, S. (2020). Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.
- UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. Haettu 16.5.2023 osoitteesta <https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>.
- Vakkuri, V., Kemell, K.-K., Jantunen, M., Halme, E. & Abrahamsson, P. (2021). ECCOLA : a method for implementing ethically aligned AI systems.

Journal of Systems and Software, 182, Article 111067.
<https://doi.org/10.1016/j.jss.2021.111067>.

- Wang, X. & Hu, Q. (2023). Visual Truth and Image Manipulation: Visual Ethical Anomie and Reconstruction of Digital Photography. In *SHS Web of Conferences* (Vol. 155, p. 03018). EDP Sciences.
- Wang, X., Xie, L., Dong, C. & Shan, Y. (2021). Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1905-1914).
- Wicklin, R. (2020). What is bilinear interpolation? *The DO Loop*. Haettu 26.4.2023 osoitteesta <https://blogs.sas.com/content/iml/2020/05/18/what-is-bilinear-interpolation.html>.
- Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Shao, Y., Zhang, W., Cui, B. and Yang, M.H. (2022). Diffusion models: A comprehensive survey of methods and applications. *arXiv preprint arXiv:2209.00796*.
- Yazıcı, Y., Foo, C. S., Winkler, S., Yap, K. H., Piliouras, G. & Chandrasekhar, V. (2018). The unusual effectiveness of averaging in gan training. *arXiv preprint arXiv:1806.04498*.
- YK:n ihmisoikeuksien yleismaailmallinen julistus. (1948). Haettu 9.6.2023 osoitteesta <https://ihmisoikeusliitto.fi/ihmisoikeudet/ihmisoikeuksien-julistus/>.
- Zhang, L. & Agrawala, M. (2023). Adding conditional control to Text-to-Image diffusion models. *arXiv preprint arXiv:2302.05543*.
- Zhao, W., Bai, L., Rao, Y., Zhou, J. & Lu, J. (2023). UniPC: A Unified Predictor-Corrector Framework for Fast Sampling of Diffusion Models. *arXiv preprint arXiv:2302.04867*.

LIITE 1 NÄYTTEENOTTOASKELEET VAIHEITTAIN



KUVIO 19 Näytteenottoaskelia: 0, 1 & 5. Kehote: *“Green Pinja spoke chair, during sunset, golden hour, snow on the ground, winter forest in the background, clouds in the sky”*. 1024px x 1024px / kuva. Ensimmäisessä kuvassa on kohinaa, mistä kuvan generoimisprosessi alkaa. Näytteenottoaskelten kasvaessa kuvan yksityiskoh-
taisuus ja täsmällisyys kehotteen kanssa paranevat.



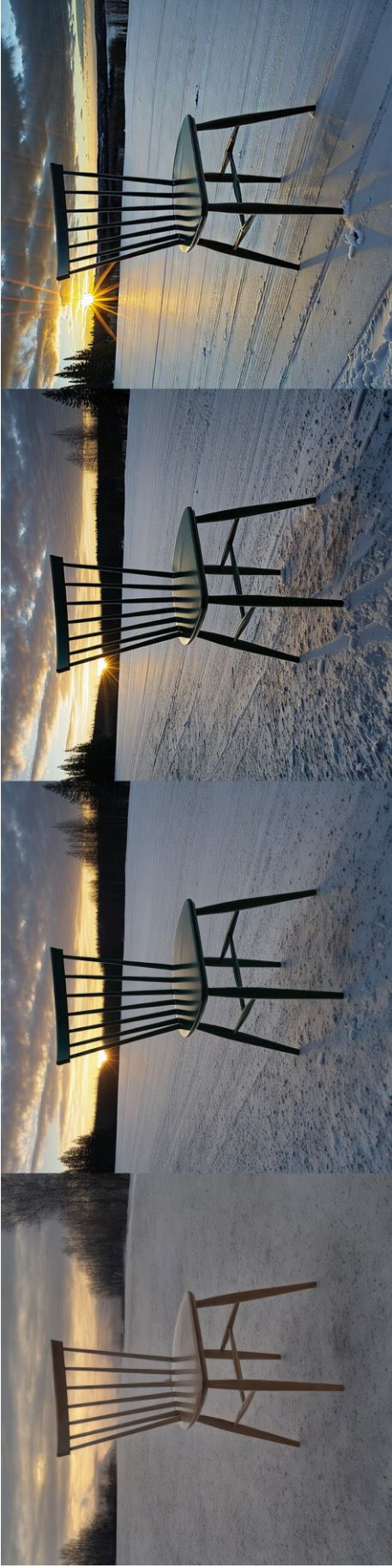
KUVIO 20 Näytteenottoaskelia: 10, 20 & 30. Kehote: *“Green Pinja spoke chair, during sunset, golden hour, snow on the ground, winter forest in the background, clouds in the sky”*. 1024px x 1024px / kuva.



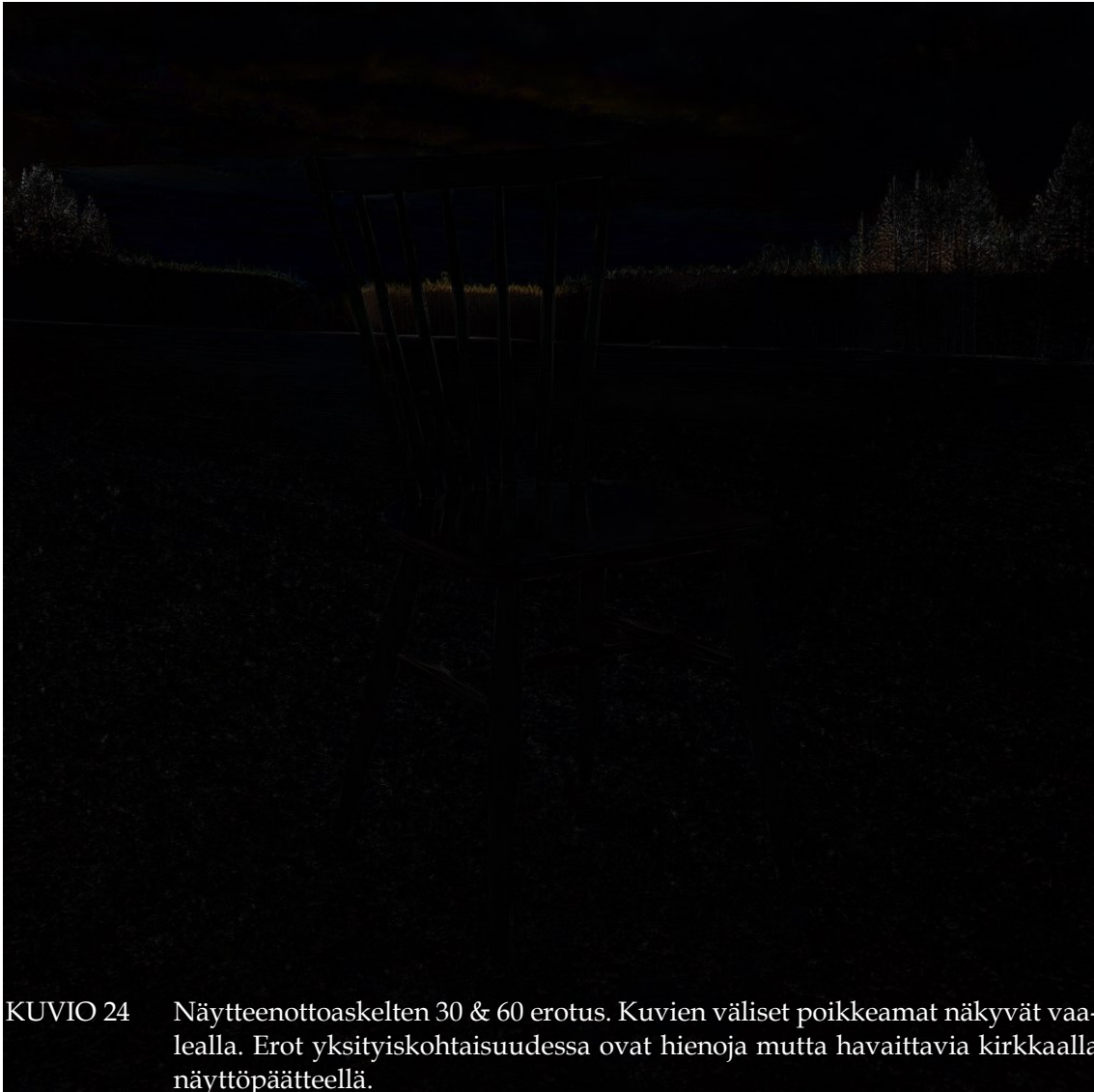
KUVIO 21 Näytteenottoaskelia: 40, 50 & 60. Kehote: *“Green Pinja spoke chair, during sunset, golden hour, snow on the ground, winter forest in the background, clouds in the sky”*. 1024px x 1024px / kuva.

LIITE 2 GENEROITU KUVA & CFG SCALEN VAIKUTUS

KUVIO 22 Generoitu tuotokuva 60 näytteenottoaskeleen jälkeen. Kehote: *“Green Pinja spoke chair, during sunset, golden hour, snow on the ground, winter forest in the background, clouds in the sky”*. 1024px x 1024px. CFG scale: 4.



KUVIO 23 Generoitu tuotekuva 60. näytteenottoaskeleen jälkeen. CFG scale: 1, 4, 7 & 14.
Kehote: "Green Pinja spoke chair, during sunset, golden hour, snow on the ground, winter forest in the background, clouds in the sky". 1024px x 1024px / kuva. CFG scalen kasvaessa kuvan värikylläisyys nousee, ja se on tyytellymmän näköinen. Se voi toimia taiteellisissa kuvissa, mutta fotorealismissa värityksen on syytä olla hillittyä.

LIITE 3 NÄYTTEENOTTOASKELTEN VERTAILU

KUVIO 24 Näytteenottoaskelten 30 & 60 erotus. Kuvien väliset poikkeamat näkyvät vaalealla. Erot yksityiskohtaisuudessa ovat hienoja mutta havaittavia kirkkaalla näyttöpäätteellä.



KUVIO 25 Generoidun kuvan yksityiskohtaisuus 30 & 60 näytteenottoaskeleen jälkeen.

LIITE 4 GENEROITU KUVA SKAALATTUNA ILMAN SUPER-RESOLUUTIOTA

KUVIO 26 Skaalattu kuva. Kehote: "*Brown, peanut, wooden, dark oak colored Pinja spoke chair, white background, concrete floor*". 256px x 256px.

LIITE 5 PIENEN KUVAN SUPERRESOLUUTIO

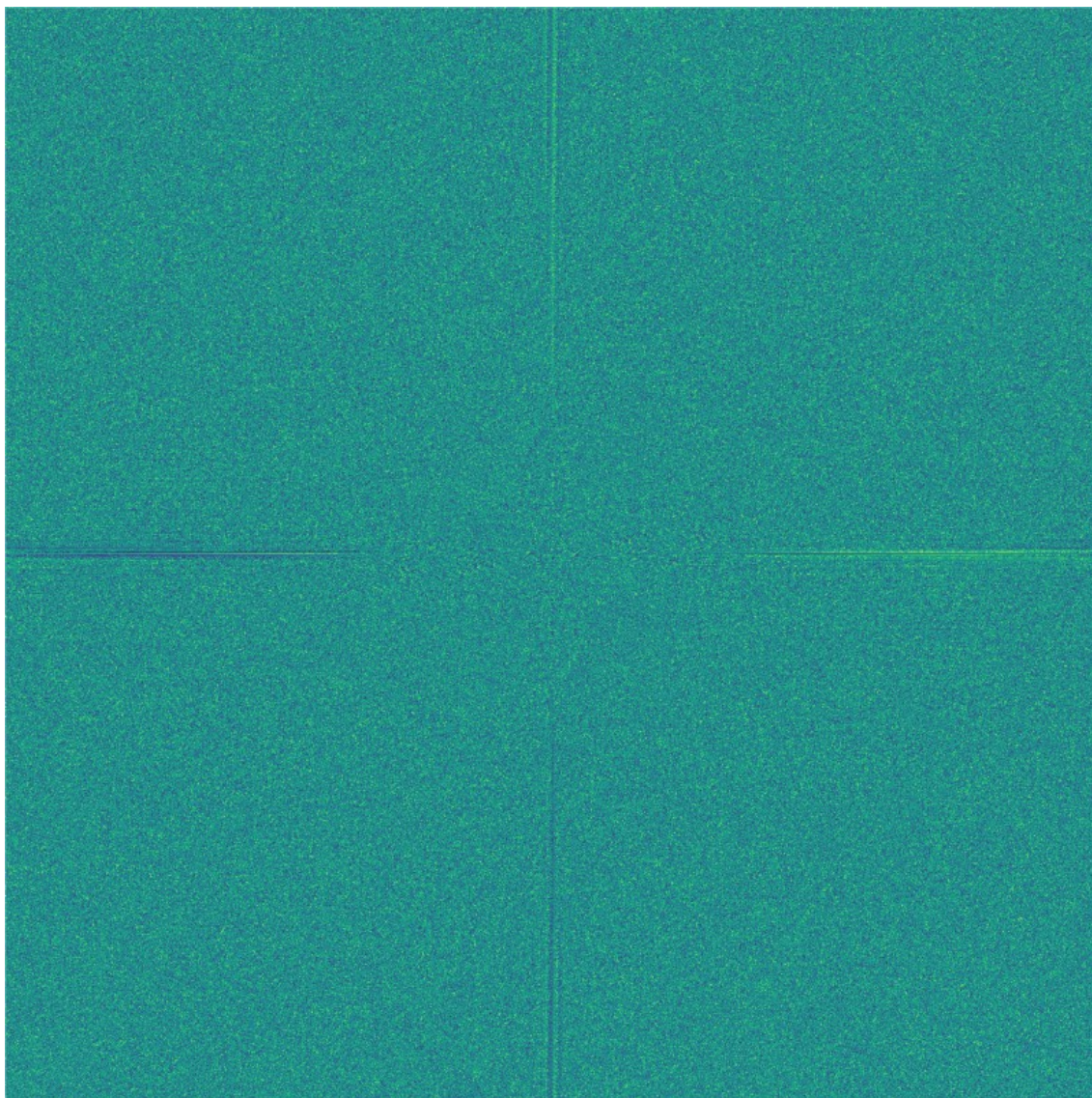
KUVIO 27 Superresoluutio (1/2). Kehote: "*Brown, peanut, wooden, dark oak colored Pinja spoke chair, white background, concrete floor*". Upscaler 1: SwinIR 4x. Upscaler 2: ESRGAN 4x (50 % intensiteetillä). 768px x 1024px (skaalattu ja rajattu kuva, alkuperäinen 256px x 256px).

LIITE 6 SUUREN KUVAN SUPERRESOLUUTIO



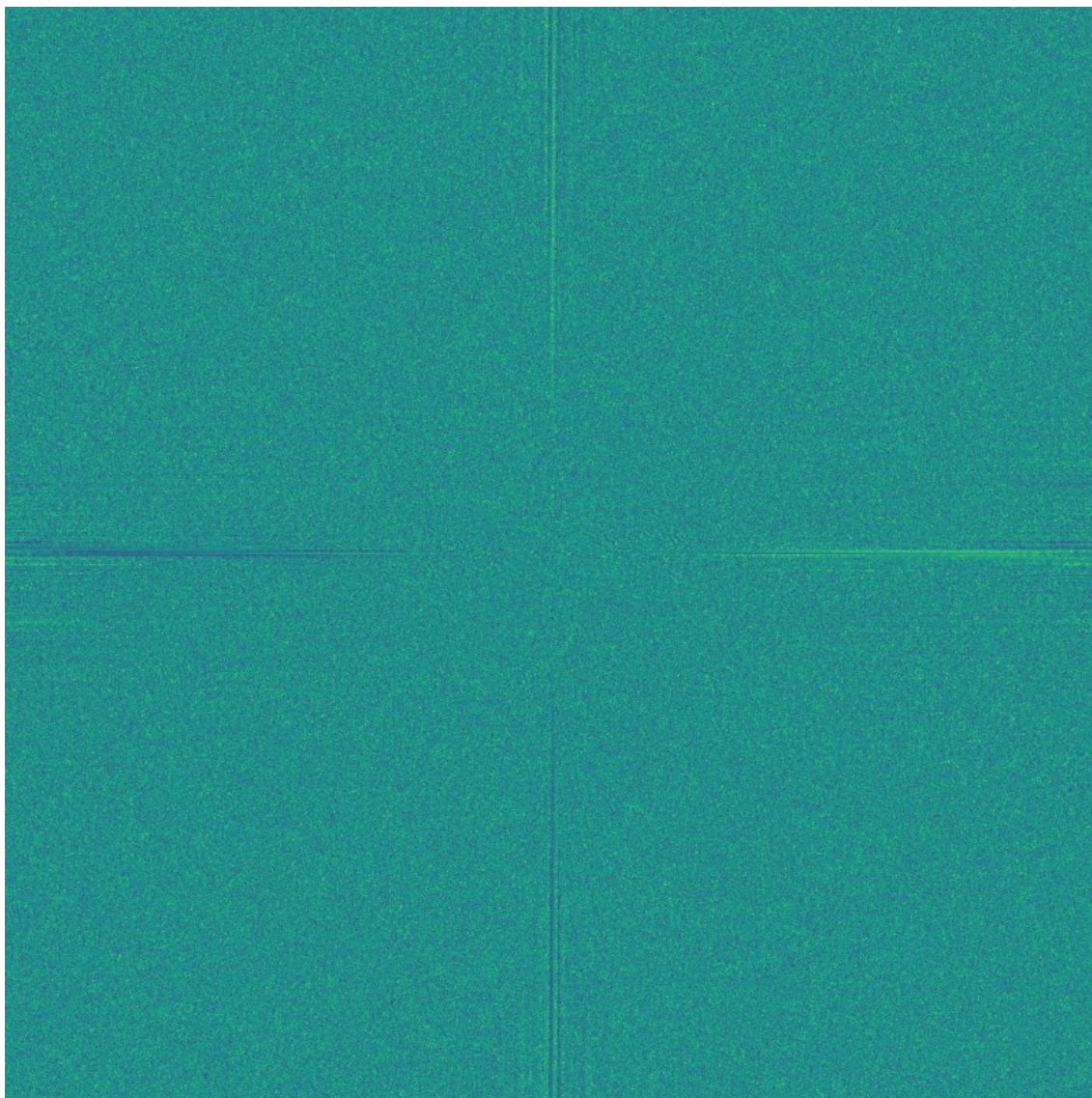
KUVIO 28 Superresoluutio (2/2). Kehote: "*Brown, peanut, wooden, dark oak colored Pinja spoke chair, white background, concrete floor*". Upscaler: ESRCAN 4x. 3072px x 4096px (skaalattu ja rajattu kuva, alkuperäinen 1024px x 1024px).

LIITE 7 ARTEFAKTIT GENEROIDUN KUVAN DFT-VAIHE- KUVASSA



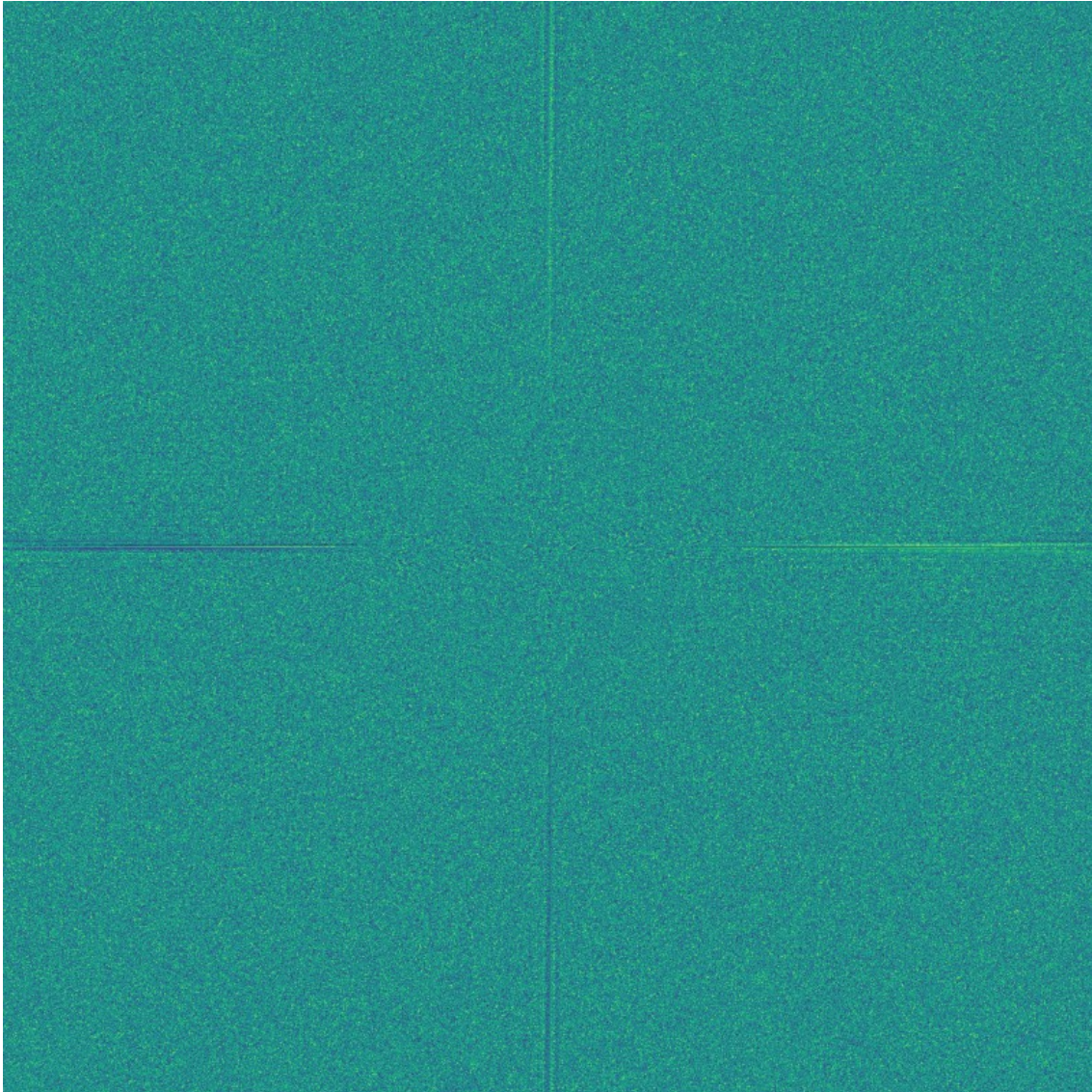
KUVIO 29 Fourier-muunnos (DFT:n vaihekuva) ESRGAN 2x superresoluutiosta ja siitä syntyvistä artefakteista generoidussa kuvassa. Artefaktit symmetrisiä viivoja vaaka- ja pystyakseleilla.

LIITE 8 ARTEFAKTIT OIKEAN VALOKUVAN DFT-VAIHEKUVASSA

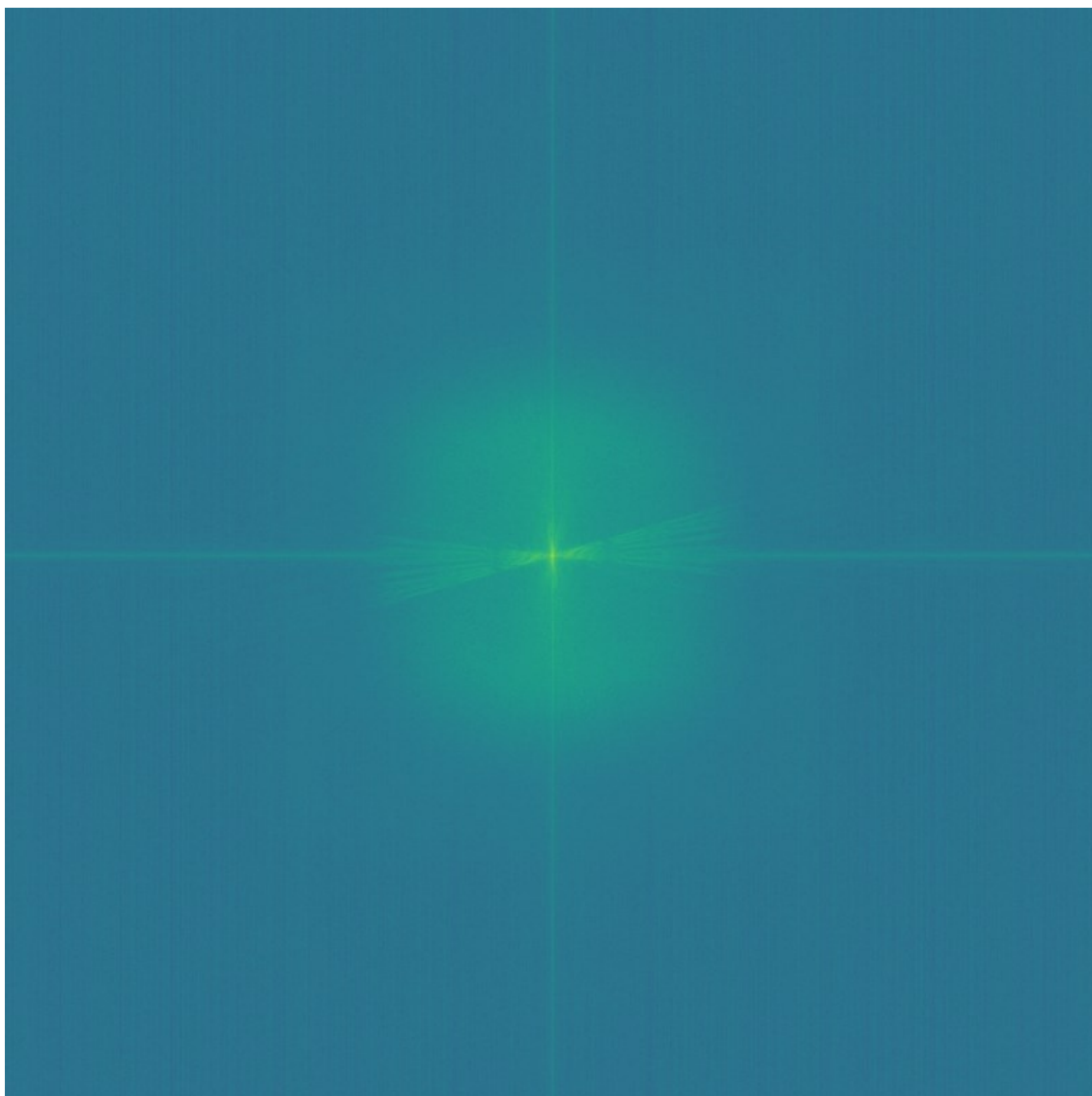


KUVIO 30 Fourier-muunnos (DFT:n vaihekuva) ESRGAN 4x superresoluutiosta ja siitä syntyviä artefakteja oikeassa valokuvassa.

LIITE 9 ARTEFAKTIT INTERPOLOIDUSSA DFT-VAIHE- KUVASSA

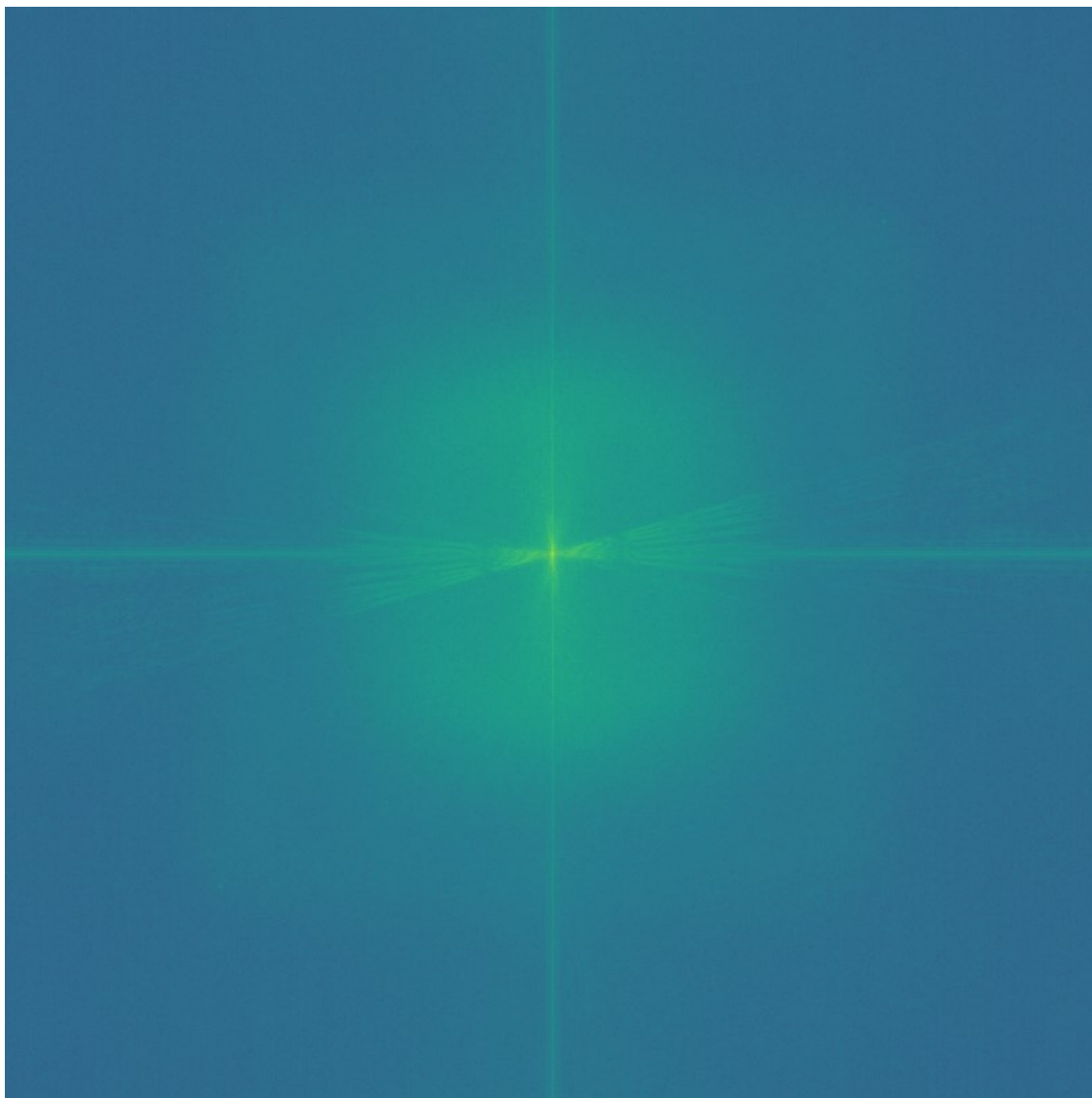


KUVIO 31 Fourier-muunnos ESRGAN 6x superresoluutio generoidussa kuvassa. Lisätty digitaalista kohinaa, ja rajattu 2048px x 2048px resoluutioon käyttäen Bilinear-interpolointia. Kuvassa yhä havaittavissa viivat vaaka- ja pystyakselilla, mutta ne eivät ole yhtä symmetrisiä, ja koneellisesti tuotetun näköisiä. Toisaalta viivat vaaka- ja pystyakselilla viestivät yhä kuvan koneellisesta luonteesta.

LIITE 10 DIGITAALINEN KOHINA JA DFT-SUURUUSKUVA

KUVIO 32 Digitaalinen kohina havainnoituna DFT:n suuruuskuvassa. Kohina ilmenee symmetrisinä viivoina, jotka kulkevat kuvassa pystysuunnassa.

LIITE 11 DIGITAALISEN KOHINAN PILOTTAMINEN JA DFT-SUURUUSKUVA



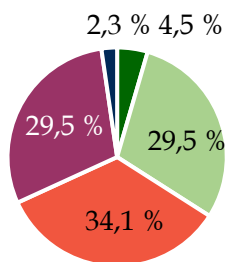
KUVIO 33 Digitaalisen kohinan piilottaminen havainnoituna DFT:n suuruuskuvana. Kuvaan tehty 6x ESRGAN superresoluutio, johon lisätty digitaalista kohinaa, ja joka on skaalattu 2048px x 2048px resoluutioon hyödyntäen Bilinear-interpolointia.

LIITE 12 KYSELYTUTKIMUKSEN TULOKSET

Mielipidekysely tekniikoiden käytöstä markkinointiviestinnässä

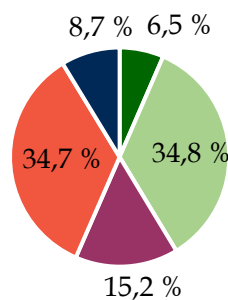
KYSYMYS 1.

Verkkosivulla myydään jotain oikeasti olemassa olevaa tuotetta. Mikä on mielestäsi hyväksyttävä tapa tämän tuotteen kuvastamiseen?



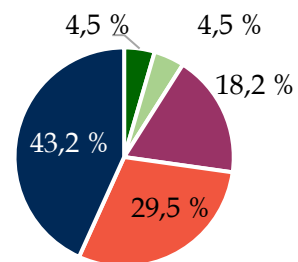
KYSYMYS 2.

Yritys toteuttaa markkinointikampanjan jonkin olemassa olevan tuotteen kohdalla. Tuotetta markkinoidaan sosiaalisessa mediassa. Mikä on mielestäsi hyväksyttävä tapa tuotteen kuvastamiseen?



KYSYMYS 3.

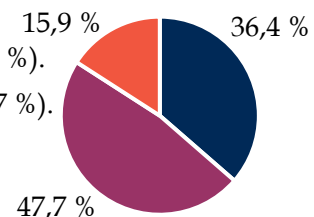
Yritys kerää tietoa ihmisten mielipiteistä ja kiinnostuksesta jotain uutta tuotekonseptia kohtaan. Mikä on mielestäsi hyväksyttävä tapa uuden suunnitella olevan tuotteen kuvastamiseen?



- Tuotekuvien tulisi olla ainoastaan oikeita valokuvia.
- Tuotekuvien tulisi olla oikeita valokuvia, mutta niiden pienimuotoinen käsittely ei haittaa.
- Tuotekuvat voivat perustua oikeisiin valokuviiin tai tarkkoihin 3D-mallinnoksiin.
- Tuotekuvat voivat olla synteettisesti generoituja, jos niiden koulutustietokanta perustuu oikeisiin valokuviiin.
- Tuotekuvat voivat olla synteettisesti generoituja, vaikka niillä ei olisi tosimaailman vastinetta.

KYSYMYS 4. Tulisiko markkinointimateriaaleissa olla mainittuna, mikäli sen sisältö on synteettisesti generoitua?

- Kyllä, myös vahvasti tyyliteltyissä kuvissa (16 ääntä, 36,4 %).
- Kyllä, mutta vain tähdätessä fotorealismiin (21 ääntä, 47,7 %).
- Ei ole tarvetta mainita (7 ääntä, 15,9 %).



KYSYMYS 5. Jos vastasit aiempaan kysymykseen kyllä, millainen maininnan olisi syytä olla?

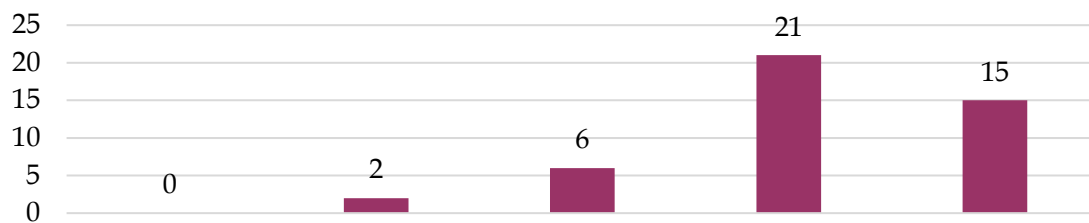


28 vastaajaa.

KYSYMYS 6. Miten fotorealistisesti generoidut kuvat (oikealla) mielestäsi muistuttavat sisällöltään alkuperäisten kuvien (vasemmalla) tuotteita? Taustojen vastaavuus ei ole merkityksellistä.



Tuotettu *Text-to-Image*-tekniikalla koulutustietokannassa olevan värin perusteella.
768px x 768px.



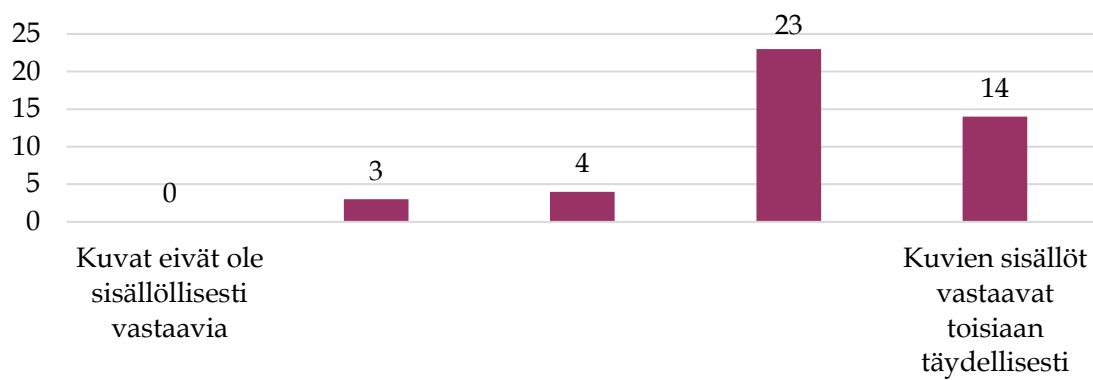
Kuvat eivät ole sisällöllisesti vastaavia.

Kuvien sisällöt vastaavat toisiaan täydellisesti.

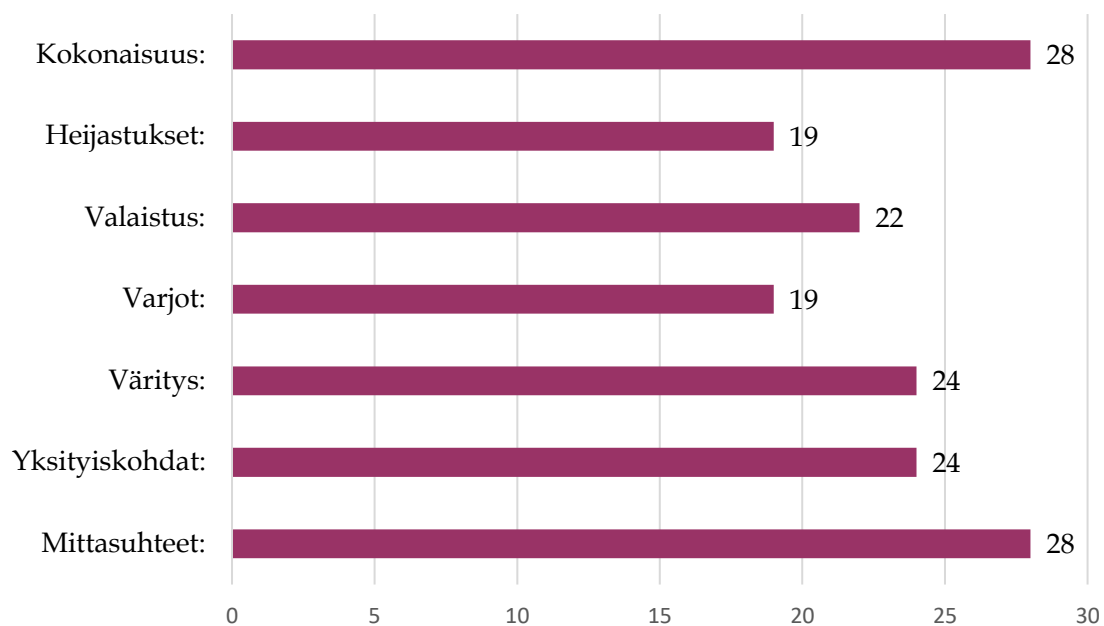
KYSYMYS 7. Miten fotorealistisesti generoidut kuvat (oikealla) mielestäsi muistuttavat sisällöltään alkuperäisten kuvien (vasemmalla) tuotteita? Taustojen vastaavuus ei ole merkityksellistä.



Tuotettu *Text-to-Image*-tekniikalla koulutustietokannassa olevan värin perusteella.
768px x 768px.

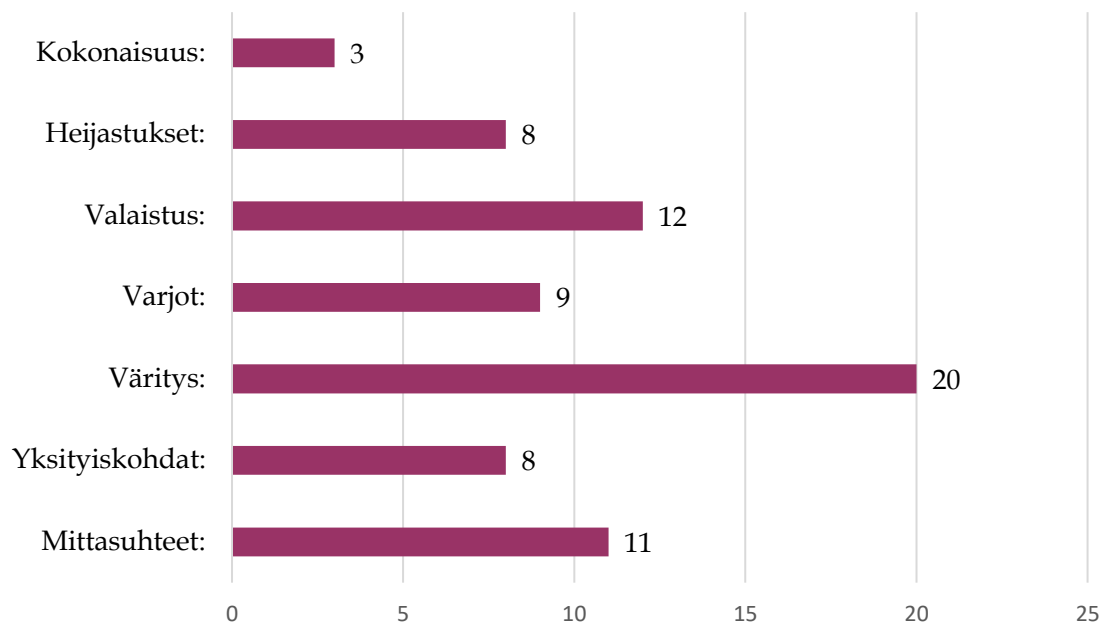


Vapaavalintainen kysymys: Mikä kysymysten 6 & 7 generoiduissa kuvissa onnistuu?



41 vastaajaa.

Vapaavalintainen kysymys: Mikä kysymysten 6 & 7 generoiduissa kuvissa epäonnistuu?



34 vastaajaa.

KYSYMYS 8 & 9. Miten fotorealistisesti generoitu kuva (alhaalla) mielestäsi muistuttaa alkuperäistä tuotetta (ylhällä)?



Kuvat kysymyksestä 8.

Kuvat kysymyksestä 9.

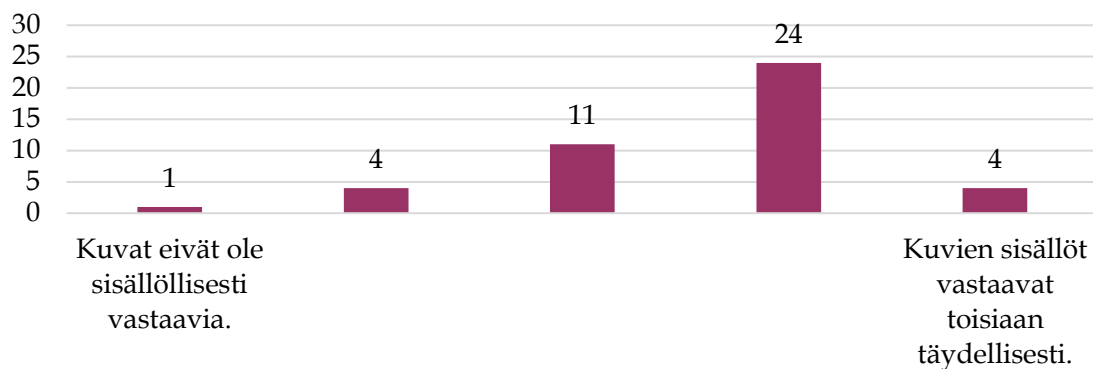
Generoitu kuva (alempi kuva) tuotettu *Image-to-Image*-tekniikalla olemassa olevaan väriin (ylempi kuva¹²) pohjautuen. 1088px x 1088px.

Generoitu kuva (alempi kuva) tuotettu *Image-to-Image*-tekniikalla olemassa olevaan väriin (ylempi kuva¹³) pohjautuen. 960px x 960px.

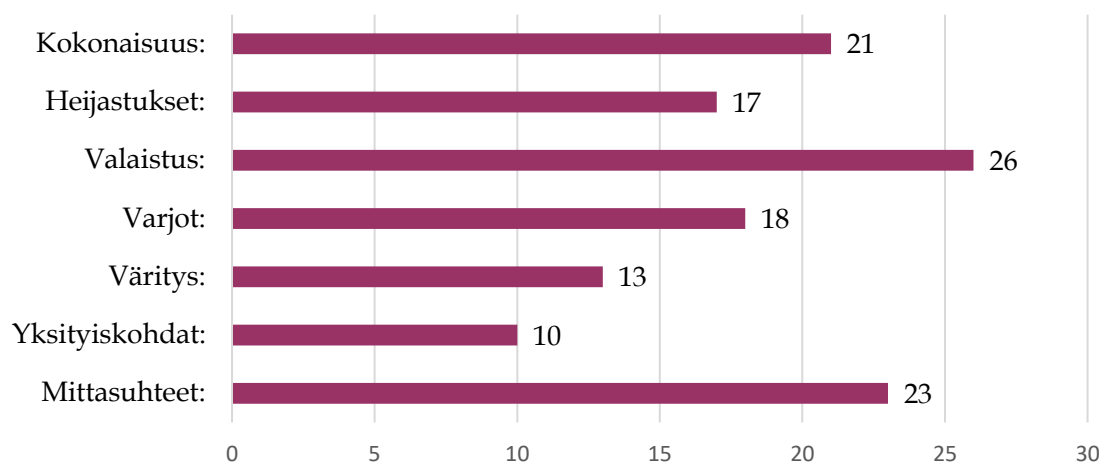
¹² Pohjanmaan Kaluste, (2023). Haettu 10.1.2023 osoitteesta <https://www.pohjanmaan.fi/tuotteet/pinja/ruokatuoli> myyntiliikkeen luvalla.

¹³ Pohjanmaan Kaluste, (2023). Haettu 10.1.2023 osoitteesta <https://www.pohjanmaan.fi/tuotteet/pinja/ruokatuoli> myyntiliikkeen luvalla.

KYSYMYS 8 TULOKSET.

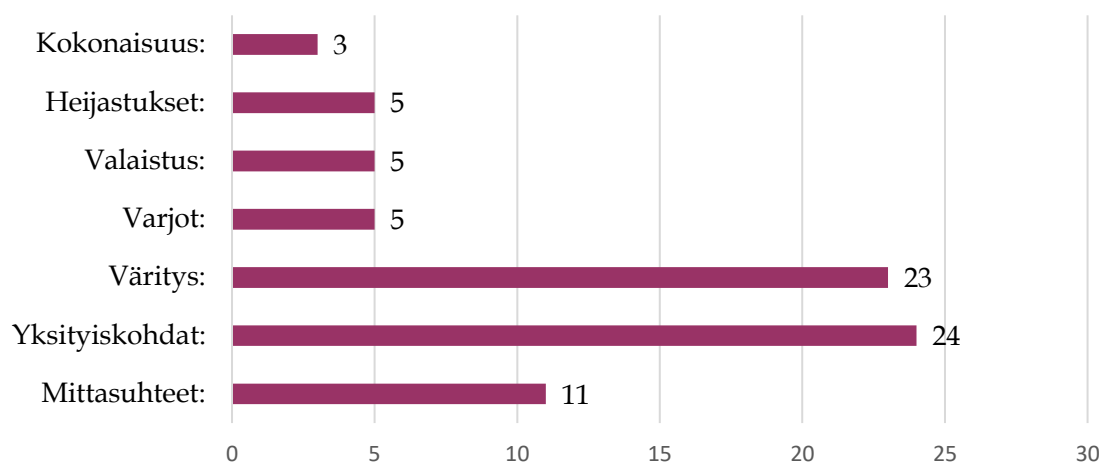


Vapaavalintainen kysymys: Mikä kysymyksen 8 kuvassa onnistuu?



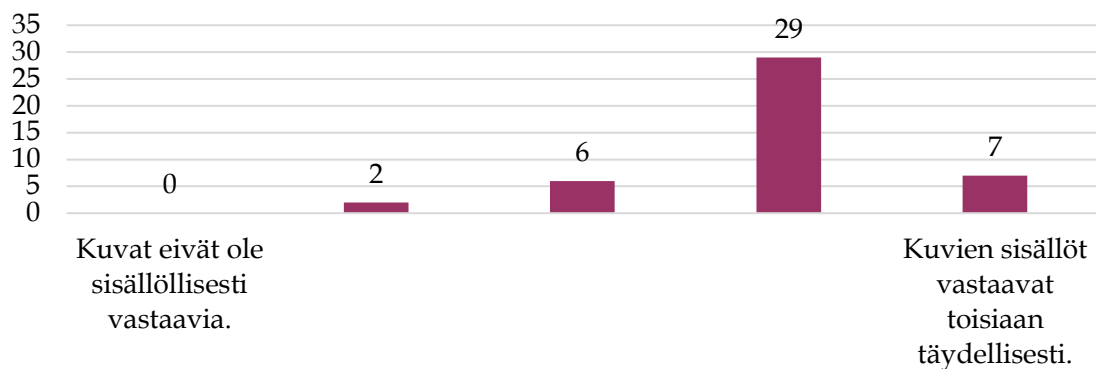
40 vastaajaa.

Vapaavalintainen kysymys: Mikä kysymyksen 8 kuvassa epäonnistuu?

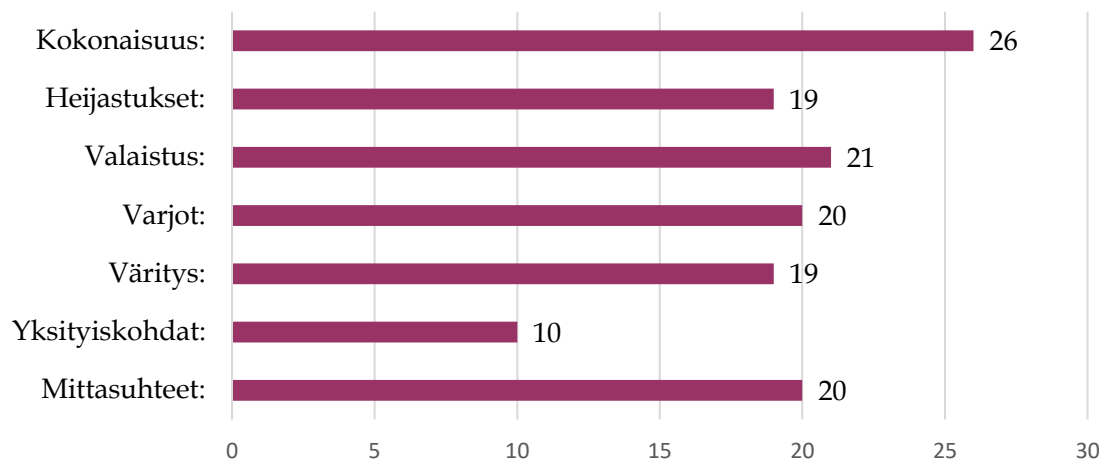


36 vastaajaa.

KYSYMYS 9 TULOKSET.

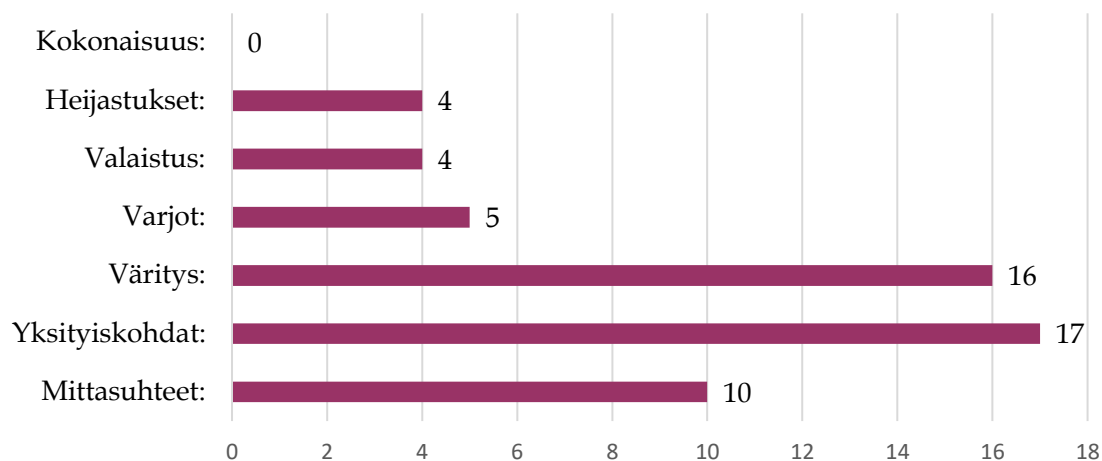


Vapaavalintainen kysymys: Mikä kysymyksen 9 kuvassa onnistuu?



42 vastaajaa.

Vapaavalintainen kysymys: Mikä kysymyksen 9 kuvassa epäonnistuu?



32 vastaajaa.

KYSYMYKSET 10 & 11. Miten fotorealistisesti uskottavia generoidut kuvat ovat sävytykseltään ja muodoiltaan?



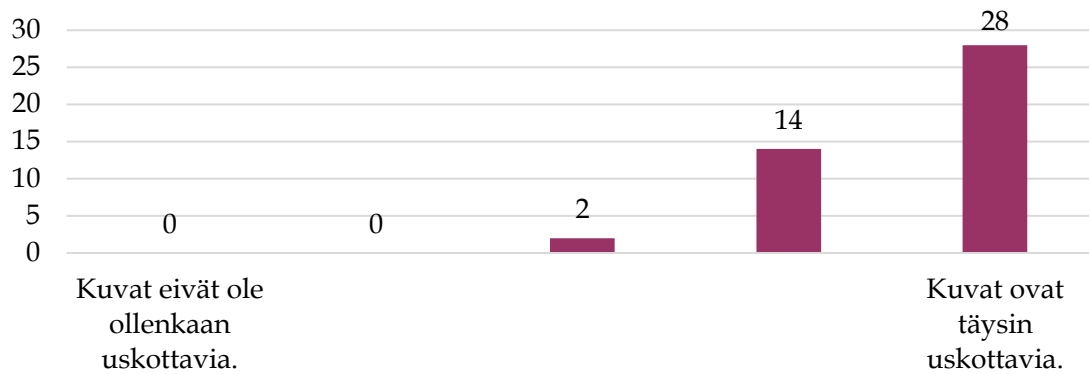
Kuvat kysymyksestä 10.

Kuvat kysymyksestä 11.

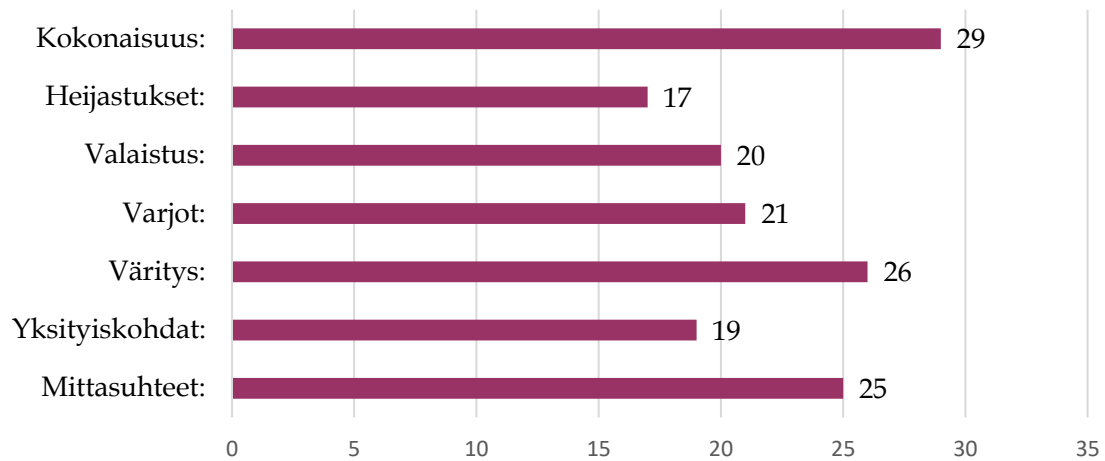
Kuvat generoitu *Image-to-Image*-tekniikalla uudella koulutustietokannasta puuttuvaan väriin pohjautuen. 960px x 960px.

Kuvat generoitu *Image-to-Image*-tekniikalla uudella koulutustietokannasta puuttuvaan väriin pohjautuen. 960px x 960px.

KYSYMYS 10 TULOKSET.

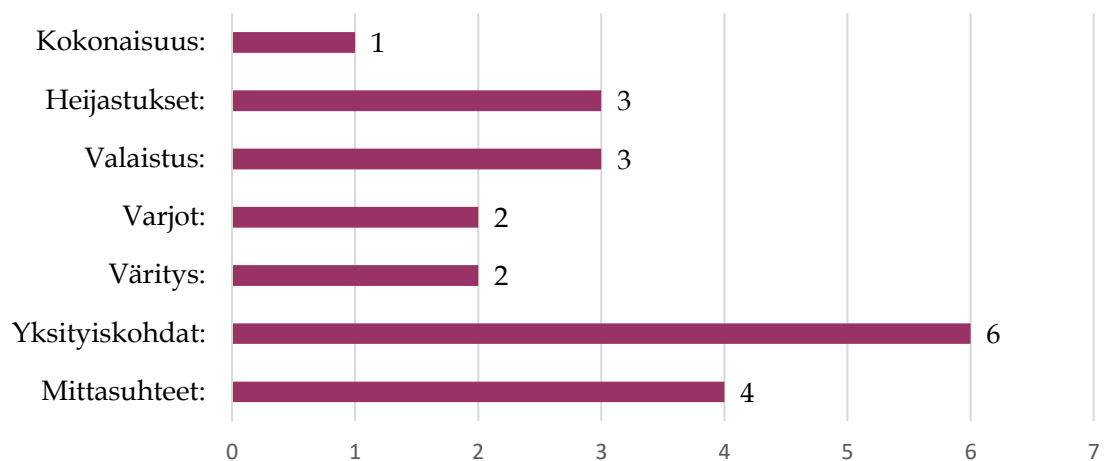


Vapaavalintainen kysymys: Mikä kysymyksen 10 kuvissa onnistuu?



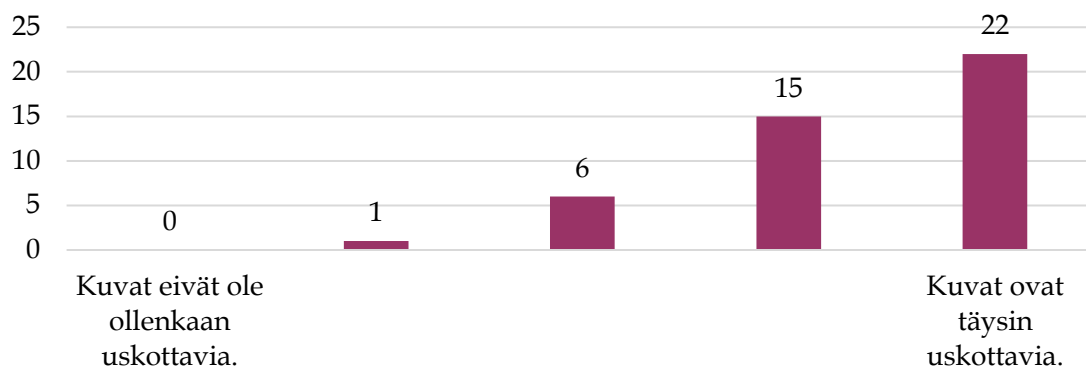
39 vastaajaa.

Vapaavalintainen kysymys: Mikä kysymyksen 10 kuvissa epäonnistuu?

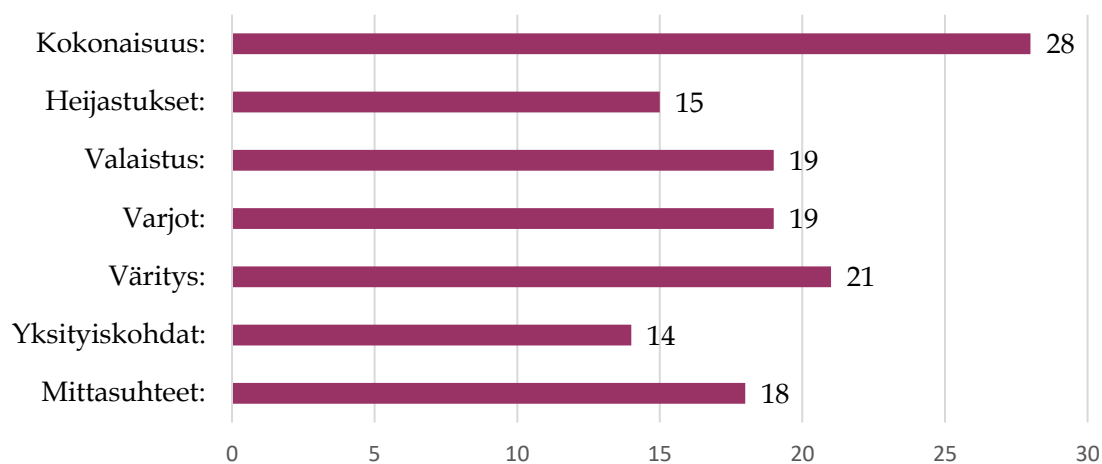


13 vastaajaa.

KYSYMYS 11 TULOKSET.

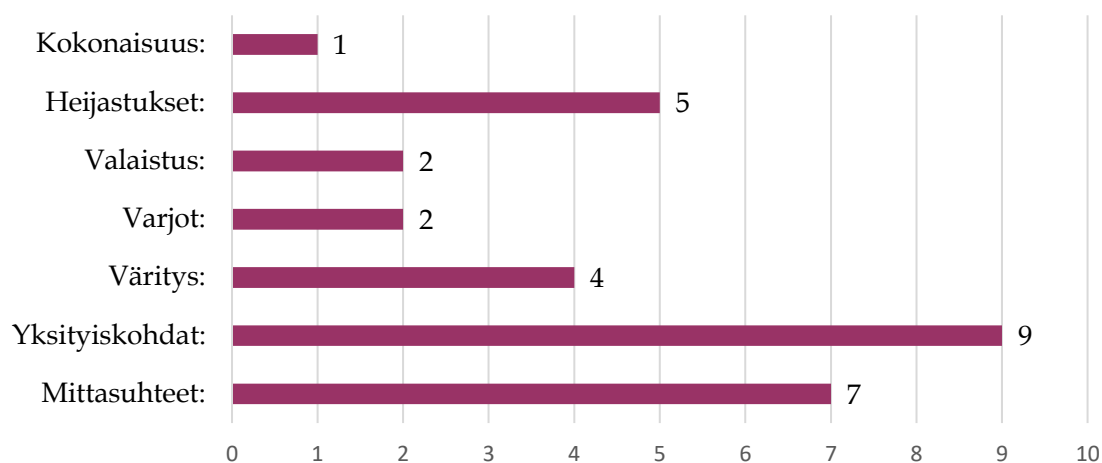


Vapaavalintainen kysymys: Mikä kysymyksen 11 kuvissa onnistuu?



35 vastaajaa.

Vapaavalintainen kysymys: Mikä kysymyksen 11 kuvissa epäonnistuu?

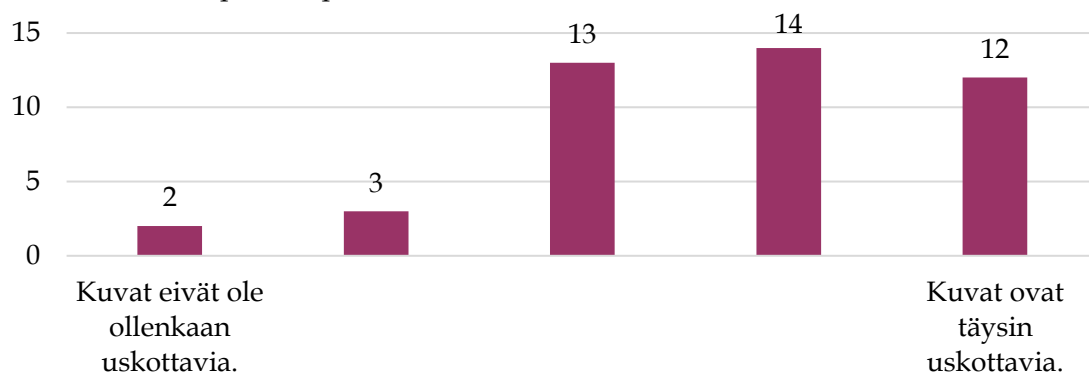


20 vastaajaa.

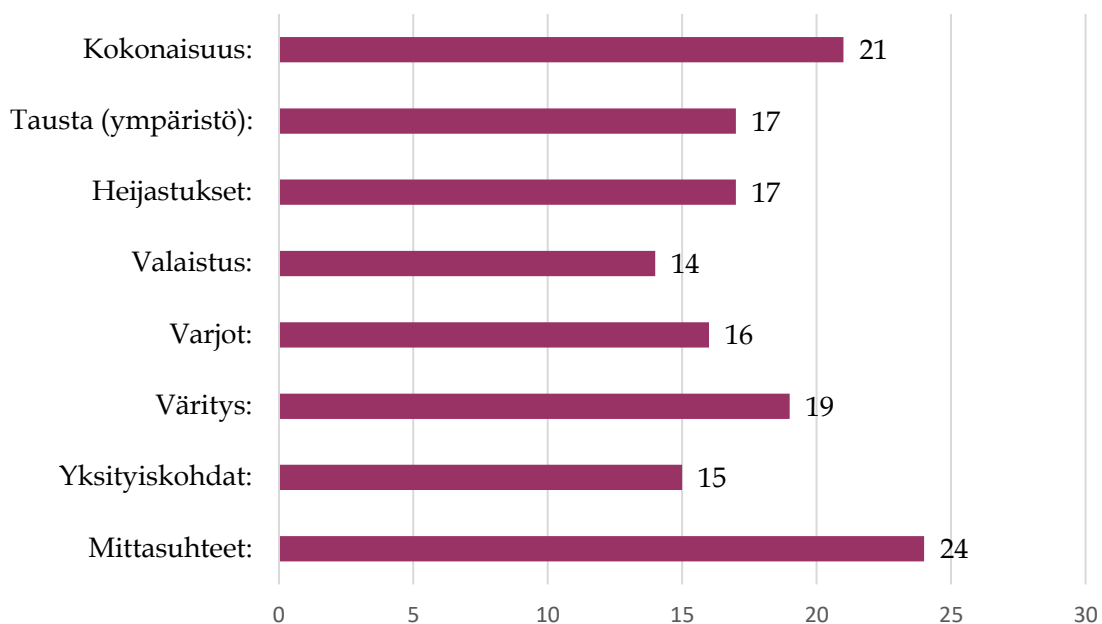
KYSYMYS 12. Miten fotorealistisesti uskottavia generoidut kuvat ovat?



Tuotettu *Text-to-Image*-tekniikalla uuden värin perusteella, mitä ei ollut koulutus-tietokannassa. 768px x 768px.

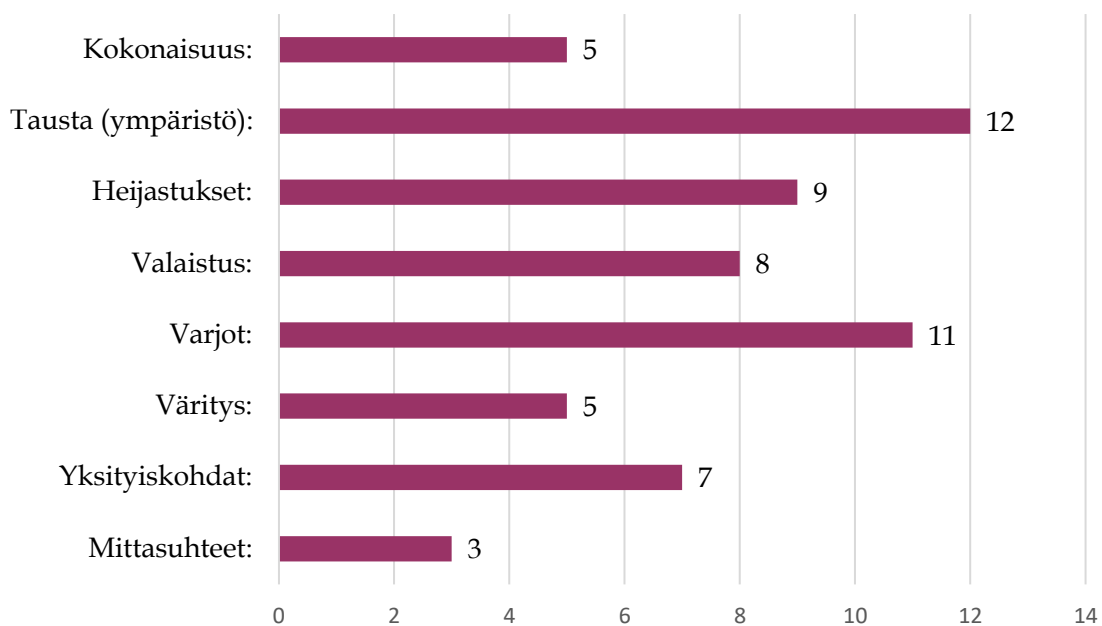


Vapaavalintainen kysymys: Mikä kysymyksen 12 kuvissa onnistuu?



39 vastaajaa.

Vapaavalintainen kysymys: Mikä kysymyksen 12 kuvissa epäonnistuu?

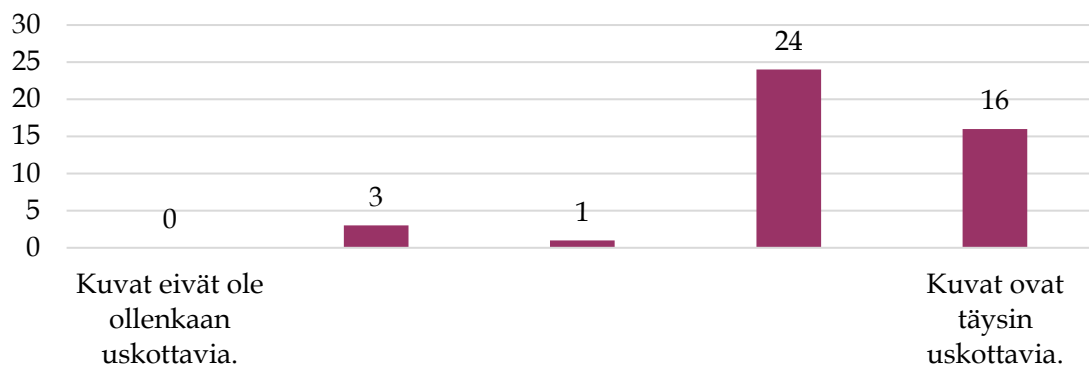


29 vastaajaa.

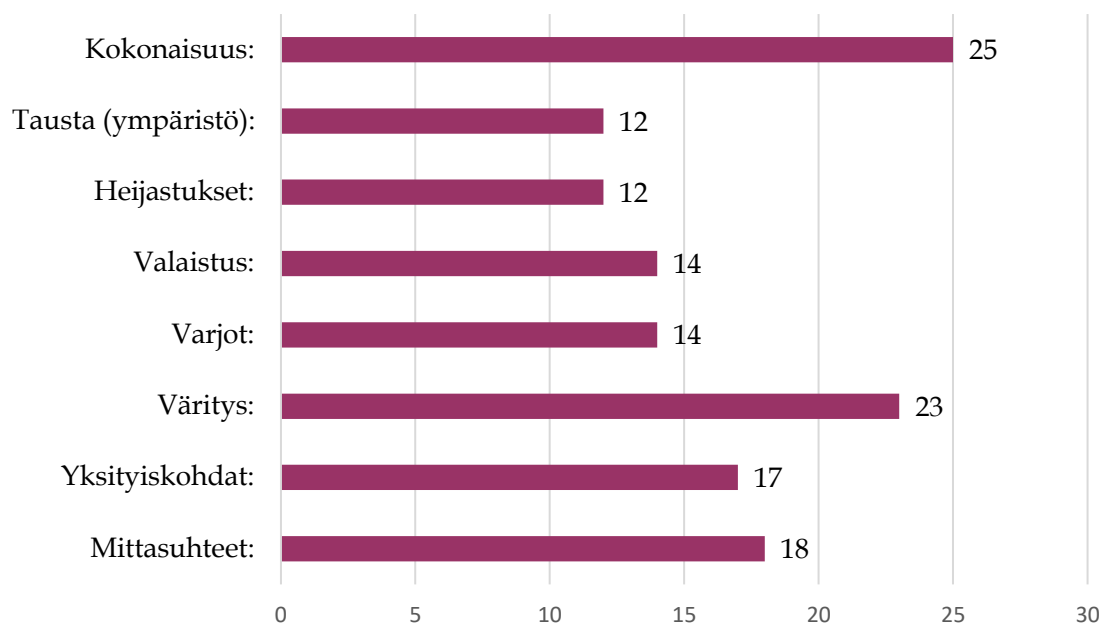
KYSYMYYS 13. Miten fotorealistisesti uskottavia generoidut kuvat ovat?



Tuotettu *Text-to-Image*-tekniikalla uuden värin perusteella, mitä ei ollut koulutus-tietokannassa. 768px x 768px.

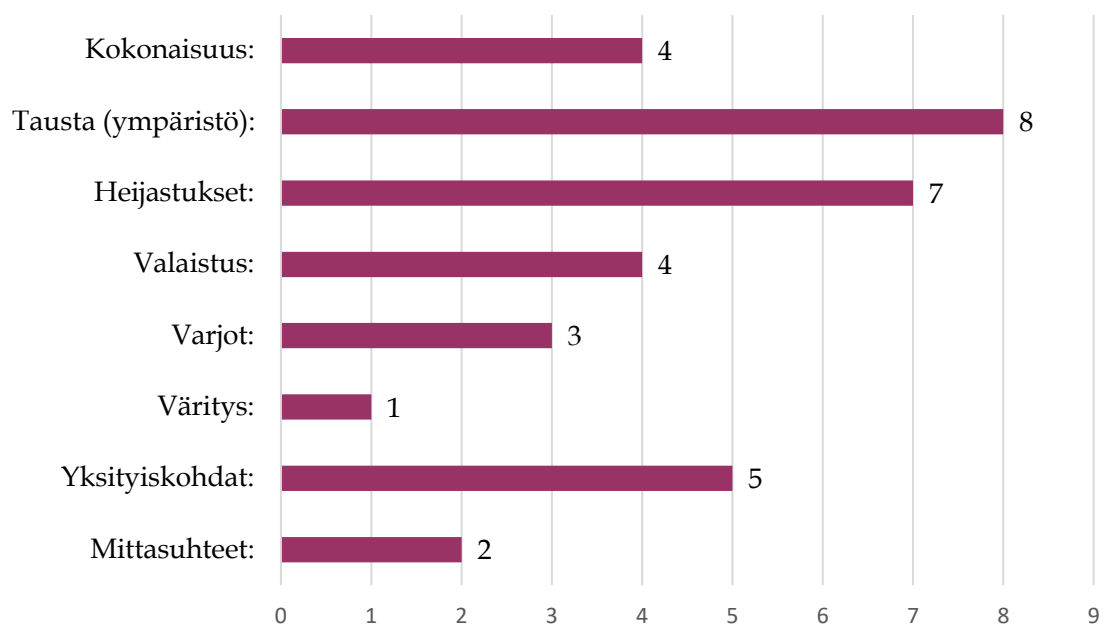


Vapaavalintainen kysymys: Mikä kysymyksen 13 kuvissa onnistuu?



36 vastaajaa.

Vapaavalintainen kysymys: Mikä kysymyksen 13 kuvissa epäonnistuu?



16 vastaajaa.

KYSYMYS 14. Miten fotorealistisesti uskottavia generoidut kuvat ovat?

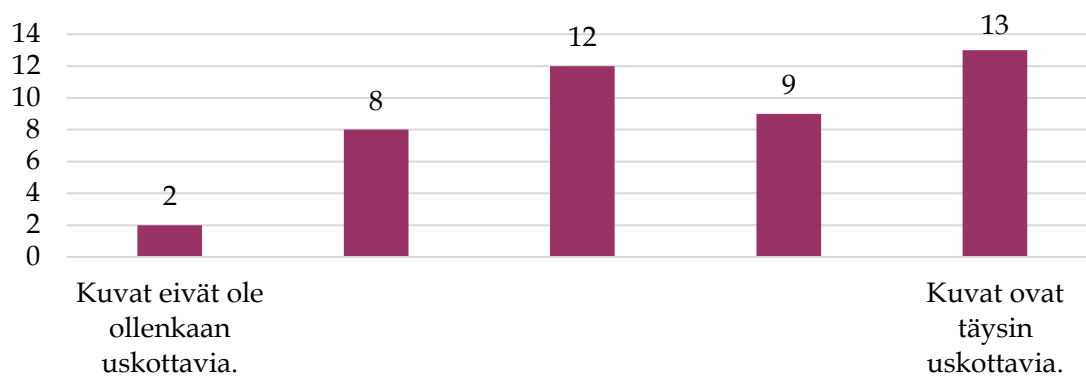


Kysymyksen 14 kuva 1. Tuotettu *Text-to-Image*-, ESRGAN-superresoluutio- ja Inpainting-tekniikoilla koulutustietokannasta löytyvän värin perusteella. 2372px x 3390px.

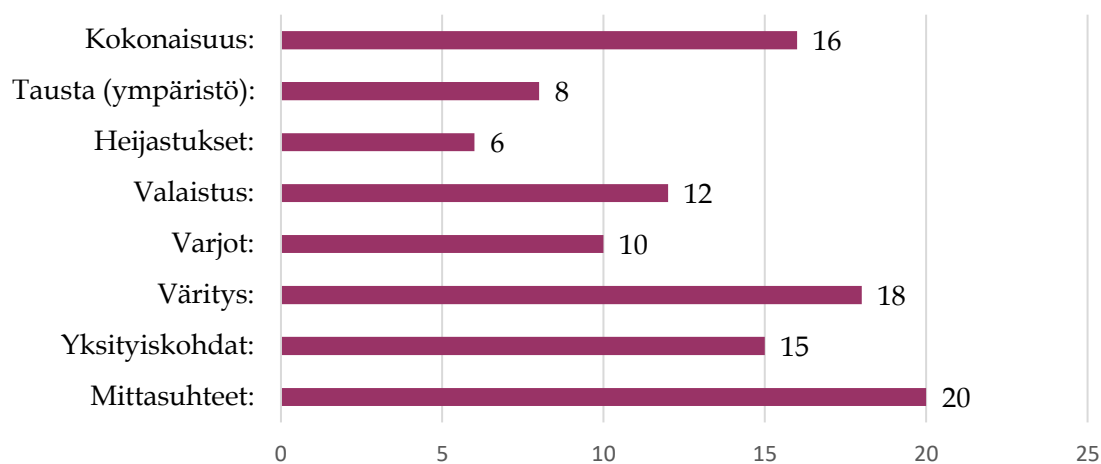


Kysymyksen 14 kuva 2. Tuotettu *Text-to-Image*-, ESRRGAN-superresoluutio- ja Inpainting-tekniikoilla koulutustietokannasta löytyvän värin perusteella. 5760px x 4032px.

KYSYMYS 14 TULOKSET.

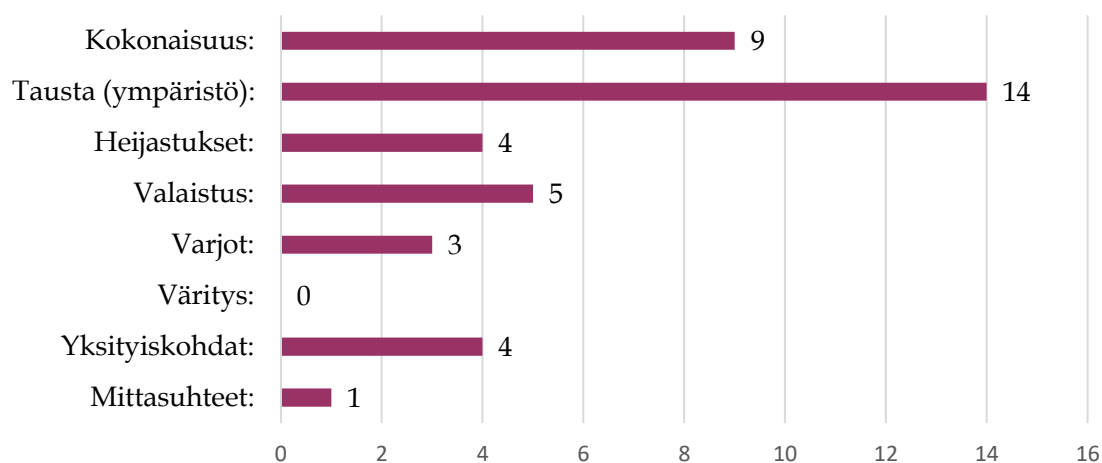


Vapaavalintainen kysymys: Mikä kysymyksen 14 kuvissa onnistuu?



35 vastaajaa.

Vapaavalintainen kysymys: Mikä kysymyksen 14 kuvissa epäonnistuu?



20 vastaajaa.

Kysymys 15. Miten fotorealistisesti uskottavia generoidut kuvat ovat?

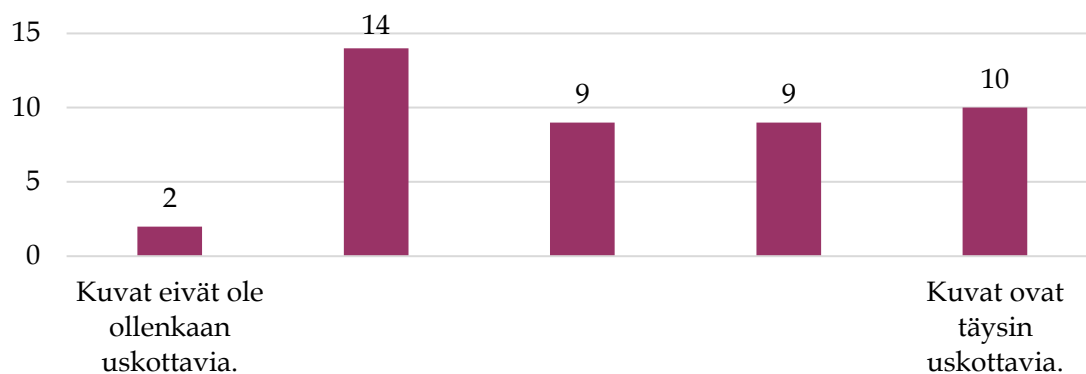


Kysymyksen 15 kuva 1. Tuotettu *Text-to-Image*-, ESRRGAN-superresoluutio- ja Inpainting-tekniikoilla värillä mitä ei ollut koulutustietokannassa. 1920px x 1344px.

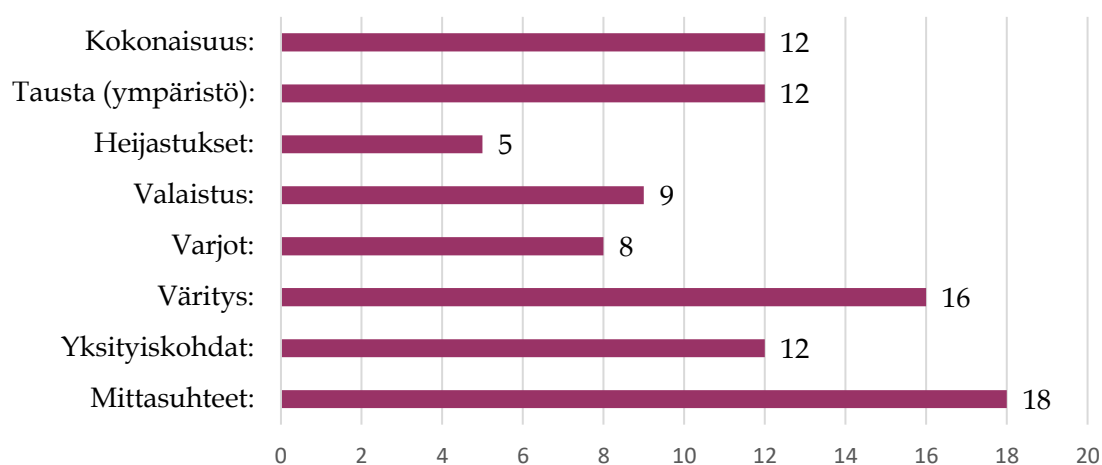


Kysymyksen 15 kuva 2. Tuotettu *Text-to-Image*-, ESRGAN-superresoluutio- ja Inpainting-tekniikoilla värillä mitä ei ollut koulutustietokannassa. 3496px x 4994px.

KYSYMYS 15 TULOKSET.

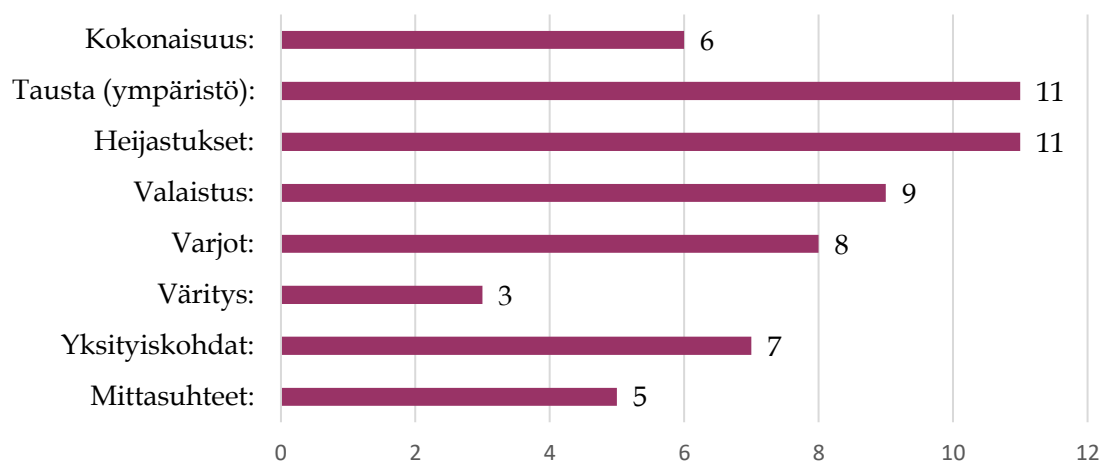


Vapaavalintainen kysymys: Mikä kysymyksen 15 kuvissa onnistuu?



31 vastaajaa.

Vapaavalintainen kysymys: Mikä kysymyksen 15 kuvissa epäonnistuu?



27 vastaajaa.