

Audiopohjainen tekoälygeneraatio musiikin tuotannossa

TURUN YLIOPISTO
Tietotekniikan laitos
TkK-tutkielma
Tietotekniikka
Toukokuu 2026
Santeri Sajamo

TURUN YLIOPISTO
Tietotekniikan laitos

SANTERI SAJAMO: Audiopohjainen tekoälygeneraatio musiikin tuotannossa

TkK-tutkielma, 25 s.
Tietotekniikka
Toukokuu 2026

Audiopohjaisella musiikin generoimisella tarkoitetaan valmiin sävellyksen generoimista äänitiedostoilla koulutetun tekoälymallin avulla. Musiikin generoiminen tällä tavalla on tullut suureen huomioon viime vuosina generatiivisen tekoälyn vauhdikkaan kehityksen johdosta. Generatiivinen tekoäly kykenee tuottamaan uutta sisältöä suurista tietoaaineistoista oppimalla. Nykyiset suuret musiikkigeneraattorit koulutetaan tuhansien, ja jopa miljoonien sävellysten äänitiedostojen avulla.

Tutkielman tavoitteena on tarkastella, miten audiopohjainen musiikin generoiminen on mahdollista, ja mitä tekniikoita nykyiset musiikkigeneraattorit hyödyntävät lopputuloksen saavuttamiseen. Työssä tarkastellaan myös musiikin generoimisen vaikutuksia kulttuurisessa, sekä ammatillisessa kontekstissa.

Työssä tarkastellaan kirjallisuutta tekoälymusiikista kokonaisuutena, mutta erityisesti keskitytään transformer-malleihin perustuviin audiopohjaisiin musiikin generoimistekniikoihin. Näihin tekniikoihin perustuvat musiikkigeneraattorit kykenevät tuottamaan monien minuuttien pituisia kokonaisia kappaleita, eli kuunneltavaksi valmiita äänitiedostoja. Generointi voi olla myös lyhyempien pätkien tuottamista avuksi musiikin luomisprosessiin, mutta tässä työssä keskitytään pääosin valmiiden tuotoksien generoimiseen.

Musiikkigeneraattori voi olla hyödyllinen työkalu niin kokeneille muusikoille, kuin musiikin tuottamisen aloituksesta kiinnostuneille. Musiikin generoiminen ja sen teknologian kehittyminen synnyttää kuitenkin haasteita nykyisille tekijänoikeuksille, sekä artistien työn sopiville korvauksille. Aihe on laajan kritiikin kohteena sekä muusikoilta että musiikin kuuntelijoilta.

Asiasanat: tekoäly, generatiivinen tekoäly, musiikki

Sisällys

1 Johdanto	1
1.1 Taustaksi	1
1.2 Tutkimuskysymykset ja rajaukset	5
2 Teknologia	7
2.1 Neuroverkot	7
2.2 Audiogeneraatio	12
2.3 Musiikkigeneraatio	15
2.4 Musiikkigeneraattorit	16
3 Kulttuuriset vaikutukset	19
3.1 Tekijänoikeuslait	19
3.2 Jakautuvat mielipiteet	20
4 Pohdintaa	22
5 Yhteenveto	24
Lähdeluettelo	26

Kuvat

1.1	Viimeisen kymmenen vuoden kasvu laskentatehossa.	4
1.2	Hakumenetelmät	6
2.1	Transformer-mallin arkkitehtuuri	11

1 Johdanto

Tässä kandidaatintutkielmassa tarkastellaan eri audiopohjaisia tekoälygeneraation malleja ja tekniikoita, keskittyen erityisesti musiikin tuottamiseen. Tekoäly, ja erityisesti generatiivinen tekoäly (engl. artificial intelligence), on noussut viime vuosina keskeiseksi aiheeksi. AI generoitu taide ja musiikki ovat herättäneet sekä teknisiä että eettisiä kysymyksiä. Työssä tutkitaan tarkemmin, miten erityisesti audiopohjainen musiikin generoiminen ja tuottaminen tekoälyllä on mahdollista ja kuinka se tapahtuu. Tutkielman päätavoitteena on tutkia audiopohjaisen tekoälygeneraation tekniikoita, sekä sen vaikutuksia ammatillisessa että kulttuurisessa kontekstissa. Keskeistä on miten eri tekoälymallit saavuttavat tuotoksen, miten mallit eroavat toisistaan, sekä miten tekoälymallit eroavat ihmisen luovasta prosessista. Lopuksi työssä tarkastellaan miltä tekoälymusiikin tulevaisuus mahdollisesti näyttää, ja millaisia vaikutuksia kehityksellä voisi olla musiikin tuottamiseen ja kulutukseen.

1.1 Taustaksi

Musiikki on tärkeässä roolissa monissa maailman kulttuureissa. Musiikin käyttötarkoitukset yltyvät taustamelusta syvällisten ajatusten jakamiseen; lyhyistä jingleistä pitkiin yksityiskohtaisiin sinfonioihin. Musiikki on historiallisesti nähty ihmiskokemuksena, jota ei esiinny muualla luonnossa. Irlantilainen muusikko ja IEEE:n globaalin tekoälyn etiikan taidekomitean perustuspuheenjohtaja tohtori Martin Clancy ehdottaa kirjassaan ”Artificial Intelligence and Music Ecosystem”, että musiikkia

voidaan ajatella ääneksi, jonka äänen tekijä on itse kategorisoinut musiikiksi. Vaikka luonnossa jotkut eläimet, kuten linnut, laulavat tai luovat muita rytmillisiä ääniä, ei tätä kategorisoida yleensä kuitenkaan musiikiksi. [1] Musiikki on siis abstrakti ihmiskonsepti.

Generoitu musiikki on mahdollista musiikin sääntöjen olemassaolon takia, jotka ihminen oppii tajuamattaan kuunnellessaan musiikkia. Tätä kutsutaan musiikin-teoriaksi. Musiikinteoria vaihtelee paikoittain, sillä nämä säännöt eivät ole absoluuttisia, vaan perustuvat vain ihmisten tottumuksiin musiikista. Eri kulttuureissa musiikki siis perustuu erilaisiin sääntöihin. [2] Tässä tutkielmassa käsitellään musiikkia länsimaisesta näkökulmasta, joka on hyvin syvällisesti tutkittua, ja jota voitaisiin kutsua jopa laskelmoiduksi.

Matemaattinen ajatus musiikista juontaa juurensa antiikin Kreikkaan. Kreikkalainen filosofi ja matemaatikko Pythagoras (570–495 eaa.) kehitti matemaattisten teorioiden ohella myös paljon teorioita musiikista. Pythagoras ajatteli, että universumi perustuu kokonaan matematiikkaan, ja siten myös musiikin käsitteitä voi selittää matematiikan avulla. Hän huomasi miellyttävien sointujen perustuvan matemaattisiin suhteisiin, ja laskelmillaan kehitti oktaavin ja muita intervaleja, eli nuottien välisiä etäisyyksiä. [3]

Nykyisin käytetty neljään nuottiviivaan perustuva nuottikirjoitus, tai notaatio, eli musiikin kirjoittaminen muotoon, jota soittaja voi soittaa säveltäjän tarkoittamalla tavalla, keksittiin 1000-luvun alussa keskieuropassa. Musiikkia kirjoitettiin muiden soitettavaksi jo ennen tätä, mutta notaatio oli paljon löysempää, eikä sisältänyt tarkkaa tietoa sävelkorkeuksista. Soittajat siis usein improvisoivat paljon, ja musiikki kuulosti hyvin erilaiselta riippuen esittäjästä. [4] Nykyisin laajimmin käytetty, varsinkin länsimaisessa musiikissa, on 12 tasaviritysjärjestelmä, joka keksittiin 1500-luvulla. Järjestelmässä oktaavi on jaettu 12 säveleen, joiden välit on laskettu logaritmisesti kuulostamaan samoilta etäisyyksiltä toisiinsa.

Algoritminen musiikki viittaa musiikkiin, jossa pääosa sävellyksestä tapahtuu ilman ihmisvaikutusta. Esimerkki yksinkertaisesta algoritmisesta musiikin säveltämisestä voisi olla kahden 6-puolisen nopan heittäminen, jossa jokaiselle noppien numerolle annetaan yksi nuotti. Näitä noppasävellyspelejä keksittiin useampia 1700-1800-luvulla, ja eri pelit käyttivät eri tekniikoita noppien avulla säveltämiseen. [5] Tässä tapauksessa ihminen on siis kuitenkin heittämässä noppaa, ja on usein säveltänyt itse jo valmiiksi osia musiikista. Tekoälymusiikki vie algoritmisen säveltämisen uudelle tasolle, jossa ihmisen vaikutus sävellykseen pienentyy huomattavasti.

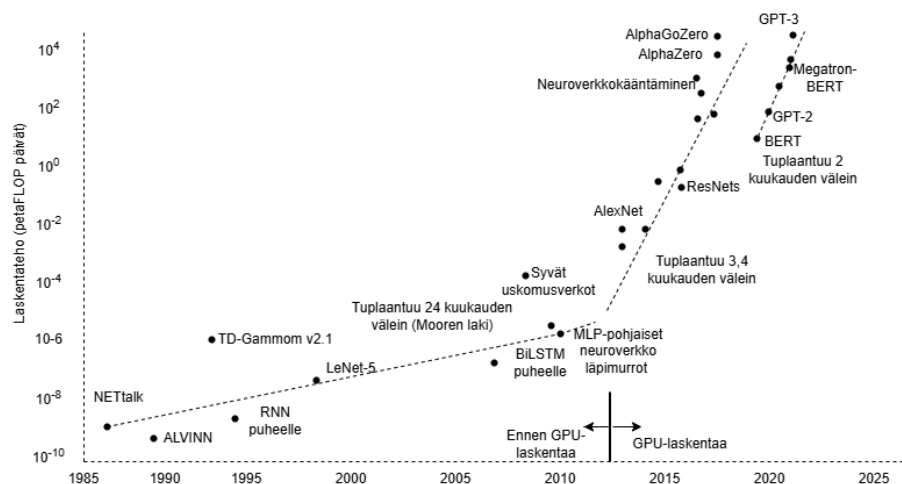
Tietokoneilla on tuotettu musiikkia jo tietokoneiden keksimisestä lähtien. Audio-työasema (engl. Digital Audio Workstation, DAW) on termi musiikin tuottamiseen tarkoitettulle ohjelmalle. DAW on siis vain alusta musiikin tekemiseen tietokoneella, ja sen tarkoitusta voi verrata vaikka kuvanmuokkaus- tai videoneditointiohjelmiin. DAW:in sisällä voi olla monia musiikin tuottamista helpottavaa ominaisuutta ja lisäohjelmaa, mutta tämä rajataan kuitenkin työkalujen puolelle. Musiikin generoimisella ei tarkoiteta kaikkea tietokoneella tuotettua musiikkia, vaan tarkalleen tietokoneen itse säveltämää algoritmista musiikkia.

Tietokoneen algoritmeihin perustuva musiikin kirjoittaminen keksittiin jo 1950-luvulla [6], mutta tuolloiset algoritmit ja niillä tuotetut sävellykset olivat hyvin yksinkertaisia. Tekniikkaa käytettiin yleisemmin vain automatisoimaan osaa musiikin tuottamisprosessista. 1980-luvulla tietokoneiden ja elektronisen musiikin kehityksen myötä myös algoritminen musiikki yleistyi.

Tekoäly konseptina oli ajatuksissa jo antiikin aikoihin. Antiikin kirjoittajat tarinoivat elävistä patsaista ja jumalien takomista pronssisista sotilaista, jotka pystyivät ajattelemaan ja toimimaan itsenäisesti. [7] Tekoäly-termillä viitataan nykyään kuitenkin ihmisälyä imitoivaan koneeseen, lähinnä tarkoittaen koneoppimisen avulla koulutettua ohjelmaa.

Koneoppiminen on tekoälytutkimuksen alaluokka, jossa tutkitaan algoritmeja, jotka mahdollistavat sen, että kone oppii datasta ja keksii oppimansa perusteella uutta dataa. Tämä mahdollistaa koneen oppimaan itse ratkaisemaan monimutkaisia ongelmia, joiden ratkaisuihin sääntöjen keksiminen olisi muuten epäkäytännöllistä ja liian monimutkaista.

Koneajattelu ja tekoäly tulivat puheenaiheeksi jo 1940-luvulla, jo heti tietokoneen keksimisen jälkeen. 1950-luvulla tekoälyyn keskitettiin paljon resursseja, mutta nopeasti huomattiin, että kehitys on paljon luultua hitaampaa. Prosessointiteho ihmisen kaltaiseen ajatteluun olisi valtava. Tästä syystä, vaikka tekoäly onkin ollut tunnettu konsepti jo pidemmän aikaa, on se tullut suurempaan huomioon ja menestykseen vasta lähivuosina, kun prosessointiteho on ollut tarpeeksi suuri ja tarpeeksi tehokkaita algoritmeja on kehitetty. AI-boomilla viitataan tekoälyn räjähdysmäiseen kasvuun, joka sai alkunsa vuosina 2010-2016, ja kiihtyi 2020-luvulla.[8] Tätä kasvua on mallinnettu kuvassa 1.1.



Kuva 1.1: Viimeisen kymmenen vuoden kasvu laskentatehossa. Kuva mallinnettu lähteen mukaan. [8]

Nykyinen tekoäly on kuitenkin vielä epätäydellistä nimeensä nähden. Tekoäly ei aidosti ajattele kuin ihminen, vaan päätökset purustuvat moniin kerroksiin algoritmeja ja valintoja, joihin malli on koulutettu. Tätä kutsutaan Deep Neural Networ-

kiksi (DNN), eli syväksi neuroverkoksi. DNN on monitasoinen verkko, joka emuloi ihmisen ajatusta sille annettujen parametrien perusteella. [9]

Tekoäly, ja tarkemmin generatiivinen tekoäly, on nykyään valtava tieteenala, josta julkaistaan jatkuvasti uusia tutkimuksia, ja jonka tekninen kehitys on ollut lähi-vuosina sokaisevan nopeaa. Vaikka tekoälytutkimuksen huomio usein keskittyy suurien kielimallien mahdollisuuksiin ja kuva- ja videogeneraatioon, on audiogeneraation kehitys myös yhtä huomattavaa. Audiogeneraatiotekniikoiden kehityksen myötä mielenkiintoiseksi osa-alueeksi on noussut musiikkigeneraatio, joka on saanut paljon huomiota lähiaikoina AI-musiikin noustessa kuunnelluimpien kappaleiden listoille ympäri maailmaa.

Suoratoistopalvelut, musiikin harkitsematomampi kulutus, sekä tekoälygeneroidun musiikin helppous ja tästä seurannut suuri levinneisyys, ovat päätekijöitä aiheen tutkimusten lisääntymiseen.

1.2 Tutkimuskysymykset ja rajaukset

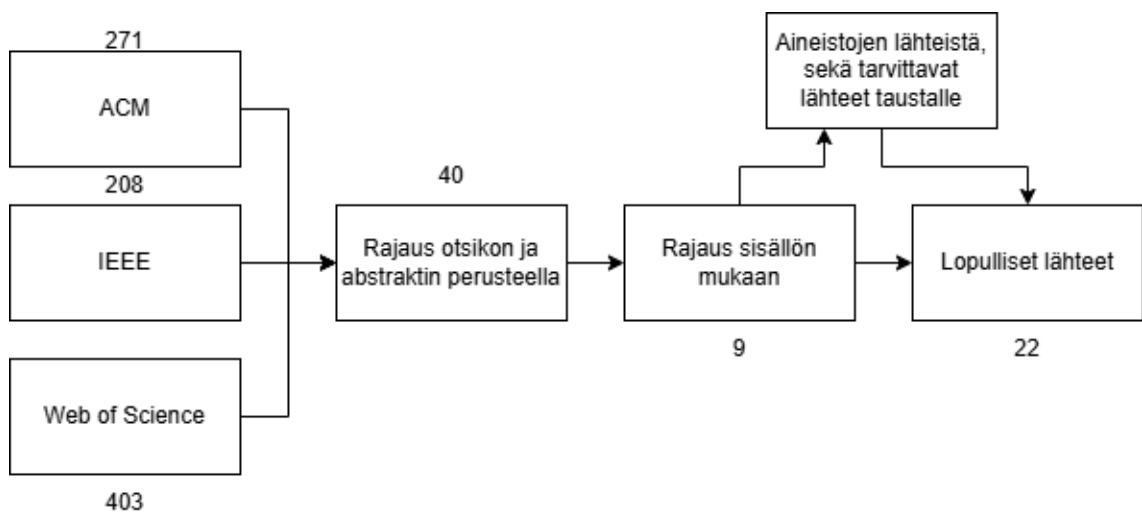
Tässä tutkielmassa keskitytään erityisesti valmiin musiikkituotoksen generoimiseen, eli äänitiedostoon, jota voi suoraan kuunnella. Tekoällyn generoima tuotos ei siis vaadi ihmiseltä muuta syötettä, kuin esimerkiksi alussa kirjoitettuja ohjeita musiikin tyylistä, genrestä, nopeudesta tai muista ominaisuuksista. Tekoälymusiikki on tutkimusaiheena lähtökohtaisesti monialainen. Tästä syystä on hyvin hedelmällistä tarkastella aihetta monesta eri näkökulmasta. Tärkeimpinä keskipisteinä ovat kuitenkin tietotekniikka ja musiikki, mutta tutkin myös ekonomisia, moraalisia ja poliittisia vaikutuksia.

Tässä tutkielmassa tarkastellaan, mikä mahdollistaa audiopohjaisen musiikin generoinnin. Tarkastellaan audiopohjaisten tekoälygeneraation mallien teknisiä puolia, sekä eri tekniikoita tuloksen saamiseen. Tutkitaan myös, mitä eroja on ihmisen ja tekoällyn tuottamalla musiikilla, ja pystyykö tekoälyllä tuotettua musiikkia erotta-

maan ihmisen tekemästä musiikista. Pohditaan, tuleeko AI-musiikin tunnistaminen vaikeutumaan lähitulevaisuudessa, vai lähenemmekö AI-musiikin kehityksen rajoja. Tarkastelun kohteena on myös miltä tulevaisuus näyttää musiikin tuottamiselle ja musiikkialalle, ja miten uudet tekoälyteknologiat ja tekniikat tulevat vaikuttamaan tuottamisprosessiin. Lisäksi tarkastellaan, miten tekoälymusiikin kehittyminen ja sen suosion nousu vaikuttaa musiikkiekonomiaan, ja miten AI-musiikin seurauksena syntyvä musiikin määrällinen paljous tulee muuttamaan musiikin myyntiä ja musiikoiden uramahdollisuuksia.

Tiedonhaku toteutettiin Google Scholarilla sekä ACM Digital Library, IEEE Xplore ja Web of Science tietokannoilla. Rajasin hakua siten, että abstraktissa esiintyisivät termit *artificial intelligence*, *music* ja *gener**. ACM-tietokannassa hakulause oli [Abstract: "artificial intelligence"] AND [Abstract: music] AND [Abstract: gener*].

Tiedonhaun ensimmäisessä vaiheessa rajasin tuloksia abstraktin ja otsikon perusteella. Toisessa vaiheessa luin sopivia aineistoja läpi, ja viimeisessä vaiheessa valitsin luetuista aineistoista sopivimmat. Lisäksi tutkielmassa on käytetty lähteinä aineistoja, jotka ovat löytyneet käytettyjen aineistojen lähteistä.



Kuva 1.2: Hakumenetelmät

2 Teknologia

Seuraavissa luvuissa tarkastellaan tarkemmin generatiivisen tekoälyn toimintaperiaatteita ja sitä, miten ne mahdollistavat musiikin generoimisen. Generatiivisen tekoälyn käyttö monilla tieteenaloilla on yleistynyt suuresti viime vuosina, ja sen käyttöä voidaan pitää jopa oletuksena useilla aloilla. Tekoäly helpottaa erityisesti työläiden ja mekaanisten osuuksien tekemistä, mikä nähdään tässä kontekstissa yleisesti positiivisena kehityksenä. [10]

2.1 Neuroverkot

Neuroverkot (engl. Artificial Neural Networks, ANN) ovat koneoppimisen alaluokka. Ne ovat ihmisen aivojen mukaan mallinnettuja laskentamalleja, joilla tietokone pystyy suorittamaan komplekseja tehtäviä, jotka vaatisivat muuten ihmisen tasoista tietoisuutta tai älykkyyttä. Neuroverkko koostuu solmuista ja liitoksista, joilla mallinnetaan vastaavasti aivojen hermosoluja (neuroneja) ja synapseja. Neuronien yhteyksille asetetaan painoja, jotka määrittävät niiden välillä liikkuvan datan tärkeyttä tai oikeellisuutta. [11]

Neuroverkko voi koostua monista kerroksista, ja tätä kutsutaan syväksi neuroverkoksi (engl. Deep Neural Network, DNN). Vastakkaisesti jos neuroverkossa on vain yksi kerros, kutsutaan sitä ohueksi neuroverkoksi (engl. Shallow Neural Network). Syväoppiminen tuli tietoisuuteen laajemmalle akateemiselle yleisölle vuonna 2015 artikkelissa "Deep Learning" (Yann Lecun et al.), jossa koottiin yhteen syväoppi-

misen tekniikoita ja käytännön käyttökohteita. Nykyään syväoppimisen mallit ovat oletuksena tekoälymallien kehityksessä. [9]

Neuroverkko pitää aivan aluksi kouluttaa. Koulutusvaiheessa säädetään neuroverkon painoiksi kutsuttuja parametreja, jotta voidaan parantaa mallin tarkkuutta. Koulutuksen jälkeen neuroverkko pystyy oikeilla painoilla tunnistamaan tärkeät osat syötteestä ja käyttämään dataa uuden datan luomiseen. Neuroverkon kouluttamiseen tavitaan suuria määriä koulutusdataa, jotta malli olisi käyttökelpoinen.

Neuroverkkomalleja on monia erilaisia, ja ne erikoistuvat erilaisiin tehtäviin. Jotkut mallit on parempia kuvien tunnistamisessa, ja jotkut toimivat paremmin vaikka kielten kääntämisessä. Yksi neuroverkkojen hyödyllinen ominaisuus on se, että niitä voidaan käyttää myös muihin tarkoituksiin kuin mihin ne alunperin kehitettiin. Tulokset ovat kuitenkin sellaisessa tapauksessa yleensä huonoja, mutta malli voi selviytyä myös yllättävänkin hyvin ongelmista, joita varten sitä ei ole kehitetty. Tämä oppiminen saa neuroverkot soveltumaan erinomaisesti komplekseihin tehtäviin, kuten musiikin tuottamiseen.

Diffuusiomallit ovat neuroverkkomalleja, joilla generoidaan uutta dataa. Ne toimivat lisäämällä Gaussin kohinaa koulutuslähteeseen, kunnes jäljellä on pelkkää tunnistamatonta kohinaa. Sitten malli opetetaan tekemään sama prosessi takaperin. Koulutuksen jälkeen malli pystyy luomaan uusia tuotoksia pelkästä satunnaisesta kohinasta. [12] Tämä tekniikka kehitettiin alun perin kuvien tuottamiseen, mutta se on löytänyt käyttöä myös äänen, ja musiikin tuottamisessa. [13]

Alkutilana voidaan käyttää äänitiedoston spektrogrammia, joka näyttää äänitiedoston taajuudet ajan funktiona. Datapisteistä kehittyy ns. latent space, eli piilevä avaruus. Piilevä avaruus on ääriviiva datasta eli kompressoitua dataa, joka ei tarkkaan kerro, miltä alkuperäinen data näyttää. Dekooderi sitten yrittää luoda alkuperäisen spektrogrammin tästä piilevästä avaruudesta. Dekooderi siis yrittää piirtää kokonaiskuvan pelkkien ääriviivojen avulla ja yrittää päästä mahdollisimman lähelle

alkuperäistä. Jotta pystyttäisiin luomaan uutta musiikkia, pitää käyttää vain pientä osaa piilevästä avaruudesta ja datapisteistä, jotta pystytään sallimaan muutoksia. [13]

Autoregressiiviset mallit ovat koneoppimisen luokka, jossa mallit ennustavat tulevaa dataa sarjassa aikaisemman datan perusteella. Autoregressio on aikasarja-analyysissä käytetty tekniikka, jossa oletetaan, että sarjan arvo pystytään laskemaan aikaisempien arvojen funktiona. Mallit ovat hyvin käteviä datan ennustamisessa ajan funktiona, ja niitä käytetäänkin esimerkiksi sään ja tieolosuhteiden ennustamiseen.

Nykyään monet generatiiviset tekoälymallit ovat autoregressiivisiä. Mallille voidaan kouluttaessa antaa vain rajallinen määrä dataa, mutta autoregressiivisyys mahdollistaa uuden aikaisemmin olemattoman datan keksimisen. Audiopohjainen musiikkigeneraatio perustuu siis myös autoregressiivisyyteen. Seuraavaksi esitellään muutamia malleja, joita voisi käyttää musiikin generoimiseen:

CNN (Convolutional Neural Network) -mallit käyttävät konvoluutiokerroksia (Convolutional Layer) sekä koontikerroksia (Pooling Layer) erottaakseen yksityiskohtia sekä vähentääkseen syvyyttä. CNN-mallit on kehitetty erityisesti kuvien ja videoiden prosessointiin ja tunnistukseen. Tekniikkaa on myös käytetty musiikin generoimiseen. [14]

RNN (Recurrent Neural Network) -mallit luovat tulosteen jokaisena ajankohtana ja käyttävät aikaisemman ajankohdan tulostetta uuden tulosteen luomiseen. Täten jokainen tuloste riippuu kaikista aikaisemmista tulosteista. RNN-malli soveltuu sekventiaalisen datan analysointiin ja soveltuu siten myös yksinkertaiseen musiikin generoimiseen. [15]

LSTM (Long Short-Term Memory Network)-malli on tietynlainen RNN-malli, joka soveltuu pidemmän aikavälin datan analysoimiseen ja generoimiseen. Tämä on erityisen hyödyllistä musiikin generoimisessa, sillä musiikillinen kokonaisuus on

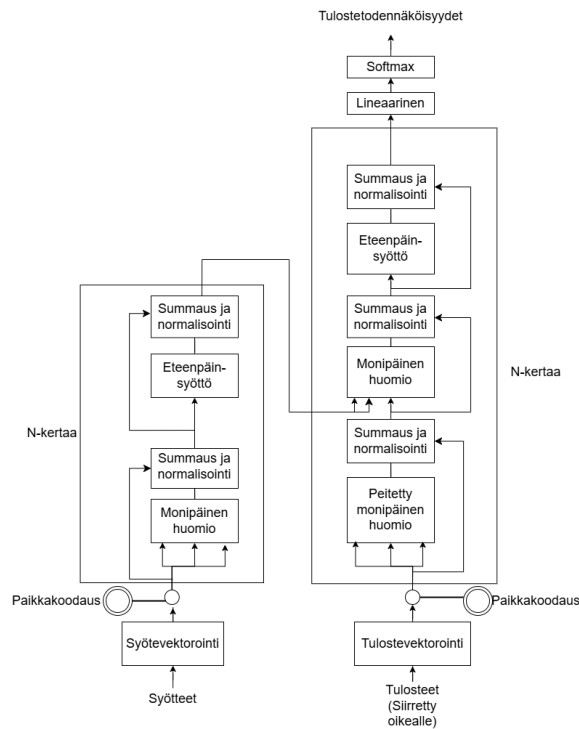
riippuvainen aikaisemmasta kontekstista. Jos aikaväli on liian lyhyt, menetetään konteksti ja seuraava osa generoitua musiikkia ei sovi aikasemmin soineen kanssa. LSTM-malleilla ei ole kuitenkaan tarpeeksi suurta muistia kokonaisen kappaleen generoimiseen, vaan ne pystyvät lähinnä tuottamaan lyhyitä musiikillisia pätkiä. [16]

Transformer-mallit

Vuonna 2017 Googlen julkaisema "Attention Is All You Need"[17] tutkimusraportti oli läpimurto koneoppimisessa. Julkaisussa esiteltiin transformer-malli (kuva: 2.1), joka pystyy itsehuomioksi kutsutun tekniikan avulla tunnistamaan datan kuvioita. Ennen transformereita neuroverkon kouluttaminen vaati suuria määriä leimattua dataa, jossa leimoilla kuvaillaan datan sisältöä. Datan kuvioiden tunnistamisen ansiosta näitä leimoja ei enää tarvita, vaan malli pystyy itsenäisesti oppimaan sille annetun datan pohjalta ja luo siitä eräänlaisen kartan, joka kuvaa asioiden läheisyyttä ja yhteyttä toisiinsa.

Itsehuomiolla tarkoitetaan, että malli merkkää sille syötettyä sekä sen tuottamaa dataa. Transformer-malleilla on myös hyvin pitkä muisti, joka mahdollistaa kaukaisien datapisteen yhdistämisen toisiinsa. Näin malli ymmärtää kontekstin sille syötetystä datasta. Mallia testattiin ensin kielten kääntämiseen ja tekstin generoimiseen, koska siitä huomaa hyvin itsehuomion toimivuuden. Pelkkien sanojen ymmärtäminen ei vielä riitä lauseen tarkoituksen ymmärtämiseen. Pitkän lauseen ymmärtämiseen tarvitaan pitkää muistia, että saadaan pidettyä kiinni kontekstista lauseen lopussakin. Olennaista on siis, että malli ymmärtää sanan merkityksen muuttuvan sitä edeltävien, mutta myös sitä seuraavien sanojen perusteella.

Data pilkotaan ensin pienempiin osiin, tokeneihin. Tätä prosessia kutsutaan tokenisaatioksi. Tokenit ovat numeromuodossa esitettyä dataa, jonka malli pystyy ymmärtämään. Esimerkiksi suurissa kielimalleissa yksi tokeni voi vastata yhtä lyhyttä



Kuva 2.1: Transformer-mallin arkkitehtuuri. Kuva mallinnettu lähteen kuvan mukaan. [17]

sanaa ja pidemmät sanat voidaan jakaa muutamaan tokeniin. Kun sanat sisältävät samoja tokeneita, malli pystyy päättämään, että näillä sanoilla on todennäköisesti jokin yhteys. Tokeneilla pystytään myös välttämään homonyymien sekoittaminen, sillä sanat, joilla on eri tarkoitus, vaikka ne kirjoitettaisiinkin samalla lailla, saavat erinumeroisen tokenin. Tokenisaatio mahdollistaa tehokkaamman kolutus- sekä päättelyvaiheen.

Kontekstin ymmärtämisen takia transformer-mallit soveltuvat erinomaisesti musiikin generoimiseen. Musiikin säveltämistä voidaan ajatella samalla lailla kuin lauseen rakentamista. Kun lauseessa sanojen merkitys ja paino muuttuu niitä ympäröivien sanojen perusteella, niin musiikissa nuotin tarkoitus muuttuu sitä ympäröivistä nuoteista. Musiikin generoimisessa oleellista on myös transformer-mallien pitkä muisti, sillä tavoitteena on generoida kokonaisia minuuttien pituisia kappaleita, ei vain lyhyitä katkelmia. [18]

Musiikin generoimiseen käytetään nykyään laajimmin transformer teknologiaan perustuvia malleja. Transformer-mallit ovat myös hyvin yleisiä muissa käyttötarkoituksissa. Valtaosa tehokkaista suurista kielimalleista, kuten OpenAI:n ChatGPT, perustuvat transformer teknologiaan. Transformer-malli onkin nähty mullistavana kehityksenä tekoälymaailmassa. [19]

2.2 Audiogeneraatio

Ennen kuin musiikkia pystyisi generoimaan valmiissa kuunneltavissa olevassa muodossa pitää pystyä generoimaan ääntä. Audiogeneroinnin kehityksessä on ollut kaksi keskeistä haastetta: äänenlaatu ja tuloksen pitkäaikainen yhtenäisyys. Pitkäaikaisella yhtenäisyydellä tarkoitetaan mallin kykyä ymmärtää itse generoimansa datan kontekstia, jotta pidempien järkevien äänitiedostojen tuottaminen olisi mahdollista.

Audiogeneraation tutkiminen on pääosin keskittynyt puheen generoimiseen ja tunnistamiseen esimerkiksi automaattisia kääntäjiä varten [14], [17]. Puheen tuoton teknologian kehittyessä on kuitenkin saatu aikaan ikään kuin vahingossa malleja, jotka pystyvät tuottamaan musiikkia. Tämä johtuu siitä, että uudet puhetta varten kehitetyt mallit oppivat tunnistamaan äänitiedostosta puhetta varten tärkeät kohdat, kuten inflektiot, kadenssit ja äänenkorkeudet, ja tämä sama periaate toimii hyvin myös musiikin kanssa.

Tarkastellaan seuraavaksi kahta esimerkkiä audiogeneraatiomalleista, jotka havainnollistavat audiogeneraation haasteita, sekä näyttävät teknologian kehityksen kulkua.

WaveNet

Vuonna 2016 julkaistu raportti ”WaveNet: A Generative Model for Raw Audio” Oord et al. [14] tutkii kehitetyn autoregressiivisen syväoppineen WaveNet mallin kykyä tuottaa uskottavaa ihmispuhetta. Malli kehitettiin puheen tuottamiseen, mutta tut-

kimuksessa huomattiin, että malli kykenee myös generoimaan musiikillisia katkelmia. WaveNet toimii esimerkkinä aikaisemmasta audiopohjaisesta koneoppimismallista.

Tutkimusten pääaiheena oli TTS (text to speech) eli tekstin muuttaminen puheeksi, mutta koska WaveNet perustuu kokonaan audiopohjaiseen generaatioon, kehitettiin mallilla myös musiikin tuottamista. Saatiin selville, että malli kykeni tuottamaan lyhyitä klassisen musiikin katkelmia, jotka olivat "mieluisan"kuuloisia, ja muistuttivat musiikkia. Malli koulutettiin musiikin generoimiseen kahden tietokannan kanssa. Toinen sisältää alle 30 sekunnin mittaisia musiikkileikkeitä, jotka on varustettu genreä, instrumentaatiota, tempoa, volyyymiä sekä tunnetilaa kuvaavilla tunnisteilla. Toinen sisältää pianomusiikkia. Käyttämällä binaarivektoreita pystyttiin ohjaamaan valmiin tuotoksen eri piirteitä, jotka johtivat haluttuun tietynlaiseen musiikkiin.

Tutkijat huomasivat, että musiikin tuottamiseen tarvittiin laajempi reseptiivinen kenttä. Jos kenttä oli liian kapea, mallit eivät pysyneet johdonmukaisina, ja generoitu musiikki ei ollut koherenttia. Mallilla ei siis ollut tarpeeksi pitkää muistia musiikin generoimiseen, ja menetti kontekstin. Musiikin generoiminen mallilla oli muutenkin mahdollista koulutusmateriaalin sisältämien valmiiden tunnisteiden, sekä yksinkertaisuuden takia.

Julkaisunsa aikaan WaveNet oli liian laskennallisesti raskas ollakseen käytännöllinen kuluttajille. Googlen DeepMind tiimi kehitti WaveNet-mallia huomattavasti alkuperäisestä, ja pian malli oli jopa 1000 kertaa nopeampi generoimaan raakoja äänitiedostoja verrattuna ensimmäisessä raportissa esiteltyyn malliin. Lokakuussa 2017 Google otti käyttöön WaveNet teknologian GoogleAssistantille generoimaan englannin ja japanin kielistä puhetta. [20]

AudioLM

Aikaisemmat audiogeneraatiomallit joutuivat valitsemaan joko korkealaatuisen, tai pitkäkestoisen äänitiedoston väliltä. AudioLM on kesällä 2023 julkaistu audiogeneraatiomalli, joka ratkaisi ongelman käyttämällä hybridimallia, jossa yhdistettiin semanttisten ja akustisten tokenien käyttöä. Semanttisilla tokeneilla kuvataan puheessa sanojen merkitystä, ja musiikissa melodiaa ja rytmiä. Akustisilla tokeneilla taas kuvataan tiedoston äänellisiä yksityiskohtia, eli akustiikkaa. Tokeneiden yhdistäminen mahdollistaa korkealaatuisten ja pitkäkestoisten äänitiedostojen generoinnin.

Käytetyt akustiset tokenit saattin SoundStream neuroaudiokoodekilla (engl. codec, coder/encoder) ja semanttiset tokenit w2v-BERT:llä. Semanttiset tokenit mahdollistavat pitkäaikaisen jatkuvuuden, ja akustiset tokenit taas korkean äänen laadun.

Tekniikan sovellettavuuden todistamiseksi mallilla suoritettiin kaksia tehtäviä: "Puheen jatkuvuus, jossa mallin odotetaan pitävän puhujan identiteetti, prosodia ja kehotteen äänitysolosuhteet ja tuottamaan uutta sisältöä, joka on syntaattisesti oikeata ja semanttisesti johdonmukaista." sekä "Pianojatkuvuus, jossa mallin odotetaan tuottavan pianomusiikkia, joka on melodisesti, harmonisesti ja rymillisesti koherenttia kehotteen kanssa."

Puhegeneraatio testi oli hyvin onnistunut, ja malli sai parannettua aikaisempiin verrattuna puheen äänen laatua, sekä ymmärrettävyyttä. Pianojatkuvustestissä akustisia parametreja muutettiin hieman, mutta muuten malli toimi samalla tavalla. Tutkijat vertasivat AudioLM rakenteen kanssa generoitua musiikkia pelkästään akustisia tokeneita käyttävien mallien tuotoksien kanssa. Tutkijat huomasivat, että vaikka äänen laatu oli molemmilla malleilla yhtä korkeaa, AudioLM onnistui kuitenkin tuottamaan jatkuvia melodioita ja selvän rakenteen pitäviä musiikkipätkiä.

Sekä WaveNet että AudioLM ovat siis kehitetty puheen generoimiseen, mutta kykenevät myös generoimaan musiikkia. Audiogeneraatiotekniikoiden kehitykset ovat keskittyneet usein puheen generoimiseen tai tunnistukseen, ja musiikin generoiminen on tullut sivutuotteena. Audiogeneraation kehittyessä alettiin kuitenkin myös kehittämään juuri musiikin generoimiseen tarkoitettuja malleja.

2.3 Musiikkigeneraatio

Musiikin tuottamisen olennaisimmat osat ovat jatkuvuus ja toistaminen. Musiikillisen tuotoksen jälkimmäiset osat seuraavat aikaisempien osien rakentamaa kontekstia. Idean toistaminen kappaleen sisällä vahvistaa kappaleen rakennetta. Toistaminen osoittaa tahtoa, ja halua tuoda sävellyksen tiettyä kohtaa esille. ”Kerran on virhe, kahdesti on jazzia”, on lainaus, joka yhdistetään jazzmuusikko Miles Davisiin. Humoristisella lausahduksella tarkoitetaan, että niin sanotun virheen tahallinen toistaminen voi muuttaa sen tarkoitukselliseksi osaksi musiikkia. Musiikin yksittäisten kohtien tarkoitus ja kuuntelijan korvaan niiden hyvältä kuulostaminen perustuu siis sen osaan kokonaisuudesta.

Aikaisemmat yritykset musiikin automaattiseen generoimiseen ovat olleet helposti tunnistettavissa koneen itse tuottamaksi musiikiksi juuri jatkuvuuden puutteen takia. Malleilla ei ole ollut tarpeeksi pitkää muistia, että tuotettu musiikki pysyisi samassa kontekstissa. Generoitu musiikki hyppi epäselvästi eri kohtiin, jotka eivät sopineet aikaisemmin soineen kohdan kanssa, vaikka yksittäisinä lyhyinä pätkinä tuotokset olisivatkin olleet mieluisan kuuloisia. Tämän muutti transformer-mallien kehitys.

On olemassa kaksi keskeistä musiikin generoimiseen käytettyä mallia: Audiopohjaiset, ja symboliset. On vielä lisäksi malleja, jotka yhdistävät nämä kaksi lähestymistapaa, joita kutsutaan hybridimalleiksi. [22] Symboliset mallit perustuvat musiikillisten symboleiden generoimiseen, joita ihminen tai kone voi soittaa, tuottaakseen

musiikkia. Mallit voivat generoida esimerkiksi nuotistoja soittimille tai kokonaisille orkestereille. MIDI (Musical Instrument Digital Interface) on tiedostotyyppi, joka sisältää tietoa musiikillisista ohjeista kuten sävelkorkeudesta ja temposta. Tietokone pystyy soittamaan musiikkia MIDI tiedoston sisältämien nuottien avulla, kun tiedosto tiputetaan johonkin audiotyöasemaan tai muuhun sopivaan ohjelmaan. [23]

Audiopohjaiset mallit eroavat muista musiikkigeneraation malleista hyvin pienen ihmisvaikutuksen tarpeensa takia. Audiopohjaiset mallit koulutetaan äänitiedostoilla ja ne generoivat uuden äänitiedosten koulutusmateriaalin perusteella. Neuroverkosto siis generoi ääntä äänen avulla. Tarkoituksena on luoda uusi musiikillinen teos koulutusmateriaalin pohjalta. Valmis tuotos muistuttaa siis vahvasti koulutusmateriaalina käytettyjä kappaleita, mutta tuotos on kuitenkin uusi, erilainen sävellys.

Kun tekoälyä koulutetaan tuottamaan musiikkia, on järkevää pysytellä yhden genren sisällä. On helpompaa luoda järkeviä sävellyksiä, kun lähdemateriaalin eroavaisuudet eivät ole liian suuria. Jos malli yhdistää jokaisen musiikin genren sävellyksiä yhdeksi, ei saada järkevää, tai yleensä miellyttävän kuuloista tuotosta.

Uudet syväoppimisen mallit musiikin generoimiseen pystyvät kuitenkin luomaan musiikkia monista genreistä ja musiikin alalajeista. Ihminen kirjoittaa alkuehtoja musiikin generoimiseen, kuten genre, tunnetila tai jokin muu mielikuva. Malli on koulutettu tuhansien sävellysten avulla, ja tekstisyötteet saavat mallin rajaamaan tuotoksen johonkin järkevään. Näissä musiikkigeneraattoreissa pitää yhdistää prosessiin myös tekstinkäsittelymalli, joka ymmärtää tekstisyötteen.

2.4 Musiikkigeneraattorit

Lähivuosina markkinoille on ilmestynyt suuri määrä musiikkingeneroimiseen tarkoitettuja ohjelmia ja sivustoja, kuten Suno AI, Googlen MusicLm, Stability AI:n Stable Audio (open source) ja Beatoven. Näistä kuitenkin ylivoimaisesti suosituin on Suno AI, joka julkaistiin julkiseen käyttöön 20. Joulukuuta 2023. [24], [25]

Sunon nykyistä mallia edelsi GitHubissa Huhtikuussa 2023 julkaistu Bark, joka on transformer pohjainen audiogeneraatiomalli. Barkilla pystyy generoimaan ääniefektjä tai puhetta tekstisyötteen avulla. [26] Suno AI:n taustalla toimii Barkin lisäksi Chirp, joka luo musiikin taustaraidan.

Suno AI on oikeastaan keskustelubotti, joka kehitettiin toimimaan helppona käyttäjäliittymänä Bark:in ja Chirp:in avulla musiikin generoimiseen. Malli lisättiin toimimaan ensin Discord nimisessä viestintäsovelluksessa, ja myöhemmin mallille tehtiin omat nettisivut (suno.com [25]). Sunon nykyisestä toimintamallista ei löydy paljoa julkista tietoa. Suno ei ole myöskään jakanut julkisesti koulutukseen käyttämiään tietokantoja.

Sunolla musiikin generoimisen aloitus tapahtuu antamalla sivustolle tekstisyöte, jonka pohjalta musiikki generoidaan. Syöte voi olla lyriikoita, tai ohjeita esimerkiksi musiikin genrestä, temposta tai tunnetilasta. Syötekenttään voi myös äänittää tai ladata valmiin äänitiedoston, mitä hyödynnetään kappaleen generoimiseen.

Valmis tuotos on usean minuutin mittainen kappale. Näitä kappaleita voi jakaa Sunon omalla alustalla, jossa käyttäjät voivat kuunnella toistensa generoimia tuotoksia. Suno AI pro -tilaaja voi myös ladata kappaleita suoratoistopalveluihin kaupallisessa tarkoituksessa. Tähän tarvitsee ulkoisen musiikinjakelijan, joka pystyy lähettämään kappaleet esimerkiksi Spotifylle tai Apple Musicille.

Tekoälygeneroidusta musiikista saatua kaupallista hyötyä kyseenalaistetaan usein. Kysymykset eettisyydestä ja ekonomisesta vaikutuksesta toisten luomien sävellysten pohjalta koulutetun mallin hyödyntämiseen kaupallisen musiikin tuottamisessa ovat lisääntyneet teknologian kehittyessä.

Suno dominoi AIM (Ai Music) markkinoita, ja on ollut monien kiistojen kohteena. Suno on myös haastettu oikeuteen Yhdysvalloissa useista tekijänoikeusrikkomuksista. Sunon asianaajat ovat argumentoineet oikeudessa, että ohjelma ei riko tekijänoikeuslakeja, koska musiikkia ei samplata, eli osia musiikista ei suoraan uu-

delleenkäytetä uuden musiikin tekemiseen. Heidän mukaansa ei ole väliä kuulostaako Sunon tuotos samanlaiselta, tai edes täsmälleen samalta kuin jokin olemassa oleva teos, koska ohjelma ei perustu samplaamiseen.

Muutkin musiikkigeneraattorit perustuvat hyvin samanlaiseen prosessiin kuin Suno AI. Musiikin generoiminen on jokaiselle siis hyvin helppoa ja nopeaa, eikä välttämättä tarvitse minkään laista aikaisempaa kokemusta musiikin tuottamisesta.

2025 lopussa pidetyssä sisäisessä sijoittajien esityksessä kerrottiin, että Suno AI sivustolla tuotettiin noin 7 miljoonaa kappaletta päivittäin. Sivusto siis tuotti esimerkiksi Spotifyn valtavan yli 100 miljoonan kappaleen kirjaston määrän uusia kappaleita aina noin kahden viikon välein. Tekoälygeneroitua musiikkia on nykyään jo melko yleistä nähdä useilla top-musiikin listoilla kuunnelluimpien kappaleiden joukossa.

3 Kulttuuriset vaikutukset

Audiopohjainen musiikkigeneraatio herättää paljon keskustelua, ja näkee enemmän käyttöä ja kuuntelua päivä päivältä. Valmiin äänitiedoston generoiminen pelkkien sanojen pohjalta lisää näkyvyyttä matalan kynnyksensä ansiosta. Tekoälymusiikkia ja muuta taidetta hyödynnetään paljon meemeissä, ja muussa lyhytkestoisessa sisällössä, jonka elinikä ei kuulukkaan olla pitkä. Tekoälymusiikkia ei nähdä usein aitona, tai tosissaan otettavana musiikkina. Tämä voi kuitenkin olla muuttumassa.

3.1 Tekijänoikeuslait

Tekijänoikeuslait ovat jääneet tekoälyn vahudikkaan kehityksen jälkeen. Tekijänoikeuslait on säädelty suojaamaan nimenomaan ihmisten tuottamaa taidetta, ja myös motivoimaan ihmisiä tekemään taidetta luvatonta uudelleenkäyttöä estämällä. Vielä ei ole päästy yhteisymmärrykseen tekoälytuotoksien originaalisuudesta, tai siitä, miten määritellään mihin ihmisen vaikutus loppuu ja mistä täysin tietokoneen tuottama tuloste alkaa. Yleinen ratkaisu ongelmaan on, että kaikki tietokoneella automatisoitu tuottaminen kuuluu eri kategoriaan kuin ihmisen tuottama sisältö. Tekoälygeneraatioon liittyvät tekijänoikeuslait vaihtelevat paljon eri valtioiden välillä, mikä nykyisessä globalisoituneessa yhteiskunnassa hankaloittaa artistien rajojen tuntemusta, ja omaan taiteeseen kohdistuneiden vääryyksien haastamista.

Keskeisimmät kysymykset tekijänoikeuksista ja tekoälytaiteen eettisyydestä voidaan jakaa kahteen aiheeseen: mallin kouluttamiseen tarvittavan datan keräämiseen

ja käyttöön, ja mallin tuotoksen samankaltaisuuksiin aikaisempiin teoksiin ja sävellyksiin. Neuroverkon kouluttamiseen tarvitaan valtavat määrät dataa, joka musiikin generoimisessa tarkoittaa tuhansien tai jopa miljoonien sävellyksien haalimista koulutuskäyttöön. Voidaankin kysyä; Mistä nämä kappaleet saadaan kerättyä, ja kuuluisiko alkuperäisille artisteille jakaa korvauksia heidän sävellyksien käyttämisestä? Tekijänoikeuslait eivät estä ohjelmoijia käyttämästä musiikkia mallin kouluttamiseen, mutta luvaton kappaleiden lataaminen ja jakaminen on silti laitonta. [27]

3.2 Jakautuvat mielipiteet

Musiikin generoiminen jakaa mielipiteitä. Jotkut näkevät sen mahdollisuutena kaikille tehdä musiikkia, riippumatta omista yhteyksistä tai saatavilla olevista työkaluista. Enään ei tarvitsisi suuria levy-yhtiöitä resursseihin ja musiikin julkaisemiseen. Toiset taas uskoo tekoälymusiikin tappavan musiikkialan ja ekonomian. Jos kaikki pystyvät tulevaisuudessa tuottamaan musiikkia, jota suurin osa ei erottaisi ihmisen tuotoksesta, pienenevätkö mahdollisuudet pienille artisteille päästä sisälle alaan ja tehdä elantoa?

Suurin osa tekoälymusiikin kannattajista eivät kuitenkaan ajattele ekonomian tai eettisyyden kannalta, vaan luomisen helppouden ja hauskuuden. Yksi yleinen mielipide on, että tekoäly on vain työkalu, jolla tehdä musiikkia. Sama kuin instrumentin, tai tietokoneen muiden ohjelmien käyttäminen musiikin tuottamiseen. Tekijän artistisuus voidaan joidenkin mielestä nähdä siinä, miten kirjoitetaan taitavasti ohjaavia kehoitteita mallin tekstikenttään.

Tekoälymusiikkia ympäröivää diskurssia voidaan verrata aikaisempiin musiikkiteknologian edistymisen herättämiin kiistoihin. Voidaan palata äänitetyn musiikin alkuaikoihin 1930-luvulle, jolloin äänitetty musiikki nähtiin niin mielenkiintoisena kehityksenä, kuin sydämettömänä ja tunteettomana verrattuna livemusiikkiin.

Monille tuttu ja musiikkialalla nykyään itsestäänselvyydeksi lujittunut AutoTune, oli alkuhetkillään hyvin kiistanalainen uusi teknologia. AutoTunella pystyy parantamaan lauluesityksen tarkkuutta antamalla ohjelman liikkuttaa nuotin vierestä laulettuja osia lähemmäksi, tai tarkalleen oikealle nuotille. Nykyään AutoTune hyväksytään laajemmin osana musiikin tuottoprosessia, varsinkin ammattitasolla, ja on usein oletettua, että valtaosa populaarimusiikista käyttää tätä teknologiaa. Vertauskuva AI-musiikin ja AutoTunen välillä on hyvin yleinen argumentti AI-musiikin puolesta.

Tekoälymusiikkia pidetään usein kuitenkin perusteellisesti erinlaisena kehityksenä muista musiikkiteknologian kehityksistä. Argumenttina on luovan artistin kontrollin menettäminen ja tarkoituksenmukaisuuden puuttuminen sävellyksestä. Eli uudet työkalut ovat aina vaatineet käyttäjältä jonkinlaista osaamista ja aktiivista tekemistä, joka vähenee huomattavasti algorimisesti generoidussa musiikissa. AI-musiikki on siis nykyään kiistanalainen asia, mutta teknologian uutuus voi myös vääristää ihmisten suhtautumista siihen.

4 Pohdintaa

AI-musiikin tämän hetkinen populaarisuus johtuu suureksi osaksi teknologian tuoreudesta. Uusi teknologia on jännittävää ja monille mielenkiintoista, ja toisaalta joillekin pelottavaa. Audiopohjainen musiikin generoiminen on suuri saavutus tekoäly teknologiassa, mutta jonka mahdollisuus tuo esiin monia ongelmia. Artistin näkökulmasta musiikkigeneraatiomallin koulutus sisällöllä ilman artistin suostumusta voidaan nähdä varastamisena. Artistien elantomahdollisuudet voivat pienentyä tekoälymusiikin yleistymisen seurauksena, joka vaikuttaisi myös musiikin kuuntelijoihin. Kuuntelijan suosikki artisti voi rahoituksen puutteen vuoksi joutua vähentämään musiikin tuottamista.

Yksi ratkaisu artistien ja kuuntelijoiden ongelmiin voi olla tekoälyllä generoidun musiikin ja ihmisten tuottaman musiikin jakaminen eri kategorioihin. Deezer on suoratoistopalvelu, joka otti kesällä 2025 käyttöön AI-leiman, jolla merkataan kaikki tekoälygeneroidut tuotokset sivustolla. Deezerin mukaan sivuston kaikista toistoista noin 0.5 % kuuluu tekoälygeneroidulle sisällölle. He huomasivat myös, että noin 70% AI-musiikin toistoista on petollisia, eli on käytetty jotain toistoja manipuloivaa taktiikkaa rahallisen voiton saamiseksi.[28]

Timbaland on yksi 2000-luvun tunnetuimpia ja vaikutusvaltaisimpia tuottajia, ja nykyään tunnetaan suurena AI-musiikin kannattajana ja edistäjänä. Timbaland on osa artistien vähemmistöä, jotka tekoälyn kieltämisen tai sivuuttamisen sijaan

ovat hyväksyneet tekoälyn osana taiteen luomisen tulevaisuutta. Ajatuksena on, että tekoäly on, ja tulee olemaan yhä olennaisempi osa luovaa prosessia.

Musiikki on taidetta ja taiteen määrittäminen on hankalaa. Ikään kuin vastapainona, AI-taiteen yleistyminen on lisännyt musiikin kuuntelijoiden tai kuvataiteen tarkastelijoiden kiinnostusta taiteen luomisprosessiin. Luomisprosessi on olennainen osa taideteosta, ja voidaan jopa väittää, että prosessi itsessään tekee lopputuloksesta taidetta.

Jacques Attali, ranskalainen ekonomisti ja tutkija, väittää kirjassaan "Noise: The Political Economy of Music", että musiikki, tai melu, on ikään kuin kulttuurin ennen näkijä, joka ensimmäisenä näyttää suuret muutokset kulttuurissa. "Muutos on kaiverrettu melussa nopeammin kuin mitä se muuttaa yhteiskuntaa." [29] Attali julkaisi kirjan vuonna 1977, mutta sen esittämät ideat pitävät paikkansa vielä nykypäivänä. Attalin väitteen ideana on, että yhteiskunnan kehityksen, kuten myös teknologian kehitykset, näkyvät ensimmäisenä ajan taiteessa, jo ennen kuin monet huomaavatkaan suuren muutoksen tapahtuneen.

5 Yhteenveto

Musiikkia on tuotettu tietokoneilla jo tietokoneiden kehityksestä lähtien. Tietokoneen tuottama algoritminen musiikki oli mahdollista jo 1950-luvulla. Audiopohjaisen musiikin generoimisen on mahdollistanut tekoälygeneraatiossa, ja tarkemmin audiogeneraatiossa tapahtuneet suuret loikkaukset viimeisen kymmenen vuoden aikana. Teknologian kehittyessä musiikin generoimisesta on tullut entistä helpompaa, nopeampaa sekä laadukkaampaa. Puheentunnistus- ja puheengeneroimisteknologian kehityksen myötä kehittyi myös musiikin generoiminen. Mallit, jotka kehitettiin puhetta varten, toimivat myös vaihtelevalla menestyksellä musiikin generoimiseen.

Aikaisemmat yritykset musiikin generoimisessa kompastuivat tekoälymallin liian lyhyeen muistiin. Musiikkingeneroimisen kannalta on tärkeää pystyä säilyttämään konteksti pitkän aikaa, jotta minuuttien pituiset kokonaiset kappaleet kuulostavat ihmisten tottumuksiin nähden järkevältä ja mielisalalta musiikilta.

Transformer-mallin kehitys vuonna 2017 oli valtava loikkaus tekoälyteknologiasa. Transformer-malli kehitettiin kielen kääntämiseen ja tunnistamiseen, mutta korjasi samalla musiikin generoimisen ongelmia. Mallin toimintaperiaate ja pidempi muisti mahdollistaa nykyisten laajimmin käytettyjen musiikkigeneraattorien, kuten Suno AI:n, toiminnan.

Tekoälymusiikin pystyy monissa tapauksissa nykypäivänä vielä erottamaan ihmisen itse tekemästä musiikista. Ero ihmisen ja tekoälyn tekemässä musiikin kuitenkin pienenee teknologian kehittyessä entisestään.

Tekijänoikeudet, jotka suojaavat artistien taiteen uudelleenkäyttöä, ovat jääneet tekoälymusiikin nopean kehityksen jälkeen. Lakeja kuitenkin päivitetään jatkuvasti, ja uusia ratkaisuja yritetään kehittää artistien oikeuksien suojelemiseksi.

Tekoälymusiikin kehitys mahdollistaa musiikin helpomman tuottamisen, joka lisää päivittäin tuotetun musiikin määrää hyvin paljon. Tämä vaikuttaa negatiivisesti pienempien artistien mahdollisuuksiin tulla kuulluksi, ja saada korvausta tekemästään musiikista.

Aihe jakaa mielipiteitä. Jotkut näkevät tekoälygeneroidun musiikin kehityksen positiivisena asiana, joka lisää ihmisten mahdollisuutta luoda musiikkia. Toiset taas näkevät kehityksen trivialisoivan musiikkia, ja vähentävän muusikoiden uramahdollisuuksia.

Lähdeluettelo

- [1] M. Clancy, *Artificial Intelligence and Music Ecosystem*. Routledge, 2023.
- [2] C. Brower, ”A Cognitive Theory of Musical Meaning”, *Journal of Music Theory*, vol. 44, nro 2, s. 323–379, 2000, ISSN: 00222909. viitattu 5. tammikuuta 2026. url: <http://www.jstor.org/stable/3090681>.
- [3] J. Fauvel, R. Flood ja R. J. Wilson, *Music and mathematics: From Pythagoras to fractals*. Oxford University Press, 2006.
- [4] H. R. Strayer, ”From neumes to notes: The evolution of music notation”, *Musical Offerings*, vol. 4, nro 1, s. 1, 2013.
- [5] S. A. Hedges, *Dice Music in the Eighteenth Century*. Oxford University Press, 1978.
- [6] T. FUNK, ”A Musical Suite Composed by an Electronic Brain: Reexamining the Illiac Suite and the Legacy of Lejaren A. Hiller Jr”, eng, *Leonardo music journal*, vol. 28, s. 19–24, 2018, ISSN: 0961-1215.
- [7] A. Mayor, ”Gods and robots: Myths, machines, and ancient dreams of technology”, 2018.
- [8] S. Zhu et al., ”Intelligent Computing: The Latest Advances, Challenges, and Future”, *Intelligent Computing*, 2023.
- [9] Y. LeCun, Y. Bengio ja G. Hinton, ”Deep learning”, *nature*, vol. 521, nro 7553, s. 436–444, 2015.

-
- [10] Z. Epstein ja A. Hertzmann, ”Art and the science of generative AI”, *Science*, 2023.
- [11] F. Rosenblatt, ”The perceptron: a probabilistic model for information storage and organization in the brain.”, *Psychological review*, vol. 65, nro 6, s. 386, 1958.
- [12] J. Ho, A. Jain ja P. Abbeel, ”Denoising diffusion probabilistic models”, *Advances in neural information processing systems*, vol. 33, s. 6840–6851, 2020.
- [13] Q. Huang et al., ”Noise2music: Text-conditioned music generation with diffusion models”, *arXiv preprint arXiv:2302.03917*, 2023.
- [14] A. van den Oord et al., ”WAVENET: A GENERATIVE MODEL FOR RAW AUDIO”, Google DeepMind, London, UK, tekninen raportti, 2016.
- [15] F. Colombo, S. P. Muscinelli, A. Seeholzer, J. Brea ja W. Gerstner, ”Algorithmic composition of melodies with deep recurrent neural networks”, *arXiv preprint arXiv:1606.07251*, 2016.
- [16] S. Hochreiter ja J. Schmidhuber, ”Long Short-Term Memory”, *Neural Computation*, vol. 9, nro 8, s. 1735–1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.
- [17] A. Vaswani et al., ”Attention Is All You Need”, Google, tekninen raportti, 2017.
- [18] C.-Z. A. Huang et al., ”Music transformer”, *arXiv preprint arXiv:1809.04281*, 2018.
- [19] R. Bommasani, ”On the opportunities and risks of foundation models”, *arXiv preprint arXiv:2108.07258*, 2021.
- [20] A. van den Oord ja T. Walters, ”WaveNet launches in the Google Assistant”, Google DeepMind, tekninen raportti, 2017.

-
- [21] Z. Borsos et al., "Audiolm: a language modeling approach to audio generation", *IEEE/ACM transactions on audio, speech, and language processing*, vol. 31, s. 2523–2533, 2023.
- [22] J.-P. Briot, G. Hadjeres ja F.-D. Pachet, "Deep Learning Techniques for Music Generation – A Survey", Sorbonne Université, tekninen raportti, 2019.
- [23] G. Mittal, J. Engel, C. Hawthorne ja I. Simon, "Symbolic music generation with diffusion models", *arXiv preprint arXiv:2103.16091*, 2021.
- [24] Y. Y. T. Nugroho ja P. P. M. D. Manggala, "The use of AI in creating music compositions: A case study on Suno application", teoksessa *7th Celt International Conference (CIC 2024)*, Atlantis Press, 2024, s. 177–189.
- [25] "Suno", Suno, Inc., viitattu 12. maaliskuuta 2026. url: <https://suno.com>.
- [26] G. Kuczko. "Bark", viitattu 12. maaliskuuta 2026. url: <https://github.com/suno-ai/bark>.
- [27] S. Chesterman, "Good models borrow, great models steal", *Policy and Society*, 44(1), 23–37, 2025.
- [28] Deezer, "Deezer launches world's first AI tagging system for music streaming", *Deezer Newsroom*, 2025.
- [29] J. Attali, *Noise: The political economy of music*. Manchester University Press, 1985, vol. 16.