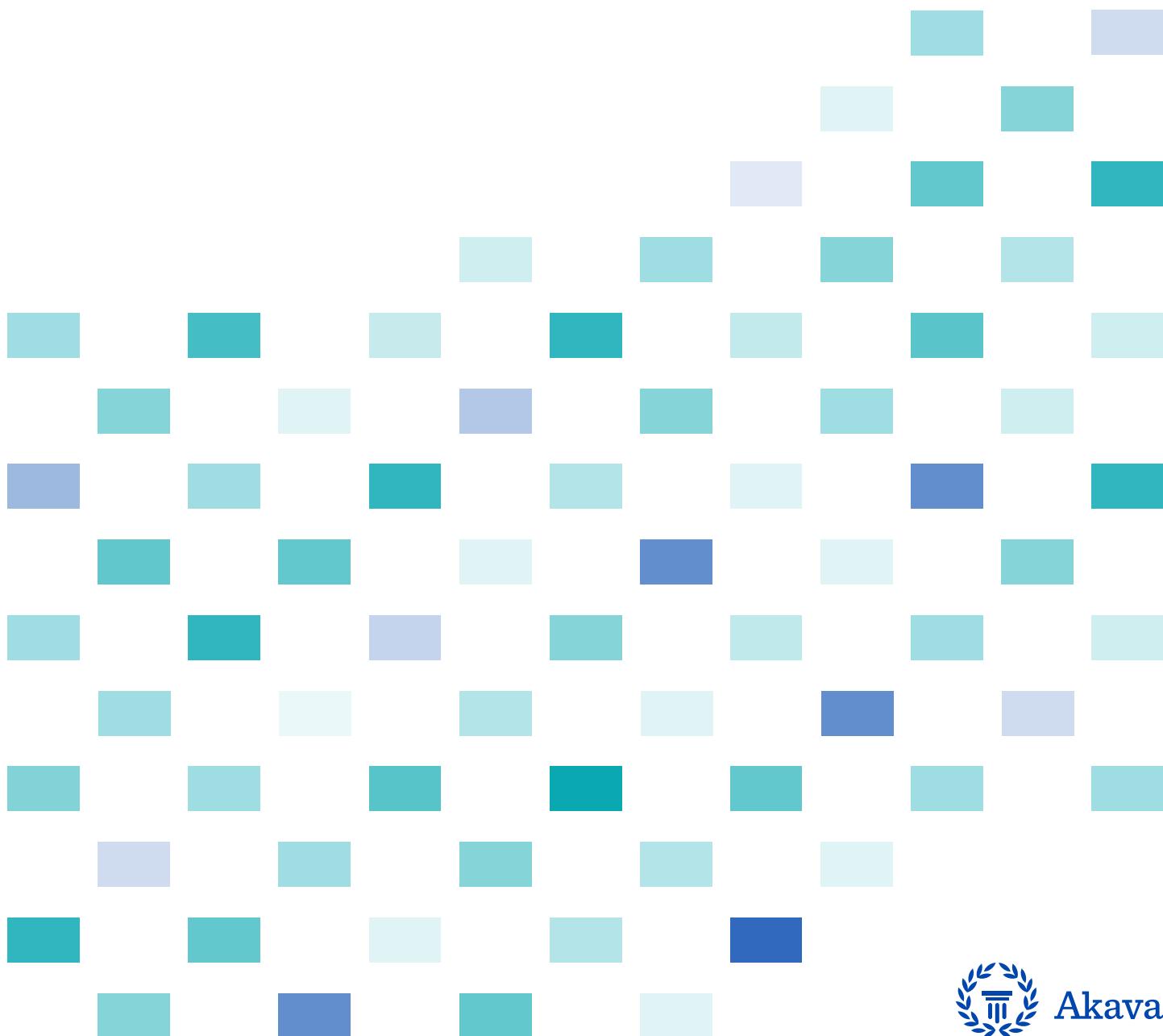


# **Tekijänoikeudella suojattu aineisto tekoälyn koulutuksessa**

Artikkeli 2/2024



## Artikkeli 2/2024

Kirjoittaja: Riku Neuvonen, julkisoikeuden yliopistonlehtori, viestintäoikeuden dosentti

Avainsanat: tekoäly, tietosuojaja

Päivämäärä: 26.2.2024

### Kirjoittajan esittely

Riku Neuvonen on julkisoikeuden yliopistonlehtori ja viestintäoikeuden dosentti.

### Yhteenveto

Artikkelissa käsitellään tekijänoikeuden roolia tekoälyn koulutuksessa, jossa hyödynnetään tekijänoikeudella suojattua sisältöä. Samalla tarkastellaan tekijänoikeuteen liittyviä rajoituksia ja säädöksiä. Artikkelin päättyy pohdintaan siitä, voidaanko tekoälyä pitää teoksen luojana ja täyttääkö tekoälyn luoma teos teoskynnyksen vaatiman omaperäisyyden kriteerin.

### Akava Works

Akavan tiedontuotanto toimii nimellä Akava Works.

Akava Works tarjoaa monipuolisesti tietosisältöjä raporteina, selvityksinä ja tutkimuksina. Tavoitteenamme on lisäksi herättää yhteiskunnallista keskustelua erityisesti akavalaisille tärkeistä ja ajankohtaisista aiheista. Akava Works -sisällöt eivät ole Akavan virallisia kannanottoja.

Akava Works -sivusto on osoitteessa [www.akavaworks.fi](http://www.akavaworks.fi).

## Sisällysluettelo

1	Tekijänoikeus tekoälyn koulutuksen haasteena.....	3
2	Tekijänoikeuden rajoitukset mahdollisuutena .....	4
3	Tekoäly tekijänä .....	5

# Johdanto

Tällä hetkellä tekoäly on vallitseva ilmiö monilla aloilla. Samalla siihen liittyy myös monia pelkoja ja uhkakuvia. Erityisesti luovilla aloilla sen pelätään ja uskotaan korvaavan luovan työn tekijät. Keskustelusta usein kuitenkin unohtuu, että tekoälyä koulutetaan olemassa olevan aineiston perusteella. Onko tämä tekijänoikeuden kannalta sallittua? Entä voiko tekoäly olla tekijänoikeuden suojaan oikeutettu tekijä?

Tekoällyn läpimurto julkisuudessa tapahtui kunnolla viimeistään vuonna 2023. Vuoden aikana OpenAI:n kehittämä, suuriin kielimalleihin (GPT) perustuva tekoälysovellus ChatGPT lunasti monet tekoälylle asetetut odotukset. Se pystyi tuottamaan tekstiä annettujen ohjeiden perusteella ja vastaamaan kysymyksiin uskottavammin kuin aiemmat tekoälyt. Perustana tälle on tekoällyn opetukseen käytetyt suuret tekstimassat. Samalla, kun tekoällyn käyttö lisääntyi, käynnistyivät spekulatiot sen vaikutuksista ja tulevaisuudesta.

Tekoäly on nähty uhkana erityisesti luovilla aloilla. Uhkakuvana on, että tekoäly pystyy tuottamaan sille annettujen ohjeiden mukaisesti kaikenlaisia sisältöjä, jolloin tarve luovan työn tekijöille vähenee. Tässä keskustelussa sivurooliin on jäänyt se, että useimpien tekoällyjen opetuksessa käytetään luovilla aloilla luotuja teoksia. Näillä teoksilla on tekijä ja teoksiin kohdistuu tekijänoikeus. Voiko tekoäly siis käyttää esimerkiksi tätä kirjoitusta omaan oppimiseensa ja omien vastaustensa pohjana? Onko näillä tekoällyn luomilla aineistolla oma tekijänoikeus?

OpenAI on tämän hetken puhutuin tekoällyn kehittäjä, sillä se on julkaissut myös kuvia tuottavan DALL-E ohjelman, josta on nyt käytössä jo toinen kehitysversio. Vastaavia ovat Midjourney ja Stable Diffusion. Näiden läpimurtosovellusten vanavedessä on tullut suuri joukko muitakin suuriin kielimalleihin ja kuvapankkeihin perustuvia tekoälysovelluksia, jotka voivat tuottaa tekstiä ja kuvia käyttäjän esittämien toiveiden mukaisesti. Tällaisiin tekoälysovelluksiin liittyy useita kysymyksiä, joita esimerkiksi EU:n tekoälysäätös (AI Act) pyrkii tulevaisuudessa sääntelemään.

## 1 Tekijänoikeus tekoällyn koulutuksen haasteena

Tekoälyalalla on runsaasti keskustelua tekoällyjen kehittämisen etiikasta. Yksi keskeinen kysymys on tekoällyjen opettaminen materiaalilla, joka on muiden tekemää. Esimerkiksi ChatGPT:lle on annettu aineistona käytännössä kaikki internetissä oleva tieto aluksi vuoteen 2021 ja nyt joiltakin osin vuoteen 2023 saakka. Tämä aineisto on lähes kokonaan sellaista, jolla on tekijä ja johon kohdistuu tekijänoikeuden suoja. Olennainen kysymys onkin, että salliiko nykyinen tekijänoikeuden sääntely tekijänoikeuden alaisen aineiston käytön tällaisessa tarkoituksessa vai onko kyse tekijänoikeuden loukkauksesta (esim. Kousa 2013).

Tekijänoikeuslähtöisessä tekoälykeskustelussa ja -tutkimuksessa lähtökohta oli pitkään pohtia, voiko tekoällyn tuottama tai avustama teos saada tekijänoikeuden suojan ja voiko tekoäly olla tekijä. Tekijänoikeudessa peruslähtökohtana on, että teoksen tulee ylittää niin sanottu teoskynnys eli osoittaa, että se on riittävän itsenäinen ja omaperäinen. Teokseksi voidaan katsoa mikä tahansa luovan työn tulos. Suomessa tekijälle syntyy

automaattisesti tekijänoikeus teokseen, joten tekijänoikeutta ei tarvitse rekisteröidä tai merkitä symboleilla. Tekijänoikeuksiin luetaan myös niin sanotut lähioikeudet, mutta niiden ottaminen tähän tarkasteluun sekoittaisi kokonaisuutta. (Harenko ym. 2016.)

Tekijänoikeus antaa tekijälle yksinoikeuden päättää teostensa käytöstä. Muilla ei ole oikeutta käyttää teosta ilman tekijän lupaa. Tähän on muutamia poikkeuksia, joihin palataan oikeassa asiayhteydessä. Tekijänoikeuteen kuuluvat taloudelliset oikeudet eli oikeus valmistaa teoksesta kappaleita ja oikeus saattaa teos yleisö saataville. Nämä oikeudet on mahdollista luovuttaa osittain tai kokonaan. Tekijänoikeuteen kuuluvat myös moraaliset oikeudet eli oikeus tulla tunnustetuksi teoksen tekijänä ja oikeus kieltää teoksen omaperäisyyttä tai taiteellista arvoa loukkaava muuntelu. Moraalisia oikeuksia ei ole mahdollista luovuttaa. (Harenko ym. 2016.)

Tekijänoikeus on tällä hetkellä voimassa tekijän koko elämän ajan ja 70 vuotta hänen kuolemansa jälkeen. Huomattava on, että tekijänoikeus ei koske ideaa, vaan muotoa. Vielä jokin aikaa sitten tekijänoikeuskeskustelussa oli osittaisena vitsinä, että tekijänoikeuden suoja-aikaa pidennetään aina kun Mikki Hiiren tekijänoikeus on vanhenemassa. Nyt alkuvuodesta 2024 ensimmäinen Mikki Hiiren muoto menetti suojansa, joten suoja-aikoihin on tuskin tulossa muutoksia.

Keskeistä tekijänoikeudessa on, että riippumatta kansallisesta lainsäädännöstä tekijänoikeus suojaa tekijän asemaa ja tekijänoikeuden haltijan oikeuksia kaikkia kohtaan. Tekijänoikeus on yksinoikeus päättää tietyn teoksen käytöstä, johon sisältyy oikeus saada korvausta teoksen käytöstä tekijän näin halutessa. Tekoällyn koulutuksessa käytetään todennäköisesti tekijänoikeudella suojattua materiaalia ilman tekijänoikeuden haltijan lupaa.

## 2 Tekijänoikeuden rajoitukset mahdollisuutena

Tekijänoikeuteen on olemassa erilaisia rajoituksia, jotka sallivat teoksen esittämisen ja kappaleiden valmistamisen eli kopioinnin tietyissä olosuhteissa. Pääosin nämä koskevat opetuskäyttöä, tutkimuskäyttöä ja lainaamista sallitulla tavalla. Sallitussa lainaamisessa eli sitaatissa on mainittava aina alkuperäinen teos ja tekijä. Tämän ohella Yhdysvalloissa on tekijänoikeudessa fair use -oppi, jonka mukaan tekijänoikeudella suojattua aineistoa voi käyttää tietyissä tilanteissa. Tämän opin mukainen käyttöala on paljon laajempi kuin eurooppalaiset erikseen mainittuihin ja rajattuihin tarkoituksiin soveltuvat tarkkarajaiset poikkeukset (Tapio 2013). Suurin osa tunnettujen tekoälysovellusten kehittäjistä on yhdysvaltalaisia, jolloin tapauksiin sovellettavan lainsäädännön valinta voi olla haasteellista ja aiheuttaa ongelmia.

Suomen tekijänoikeussäännöstö perustuu nykyään suurelta osin EU:n tekijänoikeussäännöksiin ja vastaa mannereurooppalaista tekijänoikeuskäsitystä. Tähän sääntelyn kuuluvat väliaikaiset teoskappaleet, joita saa tekijänoikeuslain 11 a § mukaan valmistaa tarkoitukseen, jolla ei ole itsenäistä taloudellista merkitystä ja joka on väliaikaista sekä välttämätön osa teknistä prosessia. Tämä soveltuu käytännössä välimuistiin tehtäviin teoskappaleisiin. Tekoällyn opetuskäyttö saattaisi liittyä tähän poikkeukseen, mutta itsenäisen taloudellisen merkityksen arviointi voi muodostua ongelmaksi. Toisaalta kaikissa tekoälymalleissa toiminta ei ole väliaikaista, vaan teos on tekoällyn käytössä pidempään ja tekoäly ottaa teoksesta osia omaan toimintaansa.

EU:n DSM-direktiivin (digitaalisten sisämarkkinoiden direktiivi) seurauksena Suomen tekijänoikeuslain 13 b §:ään (HE 43/2022 vp ja HE 313/2022 vp) on lisätty mahdollisuus valmistaa teoskappaleita tekstin- ja tiedonlouhintaa varten. Edellytyksenä on, että teokseen on laillinen pääsy eli se on tekijänoikeuden haltijan luvalla saatavilla. Tekijä voi kieltää nimenomaisesti tämän käyttötarkoituksen. Tekoälyä ei direktiivissä mainita, joten säännöksen soveltuvuus nimenomaan tekoälyn tapauksessa on avoinna.

Direktiivejä saatetaan voimaan jäsenvaltioissa kansallisesti, joten Euroopan tasolla voi olla eroja tiedonlouhinnan sääntelyssä. Toisaalta tässäkin tarkastelussa on kyse tekoälykohtaisuudesta eli ovatko kaikki tekoälyt ja niiden opetustoiminta direktiivissä ja laissa mainittua tiedonlouhintaa. Toisaalta verkossa on aineistoa ajalta ennen direktiiviä, jolloin tekijällä ei ole ollut tietoa tällaisesta toiminnasta ja mahdollisuutta kieltää sitä. Tällä hetkellä käytössä olevat tekoälyt on koulutettu aineistolla, jossa tekijöiden lupia ei ilmeisesti ole kysytty. Vastaavalla tavalla esimerkiksi Google on pitänyt pitkään omaa Google Books-palveluaan tarjolla, vaikka sen sisältämien kirjojen tekijänoikeus ei ole ollut selvitetty. Yhdysvalloissa alioikeustasolla käyttö on toistaiseksi mennyt fair use-kategoriaan. Tekoäly näin ollen on vain osa internetissä olevaa tekijänoikeudellisen aineiston harmaalla alueella liikkuvaa hyödyntämistä.

Tekoälyn koulutusaineiston sallittuun käyttöön voidaan käyttää myös lisenssimallia. Suomessa ja monissa muissa maissa sopimuslisensoijärjestelmän avulla on saatu paljon aineistoja lisensoidun käytön piiriin. Tekijänoikeusjärjestöt kuten Teosto, Kopiosto ja Sanasto edustavat suurta joukkoa tekijöitä, joiden puolesta he ovat luoneet lisensointisopimusmallin ja hoitavat lisensoinnin tekijän puolesta. Tekoälyn käyttöön on jo luotu tietokantoja, joiden sisältöä voi lisensoida tekoälyn opetukseen. Samoin OpenAI on tehnyt lisenssisopimuksia kuvapankkien kanssa.

### 3 Tekoäly tekijänä

Tekoäly tekijänä on oma kysymyksensä (Alen ym. 2018). Pääsääntöisesti tekijänoikeus suojaa luonnollisen tekijän eli ihmisen riittävän omaperäisiä teoksia. Ensimmäinen ongelma nykyisissä tekoälyissä koskee tekijän määrittämistä. Onko tekijä tekoäly vai teoksen tekoälyltä ohjeistuksen avulla hankkinut henkilö? Toinen ongelma on omaperäisyyskynnys. Monenlaisia teoksia tehdään yhdistelemällä vaikutteita ja jopa muita teoksia, mutta tekoäly selkeästi käyttää muita teoksia pohjana, ja näin tulisi selvittää, onko tekoälyn mahdollista lisätä luomaansa teokseen sellaista uutta, joka ylittäisi teoskynnyksen. Eri teosten käyttö voi johtaa myös tilanteeseen, jossa tekijänoikeus kuuluu usealla tekijälle (Mattila 2022).

Ajatus tekoälystä yhteistyönä luodun teoksen tekijänä voi soveltua myös tilanteisiin, joissa tekoäly käyttää opetuksessaan käytettyä aineistoa uuden teoksen luomisessa tavalla, jossa aiempien teosten käyttö voidaan osoittaa ja tekijänoikeus näin jakaa. Pohdinta osoittaa sen, että tällä hetkellä ei ole varmuutta oikeudellisesta perusteesta, jolla tekoäly voi käyttää tekijänoikeudella suojattuja aineistoja. Toisaalta ei ole yksiselitteisiä säännöksiä, joilla käyttö voidaan kieltää. EU:n ja sitä kautta Suomen uusi tiedonlouhintasääntely mahdollistaa tässä säännöksessä tarkoitettun käytön kieltämisen, mutta vielä ei ole tulkintaa siitä, missä määrin tekoälyn kouluttaminen ja uusien teosten luominen vastaa tätä tekijänoikeuden poikkeusta. Samalla tavoin avoimena on kysymys tekoälyn luoman teoksen tekijänoikeudesta.

Joulukuussa 2023 muun muassa New York Times on haastanut OpenAI:n ja Microsoftin oikeuteen. Perustana on, että ChatGPT:n koulutuksessa on käytetty miljoonia New York Timesin artikkeleita. Ratkaisun saamisessa kestää aikansa ja se soveltuu suoraan vain yhdysvaltalaiseen tekijänoikeussäätelyyn. Samalla tavoin aiemmin syksyllä 2023 ryhmä taiteilijoita nosti kanteen kuvia hyödyntäviä tekoälysovelluksia vastaan. EU:n piirissä on suunnitteilla uusia säännöksiä, ja vasta uudistetun tekijänoikeussäätelyn tiedonlouhintaa koskevia säännöksiä ei ole vielä testattu oikeudessa. Huolimatta suhteellisen pitkästä historiasta ja teoreettisesta keskustelusta, on tekoälyn ja tekijänoikeuden välisessä suhteessa vielä paljon avoimia kysymyksiä. Tilanne ei ole ratkeamassa pitkään aikaan ja siihen saakka voidaan olettaa, että kaikki internetissä oleva aineisto on tekoälyn koulutuksen materiaalia. Mukaan lukien tämä kirjoitus.

## Lähteet

Alen, Anette & Ballardini, Rosa & Pihlajarinne, Taina: Tekoälyn tuotokset ja omaperäisyysvaatimus - kohti koneorientoitunutta tekijänoikeutta?, *Lakimies* 7–8/2018, s. 975–995.

Hallituksen esitys eduskunnalle laeiksi tekijänoikeuslain ja sähköisen viestinnän palveluista annetun lain muuttamisesta HE 43/2022 vp.

Hallituksen esitys eduskunnalle laeiksi tekijänoikeuslain ja sähköisen viestinnän palveluista annetun lain muuttamisesta annetun hallituksen esityksen (HE 43/2022 vp) täydentämisestä. HE 313/2022 vp.

Harenko, Kristiina & Niiranen, Valtteri & Tarkela, Pekka: *Tekijänoikeus. 2. uudistettu painos*. Helsinki: Talentum Pro 2016 (3. painos julkaistaan keväällä 2024).

Kousa, Viivi: Tekoälytaide ja tekijänoikeus – haastava yhtälö?, *IPR Info* 1/2023 (27.1.2023), <https://iprinfo.fi/artikkeli/tekoalytaide-ja-tekijanoikeus-haastava-yhtalo/> (käyty 14.12.2023).

Mattila, Tuomas: Yhteistyö tekijänoikeudessa: tutkimus alkuperäisestä tekijänoikeuden haltijasta yhteistyöhön ja yhteisöllisyyteen perustuvissa luovissa prosesseissa. *Suomalainen lakimiesyhdistys 2022*.

Tapio, Veli-Markus: Fair Use ja kolmivaihetesti joustavamman tekijänoikeudellisen sääntelyn mahdollistajina. *Lakimies* 1/2013, s. 35–54.