

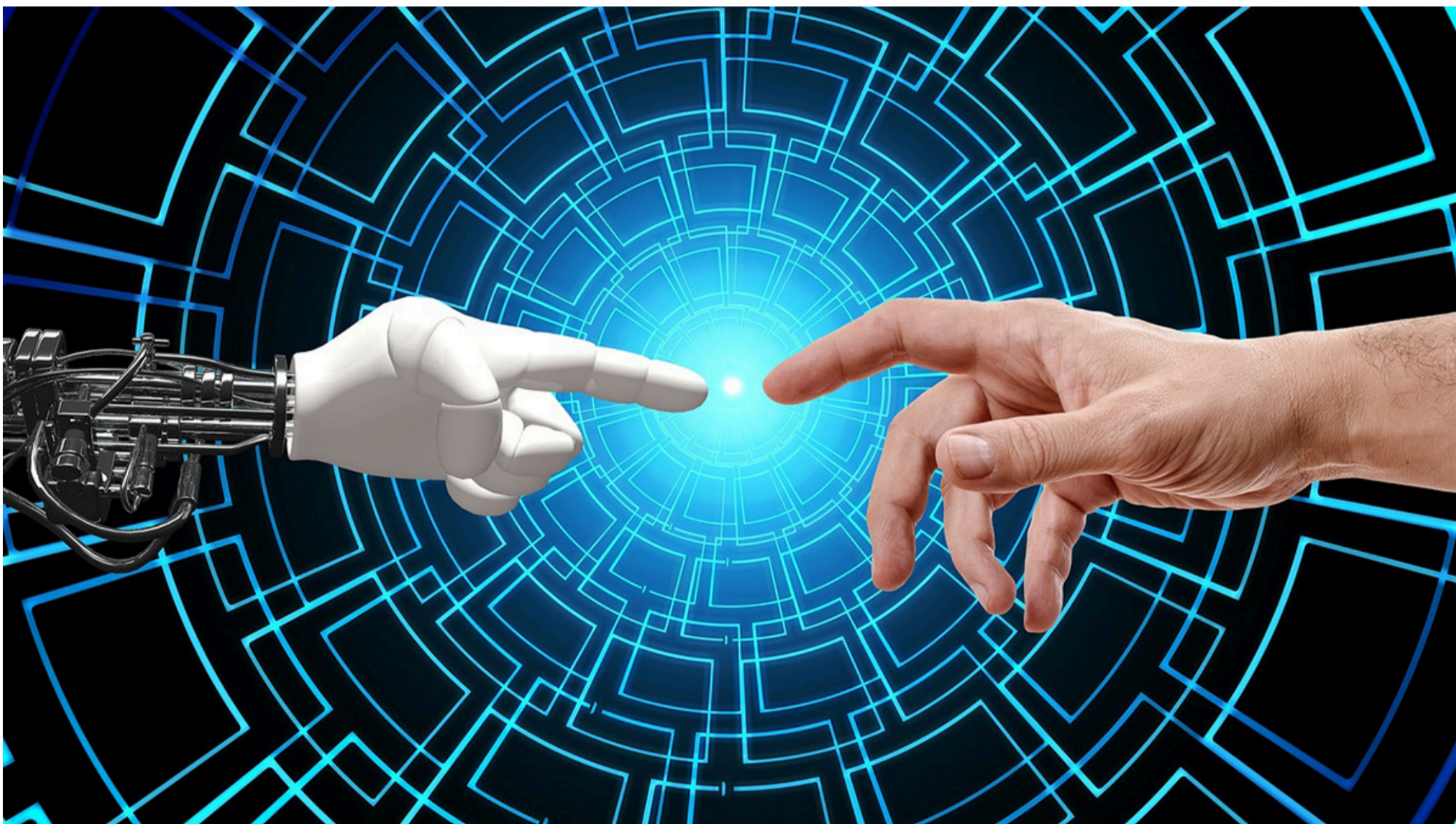
HAASTE

Haaste-lehti käsittelee rikoksentorjuntaa ja kriminaalipolitiikkaa

[Rikoksentorjunta](#) / [fi](#) / [Haaste-lehti](#) / Artikkelit

EU:n tekoälysäädös algoritmisten haittojen hallinnassa

20.12.2024 7.58



Tekoäly on vakiinnuttanut paikkansa arjessa ja työelämässä, eikä siitä seuraavia yhteiskunnallisia haittoja ole aina helppoa havaita. Vaikka EU:n tekoälysäädöksellä pyritään hallitsemaan tekoälyjärjestelmien riskejä, teknisistä parametreista etäännyvät, käytön tapoihin ja käyttöympäristöön kiinnittyvät haitat jäävät siinä varsin vähälle huomiolle. Onko meillä keinoja hallita tekoälytransformaatiota, vai muovaako teknologia yhteiskuntaa omilla ehdoillaan?

Tekoäly on terminä mitä epämääräisin. Populaarikulttuurin dystooppiset visiot, kuten Terminaattorin Skynet tai Westworldin Androidit, eroavat merkittävästi käytännön sovelluksista, kuten asiakaspalvelu-chatboteista tai älypuhelimien tekoälyassistentteista. Nämä kaikki voidaan kuitenkin luokitella tekoälyksi: konepohjaisiksi järjestelmiksi, jotka pystyvät toimimaan ainakin jossain määrin itsenäisesti saavuttaakseen määriteltäviä tavoitteita.

Vaikka tekoälyjärjestelmät ovat viimeisten vuosien aikana kehittyneet aino harppauksin, aivan lähitulevaisuudessa tuskin siintää tekoälyvallankumous. Tekoälyjärjestelmät ovat kuitenkin vakiinnuttaneet asemansa nykypäivän yhteiskunnassa. Tekoäly nopeuttaa päätöksentekoa, helpottaa arkea ja tehostaa työntekoa: luo sisältöjä, tarkistaa tekstejä, jopa ehdottaa diagnooseja. Nopeuden ja tehokkuuden kääntöpuolia, erilaisia haittoja, on samalla tunnistettu yhä enemmän.

EU pyrkii hallitsemaan tekoälyn haittoja ja riskejä

Haitat voidaan määritellä monin eri tavoin, mutta yhteiskunnallisten haittojen tutkijoiden kohtuullisen yhtenäisen konsensus on, että haitat ovat sellaisia yhteiskunnissa esiintyviä ilmiöitä tai asioita – tai joissain tapauksissa ilmiöiden tai asioiden puutteita – jotka heikentävät yksilön mahdollisuuksia elää hyvää, kukoistavaa elämää. Koska tekoälyteknologioilla on valtavasti potentiaalia muokata yhteiskuntaa, myös haittapotentiaali on merkittävä.

Tekoälyteknologioiden riskejä pyritään hallitsemaan EU:ssa jämäkästi. Vastikään astui voimaan tekoälysäädös, jonka tavoitteena on varmistaa, että tekoälyteknologiat ovat turvallisia eivätkä aiheuta kohtuuttomia riskejä ihmisoikeuksien toteutumiseksi. Tähän pyritään sääntelemällä tekoälyjärjestelmien ja -mallien tuotantoa ja markkinoita niiden arvioidun riskitason mukaisesti.

Sääntely on askel oikeaan suuntaan. Siinä asetetaan riskiperusteisia vaatimuksia tekoälyjärjestelmien tuottajille, jotta markkinoille päätyvät järjestelmät olisivat mahdollisimman turvallisia. Sinänsä on mielenkiintoista, että tästä pyrkimyksestä huolimatta säädöksessä ei kuitenkaan lähdetä siitä ajatuksesta, että velvoitteet täyttävä järjestelmä ei aiheuttaisi riskejä ihmisoikeuksille.

Tekoälyn käytön pitkäaikaisia haittoja ei kuitenkaan tunnisteta

Tekoälyjärjestelmien käytöstä seuraavien, erityisesti ajan kuluessa ilmenevien haittojen tunnistaminen jää sääntelyssä hyvin kapea-alaiseksi, jos ei kokonaan puuttumaan. Algoritmisten, erityisesti teknologioiden välittömistä vaikutuksista etäännyvien seurannaishaittojen hallitseminen ja vähentäminen lainsäädännön keinoin voi olla kieltämättä vaikeaa, sillä ne eivät välttämättä palaudu suoraan tuotannonaikaisiin ratkaisuihin. Niiden synty voi paikantua monimutkaisten yhteiskunnallisten, teknologisten ja taloudellisten prosessien vuorovaikutusverkkoon.

Haittoja on hyvin monenlaisia. Nykyään on jo lähes yleistietoa, että tekoälyjärjestelmien tekemät päätökset voivat olla syrjiviä, ja että algoritmit vaikuttavat koko ajan voimakkaammin siihen, minkälaista informaatiota ihmiset saavat. Tekoälyjärjestelmät vaikuttavat kuitenkin myös tavoilla, joita ei ole vielä yhtä laajasti tiedostettu. Automatisoidut prosessit esimerkiksi nopeuttavat eri organisaatioiden ja myös valtionhallinnon toimintaa, ja samalla ihmisten rooli päätöksenteossa kaventuu.

Tämän seurauksena päätöksenteon parametrit väistämättä muuttuvat. Aiemmin lähes itsestään selvinä pidetyt tavoitteet esimerkiksi päätösten oikeudenmukaisuuteen tai päätösharkinnan yksilöllisyyteen liittyen muuttavat merkitystään. Ajan saatossa muutokset voivat vaikuttaa käsityksiin siitä, mitä pidetään tavoiteltavana ja hyvänä, kenties myös siihen, mitä laillisena. Esimerkiksi hallintolaissa määriteltyjä periaatteita, jotka sääntelevät hyvää hallintoa, voidaan joidenkin vuosien kuluttua tulkita hyvin eri tavalla kuin nykyään. On helppoa nähdä, että kaikki muutokset eivät automaattisesti tarkoita parannuksia.

“ Ajan saatossa muutokset voivat vaikuttaa käsityksiin siitä, mitä pidetään tavoiteltavana ja hyvänä, kenties myös siihen, mitä laillisena.

Yksilön kannalta katsottuna automaattinen, datapohjainen päätöksenteko voi olla vaikeasti ymmärrettävissä. Ajan kuluessa ja datan lisääntyessä on oletettavaa, että päätöksissä hyödynnetään henkilöistä kertynyttä – ja kerättyä – dataa yhä enemmän. Tämä kaventaa yksilöiden mahdollisuuksia vaikuttaa omaan elämäänsä. Samaan aikaan internetin ja sosiaalisen median haku- ja suositusalgoritmit suodattavat informaatiota tavalla, joka muokkaa ihmisten ajattelua ja kaventaa ymmärrystä – ja pahimmillaan vahvistaa polarisaatiota ja ääriajattelua. Kaikki tämä voi heikentää ihmisten luottamusta yhteiskunnan instituutioita kohtaan, mikä voi vaikuttaa laajasti yhteiskunnan toiminnan mahdollisuuksiin.

Laajamittainen ennakkollinen vaikutustenarviointi olisi tärkeää

EU:n tekoälysäädöksen toivotaan pystyvän estämään monien tekoälyriskien realisoituminen. Siinä keskitytään muun muassa varmistamaan, että suuren riskin tekoälyjärjestelmien opetus- ja testausdata on laadukasta ja riskienhallinta järjestetty asianmukaisesti. Sääntelyn voikin odottaa vähentävän riskiä esimerkiksi opetusdatan sopimattomuudesta johtuviin virheellisiin päätöksiin – ja sellaisista seuraavia haittoja – melko tehokkaasti. Sääntelyratkaisuilla ei kuitenkaan päästä puuttumaan haittoihin, jotka ilmenevät ajan kuluessa ja teknologioiden muokatessa ympäristöään ja käyttäjiään.

Tällaisten haittojen tunnistamiseksi tarvittaisiin myös kattavia vaikutusten arviointeja ei vain teknisten toimintojen vaan myös käytännön sovellusten ja ympäristöjen osalta. Ilman ennakkollisia vaikutusarvioiteja monet haitat, joita sinänsä voitaisiin ennalta rajoittaa, mahdollisesti nousevat esiin vasta, kun niitä on vaikea tai mahdoton korjata.

Nykymuodossaan EU:n tekoälysäädöksessä vaaditaan ihmisoikeusvaikutusten arviointia tekoälyjärjestelmän käyttöympäristössä silloin, kun julkishallinnollisessa tehtävässä otetaan käyttöön suuririskinen tekoälyjärjestelmä. Muita tilanteita vaatimus ei koske, eikä muiden kuin suuririskiseksi arvioidujen järjestelmien vaikutuksia tarvitse arvioida edes julkishallinnossa. Suuren riskin järjestelmien listaa ei sääntelyssä ole turhan pitkäksi kirjoitettu. On selvää, että valtava määrä haittoja jää tämän myötä ennalta tunnistamatta – ja siten myös rajoittamatta. Jotta algoritmisia haittoja voitaisiin hallita paremmin, vaatimuksia vaikutusarviointista voisi olla perusteltua laajentaa ja terävöittää.

Vaikka EU:n ensiaskeleet tekoälyn sääntelyssä ovat lupaavia, hyvät ratkaisut kehittyvät harmillisen hitaasti verrattuna nopeaan teknologiseen murrokseen. Tarvitaan haitat kattavasti tunnistavat pelisäännöt ohjaamaan tekoälyteknologioiden tuotantoa ja käyttöä kestävässä suuntaan. Tekoäly ei ole vain teknologiaa – se on yhteiskunnallinen voima, joka muokkaa tapoja, joilla elämme, työskentelemme ja ajattelemme. Tätä voimaa olisi suotavaa hallita viisaasti ja vaikutuksia ennakoiden.

Nea Lepinkäinen

Kirjoittaja on tutkijatohtori Turun yliopiston oikeustieteellisessä tiedekunnassa.

Lepinkäinen, N. (2024). Algoritmiset haitat: [Tekoälyn riskit ja sääntelyn haasteet kiihtyvässä yhteiskunnassa](#). Turun yliopiston julkaisuja.

Kuva: Pixabay

[Haaste 4/2024](#)