

Litteraation numero: 103161/55
Äänitteen nimi: Hyva kysymys 24_v2
Äänitteen kesto: 38 min
Litteroitu: Tutkimustie Oy 05.09.2022

Puhujien merkinnät: Ville
Atte
Miapetra

Muut merkinnät: ... Puheenvuoro jää kesken, jatkaa puheenvuoroa
(-) Pois jäänyt sana tai sanan osa
(--) Pois jäänyt jakso
(sana) Epäselvä sana tai jakso
[hakasulkeet] Litteroijan kommentti

Huomautukset: --

[tunnusmusiikkia]

Ville: Tämä on valtioneuvoston selvitys- ja tutkimustoiminnan Hyvä kysymys -podcast. Tällä kertaa keskustellaan syrjinnän melko tuoreesta muodosta, nimittäin tekoälyä ohjaavien algoritmien vinoumista, jotka voivat luoda pahimmassa tapauksessa jopa yhteiskunnallisia ongelmia. Minä olen Ville Perttula, ja mulla on tänään vieraana tekoälyn vinoumiin keskittyvää selvitystä tekemässä olleen ajatuspaja Demos Helsingin tutkimuskoordinaattori Atte Ojanen. Hei Atte.

Atte: Hei, mukava tavata.

[tunnusmusiikkia]

Ville: Viime vuosina on saatu lukea mediasta tekoälyn tekemistä ratkaisuista, joita voidaan pitää vinoutuneina, tai paremminkin puolueellisina. Kun tutkijat laitoivat tekoälyohjelman valikoimaan rikollisia ihmisten kasvokuvien perusteella, tekoäly valikoi rikollisiksi enemmän tummaihoisia miehiä kuin valkoihoisia miehiä. Tällaisia samankaltaisia tekoälyn toimintaan liittyviä puolueellisuusvinoumia on löytynyt esimerkiksi rekrytointiohjelmista ja myös luotonantopuolelta, jossa on automaattista päätöksentekoa. Mutta Atte, lähdetään liikkeelle perusteista. Miten tekoälyyn syntyy vinouma?

Atte: Joo, eli siinä voi ehkä tunnistaa ainakin neljä tällaista eri kohtaa. Eli ylipäätään se järjestelmän kehittäminen ja sen perusteet, siinä voi olla huonosti määritelty tavoitteita sille järjestelmälle ja sen toiminnalle, varsinkin jos siinä ei ole kuultu asianomaisia ihmisiä tässä suunnittelussa. Sitten toisekseen on tämä ehkä tunnetuin syy eli datavinoumat. Ja ne on sitten puutteita tai virheitä siinä algoritmin opetusdatassa, sen edustavuudessa, tai huonoissa nimikkeissä. Sitten kolmanneksi itse algoritmi ja se malli, sen määrittäminen, testaaminen, jälkiprosessointi, voi johtaa vinoumiin. Siellä voi olla huonosti valittu ennustemuuttuja, reiluusmetriikoita ja vertailuluokkia. Ja sitten viimeiseksi, käytännön läheisemmällä tasolla, se itse järjestelmän käyttöönotto suunnittelemattomassa ympäristössä tai väestössä. Voidaan käyttää virheellisesti, ei ylläpidetä hyvin, ja se voi olla lisäksi läpinäkymätön. Nämä on kaikki riskitekijöitä sille syrjinnälle, ja varsinkin tämä läpinäkymättömyys on keskeinen lisähaaste.

Ville: Eli ensin on vinouma, ja se vinouma on aineistossa, josta tämä algoritmi tai ohjelma on tehty. Ja se saattaa johtaa syrjintään. Ensin on vinouma, ja sitten saattaa tapahtua syrjintää.

Atte: Joo, kyllä. Nimenomaan. Että syrjintä on vähän niin kuin alakäsite sitten tälle vinoumille, että kaikki vinoumat ei välttämättä ole syrjiviä.

Ville: No miten yleisestä ongelmasta tai mahdollisesta uhasta me tässä puhutaan? Onko tämä tällä hetkellä hyvinkin yleistä? Ja ehkä jopa semmoista, mitä me ei tiedetä, että tuolla tekoälyohjelmissa on?

Atte: Kyllä mä sanoisin, että voidaan pitää aika yleisenä tätä uhkaa. Että jos mietitään näitä esimerkkejä, niin Suomessa meillä on esimerkiksi ollut tällainen luottokelpoisuuden arviointiin liittyvä case, jossa tosiaan käytettiin vain tällaisia tilastotietoja ihmisistä. Ja sitten tosiaan yhdenvertaisuus- ja tasa-arvolautakunta piti tätä kuitenkin syrjintänä. Eli kulutusluottoa ei voi evätä pelkästään sukupuolen tai asuinpaikan, niihin liittyvien tilastollisten asioiden perusteella. Ja siitä määrättiin jopa sadantuhannen euron uhkasakko. Ja sitten meillä on esimerkiksi Alankomaista esimerkki, jossa siellä paikallinen sosiaali- ja työministeriö käytti tällaista SyRi-nimistä algoritmia sosiaaliturvapäätösten tunnistamiseen, ja arvioi sitten henkilötietojen perusteella ihmisten tällaista petosriskiä. Ja sitä käytettiin erityisesti tällaisissa maan köyhimmissä naapurustoissa. Ja sitten se tuotti virheellisiä päätöksiä, ja toimi tosi läpinäkymättömästi. Ja lopulta sitten tosiaan tuomioistuimien siellä Haagissa tuomitsi sen käytön laittomaksi nojaten just tällaisten perusoikeuksien rikkomuksiin, ja tietysti yksityisyyden ja mahdollisen syrjinnän osalta. Ja sitten toki, niin kuin tuossa aikaisemmin viittasit, niin tosiaan Yhdysvalloissa on kanssa aika paljon näitä varottavia esimerkkejä juuri vaikka kasvojen tunnistuksesta, ja monista muista vaikka rekrytointiin liittyvistä esimerkeistä. Toki Suomessa ei ehkä olla ihan näin pitkällä, mutta näihin on syytä varautua.

Ville: Oliko tässä luotonantotapauksessa, mistä äsken kerroit, niin syntyikö tämä syrjintä ikään kuin siitä, että se teki liian tammöisellä mustavalkoisella aineistolla niitä päätöksiä?

Atte: No joo, siinä mielessä, että ainakin käytännössä siinä kyseissä casessa, jos tämä hakija olisi ollut esimerkiksi suomenruotsalainen tai nainen, hän olisi saanut sen lainan. Mutta hän oli mies, niin se oli ilmeisesti se ratkaiseva tekijä siinä. Näin puhtaasti korrelaatiot ei sitten..

Ville: ..kerro ihan koko totuutta tästä lainanhakijasta.

Atte: Niin, kyllä.

Ville: Joo, se on ihan helppo ymmärtää. Demos Helsinki on tehnyt nyt selvityksen tekoölyn vinoumista ja niiden välttämistä Tampereen ja Turun yliopistojen kanssa. Onko tämän selvityksen lähtökohta ollut se, että nyt on korjausliikkeen aika?

Atte: En tiedä puhuisinko ehkä varsinaisesta korjausliikkeestä, mutta sikäli kun algoritmien käyttö lisääntyy yhteiskunnassa, niin kyllä mä näen, että algoritmisiin vinoumiin ja siihen syrjintään on keskittyttävä enemmän toki. Varsinkin julkisella puolella. Ja tämä johtuu mun mielestä ehkä keskeisesti niistä riskeistä, että tällainen koneoppiminen suurista datamassoista, niin se lisää erityisesti riskiä tällaiseen moniperustaiseen välittömään syrjintään. Ja voi sitten huomattavasti pahentaa meidän olemassa olevia yhteiskunnallisia eriarvoisuuksia. Ja aika uudessa skaalassa, ja ehkä varsin näkymättömästi. Että siinä mielessä tällaisen syrjinnän automaatio on keskeinen uhka. Ja sitten ehkä toisekseen, jos mietitään korjausliikettä, niin ehkä se mitä me ollaan haluttu tuoda tähän mukaan, on myös tämä yhdenvertaisuuden edistäminen, joka on ehkä tähän mennessä osittain puuttunut tästä keskustelusta algoritmista vinoumista.

Ville: Niin, kyllä. Että nyt on ainakin jollain tasolla tammöinen, on aika tulla tietoisiksi näistä uhista.

Atte: Kyllä.

Ville: Ja saada siihen vähän tapoja, joilla niitä voidaan myös hallita. Mutta mikä on sun käsitys siitä, missä vaiheessa Suomessa ollaan tekoölyn kehityksen kanssa?

Atte: No, tavallaanhan Suomi näyttää hyvin näissä tällaisissa vertailuissa, mitä on tekoölyn käytöstä. Mutta ainakin jos tarkastellaan julkista sektoria, niin kyllä sitä tilannetta voi ehkä kuvailla kuitenkin aika maltilliseksi, mitä tulee siihen tekoölyn käyttöönottoon. Mutta toisaalta tämä on, meidän nähdäksemme, myös mahdollisuus varautua syrjintään, sekä myös siihen yhdenvertaisuuden edistämiseen, siitä tekoölyjärjestelmän kehityksen alusta asti, jolloin ne toimet näihin asioihin puuttumiseksi on kaikkein tehokkaimpia. Ja kuitenkin ainakin ne organisaatiot, mitä me tässä meidän selvityksessä tutkittiin, niin oli etenemässä tekoölyn soveltamisen kanssa muutaman vuoden aikana. Eli tilanne voi näyttää aika erilaiselta sitten kuitenkin tulevaisuudessa, ainakin mitä tulee julkiseen puoleen. Yksityisellä sektorilla ne järjestelmät ja organisaatiot, mitä tutkittiin, niin oli ehkä, toimivat vähän kapeammalla alueella, ja kohdistuivat ehkä, ei niin yhteiskunnallisesti merkityksellisiin asioihin. Mutta sielläkin on toki riskejä.

Ville: Kyllä, juuri näitä aloja, mitkä tässä on tullut jo mainittuakin. Mutta te tutkijat ehdotatte tässä selvityksessä, että tietoisuutta tekoölyyn liittyvistä syrjintäriskeistä pitäisi lisätä. Yleisesti, mikä on sun käsitys siitä, että miten hyvin näitä riskejä nyt jo tunnetaan?

Atte: Tavallaanhan niitä tunnetaan ehkä aika hyvin yleisellä tasolla, että tämä ei ole sinällään ihan uusi debatti. Että se ilmiö on sinällään tunnistettu yhteiskunnassa. Ja myös näissä, esimerkiksi organisaatioissa, joita me haastattelimme, niin sielläkin oltiin tietoisia asiasta. Mutta sitten ne konkreettiset toimenpiteet ja käytännöt vielä puuttui. Ja ehkä tätä jossain mielessä voi myös peilata yhdenvertaisuuteen ylipäätään, että sehän on myös aika laajalti tunnistettu yhteiskunnallinen ongelma, mutta sitten ne konkreettiset toimenpiteet on vaikeampia. Ja meidän on ehkä vaikea asettaa itseämme sitten vaikka vähemmistöjen saappaisiin ja edistää sitä yhdenvertaisuutta käytännössä. Ja sitten ehkä sanoisin vielä, että kun puhutaan siitä tietoisuuden edistämisestä, niin ehkä me viitataan juuri sitten tällaisiin nyansseihin. Eli siihen, että esimerkiksi meidän mielestämme tämä ei ole pelkästään tekninen ongelma, eikä siihen ole pelkästään tällaisia universaaleja teknisiä ratkaisuja olemassa, vaan ne on kuitenkin tällaisia yhteiskunnallisia ongelmia, jotka vaativat yhteistyötä, luonteeltaan sosioteknisiä. Ja tosiaan niihin syrjiviin vaikutusten realisointumiseen vaikuttaa monet sosiaaliset, kulttuuriset ja teknologiset tekijät. Eli vaikkapa se tosiaan järjestelmän lähtökohtainen käyttötarkoitus ja oikeutus, sen käyttäjien toiminta ylipäätään, väestö, johon sitä sovelletaan, sen käytön skaala, ja sitten kaikki muutokset näissä. Eli tosiaan syrjivät vinoumat sitten syntyvät useimmiten sen tekoölyjärjestelmän arvoketjussa erilaisten vinoumien yhteisvaikutuksesta. Eli teknologia ei sinällään ole neutraalia koskaan.

Ville: No tässä selvityksessä kerrotaan, että julkinen puoli on vasta pilottivaiheessa, mitä tulee näihin tekoölyjärjestelmiin. Niin tuoteko te tässä myös tykö sen, että tämä on vasta niin uusi juttu, että nämä senkin takia saattaa olla julkiselle puolelle riskejä, jotka tiedetään, mutta joihin ei ole vielä pystytty reagoimaan?

Atte: Joo, kyllä. Että myös nämä haastatellut ihmiset ei nyt sinällään, että heillä on muutaman vuoden kokemus asiasta ja näin. Mutta tosiaan, että ollaanhan tässä uuden riskin ja uhan kanssa sinällään liikkeellä.

Ville: Kyllä. Tässä selvityksessä kerrotaan myös, että olemassa olevilla tekoölyjärjestelmillä on jo suuria yhteiskunnallisia vaikutuksia ihmisten perusoikeuksiin asti. Ja tekoölyvinoumista siis voi tulla myös yhteiskunnallisia vinoumia, mitkä liittyvät tähän syrjintään sitten. Mutta milloin voidaan puhua tekoölyn aiheuttamasta yhteiskunnallisesta ongelmasta? Tämä kuulostaa vakavalta asialta, mitä se toki onkin, mutta milloin ollaan näin pitkällä?

Atte: Niin, hyvä kysymys. Ja mielenkiintoistahan on se ehkä, että syrjintää on yleensä lähestytty tällaisena yksilötason ongelmana, että jotakin yksilöä syrjitään suoraan. Mutta kyllä meidän nähdäksemme nämä syrjivät vinoumat on sitten käytännössä aina yhteiskunnallisia ongelmia. Varsinkin, kun ne sitten ilmenevät välillisesti epäsuorasti ihmisryhmien tasolla. Ja ehkä ylipäätään tämä diskurssi algoritmista vinoumista, niin kyllähän se on vahvasti tällaisissa ryhmätason eroissa, ja kaikissa tavoissa siihen puuttua. Niin kyllä siinä sitten käytännössä vertaillaan näitä ihmisryhmiä, ja eroja heidän välillään sitten näissä järjestelmien toiminnassa. Että se on ehkä se taso, missä ne on helpoin havaita sitten ne mahdolliset vinoumat. Eli siinä mielessä tämä ehkä eroaa jossain mielessä sellaisesta perinteisemmästä syrjinnän diskurssista, ja saattaa myös aiheuttaa uusia ongelmia ja haastaa sitä.

[tunnusmusiikkia]

Ville: Mennään lainsäädäntöön. Meillä on Suomessa yhdenvertaisuuslaki, joka on aika keskeisessä osassa tässäkin selvityksessä. Mutta miten hyvin tämä laki ottaa huomioon tämmöiset tekoälyohjelmien vinoumat?

Atte: Niin, me ehkä lähestyttiin sitä vähän toisinpäin, eli miten hyvin tekoälyn kehityksessä ja käytössä huomioidaan yhdenvertaisuuslaki, jonka tavoite on sitten edistää yhdenvertaisuutta ja ehkäistä syrjintää, ja tehostaa sen syrjinnän kohteeksi joutuneiden oikeusturvaa. Tietenkään siinä laissa ei eksplisiittisesti viitata tekoälyyn, mutta ehkä tekoälyn perspektiivistä on hyvä, että siinä yhdenvertaisuuslaissa kuitenkin, siinä on se edistämismuoto viranomaisilla, koulutuksen järjestäjillä ja työnantajilla. Ja tämä on sovellettavissa myös tekoälyyn. Ja toiseksi, Suomessa tämä yhdenvertaisuuslain kiellettyjen syrjintäperusteiden lista on käytännössä avoin, eli siellä mainitaan just nämä ikä ja tausta ja muu, mutta myös muut niihin verrattavissa olevat asiat. Eli tämä mahdollistaa suhteellisen laajat toimenpiteet sen, ylipäätään ehkä syrjinnän havaitsemiseksi, ja sitten sen yhdenvertaisuuden edistämiseksi, mikä on sitten ehkä tekoälyn kohdalla erityisen tärkeää sikäli, kun syrjintä voi olla aika epäsuoraa usein.

Ville: No yksi teidän selvityksen havainnoista oli se, että syrjimättömyyden minimoimiseen ei ole vakiintunut selkeää viranomaisyhteistyön mallia esimerkiksi tietosuojavaltuutetun kohdalla. Mutta tähän kuulostaa aika vakavalta, mitä sä olet itse siitä mieltä?

Atte: No näin voisi sanoa, joo. Mutta toisaalta, niin kuin ollaan puhuttu, on tämä melko uusi aihe, että se on ehkä luontevaa. Ja se lainsäädäntö vielä kehittyy. Mutta toisaalta myös esimerkiksi tämä tietosuojavaltuutetun rooliesimerkki, niin sehän on ihan positiivinen siinä mielessä, että tätä mukailemalla me sitten voitaisiin saada ehkä parempaa yhteistyötä tekoälyn kehittäjien ja julkisen vallan välillä, mitä tulee näihin yhdenvertaisuuskysymyksiin. Että se on ehkä yksi tällainen tie, mitä me maalailaan tässä.

Ville: Kyllä. Entäs tuo yhdenvertaisuusvaltuutetun rooli, pitäisikö sen roolia kasvattaa johonkin suuntaan Suomessa? Miten hyvin yhdenvertaisuusvaltuutettu pystyy puuttua näihin tämmöisiin tekoälysyryntätapauksiin?

Atte: No joo, kyllä me ehdotetaan raportissa, että sitä resursseja ja kykyä keskittyä näihin algoritmisen syrjinnän ongelmiin olisi syytä kasvattaa yhdenvertaisuusvaltuutetun kohdalla. Että se on siellä huomioitu, mutta hirveästi resursseja tähän ei ole.

Ville: Onko tämä siis resurssikysymys vai valtuuskysymys?

Atte: No kyllä mä sanoisin, että se on ensisijaisesti resurssikysymys. Mutta toki mä näen, että siinä on myös jotakin valtuuskysymyksiä. Eli esimerkiksi voitaisiin nähdä, että yhdenvertaisuusvaltuutetulla voisi olla laajempi kyky toteuttaa jotain auditoiteja näihin järjestelmiin ajoittain, valtuus pyytää tietoja, heille voitaisiin toimittaa jotakin dokumentaatiota näistä järjestelmistä. Mutta sitten sanoisin myös, että tällaisen suoran valvonnan sijasta, niin ehkä näen tilaa myös sellaiselle yleiselle tietoisuuden lisäämiselle näistä riskeistä, koulutukseen siitä, ja muuhun avustamiseen. Että eihän kukaan yksityinenkään tekoälykehittäjä sitten tieten tahtoen kehittä syrjivää tekoälyjärjestelmää, että siinä mielessä tällainen sparraus olisi tervetullutta.

Ville: Varmaan juurikin näin. Miten tietosuojavaltuutettu sitten istuu tähän palettiin? Eikö näistä osa voisi mennä tämänkin viranomaisen tontille?

Atte: Joo, siinä on kyllä ehdottomasti jonkin verran tällaista päällekkäisyyttä. Mutta toki tietosuojavaltuutetun kohdalla ne kysymykset sitten keskittyvät ensisijaisesti siihen henkilötietoihin ja niiden käyttöön. Mutta kyllä me esimerkiksi tässä hankkeessa kehitetty arviointikehikko nähdään juuri tällaisena asiana, joka voi täydentää myös sellaista vaikutusten arviointia, mitä suoritetaan myös tietosuojaan kohdalla. Eli näissä on kyllä yhteneväisyyksiä.

Ville: Entäs näitten tekoälyn yhdenvertaisuusvaikutusten arviointi, niin pitäisikö niistä tehdä jollain tavalla pakollisia? Tai pitäisikö siihen vakiinnuttaa jonkunnäköinen malli? Kanadassa esimerkiksi on tehty näin. Niin olisiko tämä jonkunnäköinen ratkaisu?

Atte: Joo, kyllä tämä on jotain, mitä me ehdotetaan siellä raportissa. Ja oikeastaan tulkinnasta riippuen, en näe, että se on edes mitenkään kovin radikaali ehdotus. Että tämän voi nähdä jo asiana, jota yhdenvertaisuuslaki jossain mielessä vaatii, jos sitä sovelletaan kunnolla. Ja ehkä myös yksityisen puolella voisi olla jotain tarvetta tällaiselle. Mutta tosiaan niin kuin sanoit, niin Kanadassahan on tosiaan tällainen automaattisen päätöksenteon direktiivi, ja sen alla kehitetty tällainen algoritmisen vaikutusten arvioinnin työkalu, jota sitten Kanadan keskushallinnon virastojen on käytävä läpi ja täytettävä ja toimitettava eteenpäin silloin, kun he ottavat käyttöön tällaisen automaattisen päätöksenteon järjestelmän. Ja hirveästi tästä ei ole tietoa, että kuinka hyvin se nyt on toiminut, käsittääkseni ihan mahdollisesti näitä selvityksiä siellä ei ole tehty. Mutta jonkin verran niistä löytyy sitten kuitenkin dokumentaatiota, ja se lienee tärkeintä. Ja sitten

ehkä toisena esimerkkinä Alankomaissa ehkä hyvin samanlainen asia, jopa korkeammalta tasolta, että siellä keväällä Alankomaiden parlamentti tosiaan teki pakolliseksi tällaisen ihmisoikeuksien arvioinnin, kun algoritmit otetaan käyttöön julkisella sektorilla. Eli nämä on ehkä tällaisia ihan hyviä esimerkkejä.

Ville: Ja oliko näissäkin lähtökohtana syrjinnän estäminen?

Atte: Joo, kyllä. Joskin toki, ehkä sanottaisiinko, että ainakin Kanadan kohdalla se oli vähän laajempi, että siellä oli myös tällaisia yksityisyyteen ja läpinäkyvyyteen liittyviä asioita.

Ville: Kyllä. Siitä kun tarjosit noin hyvän aasinsillan, niin mennään suoraan tähän päätöksenteon läpinäkyvyyteen. Ja sehän on iso ongelma, mitä tulee näihin tekoälyjärjestelmiin ja automaattiseen päätöksentekoon. Eli kyse on siis siitä, että ihminen, joka on päätöksen kohteena, niin ei välttämättä tiedä, miten päätös on syntynyt. Ja tällainenhan voi olla aika ongelmallista, esimerkiksi jos hakee luottoa ja saa vain hylkäävän päätöksen, että et saa luottoa. Niin saattaisi olla aika tärkeää, että kansalainen sitten tietäisi, että minkä takia tätä luottoa ei ole saatu. Mutta mitä näin tutkijana ajattelet tästä asiasta, miten tätä päätöksentekoa voitaisiin läpivalaista?

Atte: Joo, tämä on toki keskeinen haaste juuri sen kannalta, että sitten tämä päätöksenteon kohde ei välttämättä, tai ihminen ei pääse sitten oikeuksiinsa, ne ei toteudu, ja on vaikea valvoa, haastaa ja korjata näitä tehtyjä päätöksiä sitten, jos ne on läpinäkyvyyttä. Mutta tosiaan tähän on olemassa eri menetelmiä, eli tällaisia vaikka teknisiä selitysmenetelmiä, jolla voidaan yrittää sitten vastata näihin haasteisiin tuottamalla selityksiä siitä järjestelmän toiminnasta, ja niistä taustalla olevista syistä niille tuloksille. Mutta niiden käytössä on kuitenkin erilaisia haasteita. Ja sitten tällainen mallin yksinkertaistaminen voi myös heikentää sen osumatarkkuutta, ja altistaa sen myös väärinkäytölle, että se ei ole mitenkään yksinkertaista. Ja sitten toki tätä läpinäkyvyyttä voidaan edistää myös muilla ei-teknisillä toimilla, niin kuin sen mallin avoimella auditoinnilla, vaikka kansalaisyhteiskunnan tai viranomaisten tahoilta. Sitten dokumentaatiolla siitä toiminnasta, eri rekistereillä, ja sitten designratkaisulla. Josta siis päästään siihen, että oikeastaan ehkä se läpinäkyvyys ei ole se välttämättä kaikkein keskeisin sana, vaan ehkä ymmärrettävyys ja selitettävyys ihmisille, jotka on päätöksenteon kohteina. Eli esimerkiksi me tässä meidän arviointikehikossa puhutaan, tai kysytään, että ilmoitetaanko henkilöille, joihin tekoälyjärjestelmä vaikuttaa, sen käytöstä, päätöksentekoprosessista ja niiden perusteista läpinäkyvästi sekä saavutettavasti. Ovatko järjestelmän toiminta sekä päätökset selitettävissä ja ymmärrettäviä asianomaisille. Eli toisin sanoen, tällainen tosi tyhjentävä ja monimutkainen looginen selitys siitä algoritmin toiminnasta, ei välttämättä ole aina tarkoituksenmukainen tapa informoida sitten ihmisiä kuitenkaan.

Ville: Mutta se, että nimenomaan päätös on ymmärrettävissä ja syyt, mitkä johtaa päätökseen, niin ne.

Atte: Kyllä. Ja ylipäätään ehkä se, tässä ollaan päätöksenteon kohteena, että sekään ei ole aina edes selvää.

Ville: Niin. Miten sä suhtaudut siihen, että julkisen puolen organisaatioita, jotka tekee isoja perusoikeuksiin liittyviä päätöksiä näin ihmisen kohdalla, että nämä ostaa valmiiksi kehitettyjä tekoälyjärjestelmiä, joihin ei välttämättä pysty täysin vaikuttamaan siihen, miten se toimii? Eli järjestelmät ostetaan ulkopuolelta, mikä on tässä maailmanajassa täysin realismia, näinhän se tapahtuu. Mutta liittyykö tähän jonkinlaisista riskiä, ettei päästä sitten käsiksi välttämättä näihin mahdollisiin vinoumiin siinä määrin kuin pitäisi päästä?

Atte: Joo, kyllä mun mielestä ehdottomasti liittyy. Ja niin kuin sanoit, niin realismia toki lienee, että käytetään ainakin osittain valmisjärjestelmiä ja sellaisia hankitaan. Mutta varsinkin kun puhutaan yhdenvertaisuudesta, niin kyllä se nähdäkseni vaatii sellaista järjestelmän (taidorointia) [epäselvää] ja käyttökohtaista harkintaa, sen soveltamista. Mutta ehkä se riippuu myös käyttökohteesta. Eli jos kyse on vaikka jostakin, en tiedä, sanotaan vaikka sormenjälkien analysoinnista, niin se ei ehkä ole ihan niin riskialtista kuin sitten vaikka joku kasvojentunnistusjärjestelmä tai vastaavaa. Varsinkin jos se on reaaliaikaista, mikä oikeastaan ei pitäisi EU-lainsäädännönkään nojalla olla mahdollista. Mutta että joka tapauksessa tässä on eroja siinä, että kuinka vakavista uhkista puhutaan. Mutta tosiaan, kyllä se on haaste ehdottomasti yhdenvertaisuudelle, ja me on se tässä arviointikehikossakin yritetty huomioida sinällään tämä yhteistyö yksityisen kehittäjäpuolen sekä julkisen puolen välillä. Että ainakin se kommunikaatio heidän kesken on tärkeää. Ja julkinenkin puoli, vaikka he hankkisi tällaisia vähän valmiimpia sovelluksia, niin kyllä me nähdään, että tässä tällaisia, juuri meidän tässä kehikossa identifioimia asioita voidaan sieltä käyttää vähän jonkinlaisina ehtoina tässä hankinnassa, tai jotenkin ohjaamassa sitä.

Ville: Ja kuten sanottu, niin tuskinpa sinne kukaan tahallaan näitä vinoumia jättää. Että ehkä siinä on enemmän kyse kuitenkin asioista, joita ei välttämättä ole osattu ajatella. Ja tietysti se on myös toimittajan etu sitten, että se ohjelma toimii niin kuin pitää.

Atte: Kyllä.

Ville: Mutta mikä on sun näkemys siitä, miten erilaiset ryhmät, vaikka vähemmistöt, joihin tällainen syrjintä saattaa kohdistua, niin miten tällaiset ihmiset saataisiin mukaan tähän kehitykseen? Ja olisiko se ylipäätään ratkaisu näihin vinoumiin, että tällaiset erilaiset ryhmät olisi mukana?

Atte: Näkisin ehdottomasti tärkeänä tämän, että sanottaisiinko, että tällaista algoritmista reiluutta saataisiin demokratisoitua, ja monipuolistettua sitä keskustelua ja arvoja, joita tähän liittyy. Että varsinkin yksityisellä puolella se saattaa olla aika yksipuolista. Ja tämä johtuu siitä, että varsinkin nämä syrjivät vinoumat syntyvät sitten monien tahojen yhteisvaikutuksesta, että siinä olisi hyvä saada monia ääniä kuuluviin, että me ylipäätään huomataan nämä mahdolliset riskit ja vaikutukset, mitä siellä on taustalla. Eli tästä mielestä tällainen kehittäjän moninaisuus ja koulutus on tärkeää. Mutta mitä tulee tähän osallistumiseen, niin ehkä katson sitä tällaisten tikapuiden kautta, että me voidaan nähdä siellä pohjalla tällaisia kuulemisia, kyselyitä ja muita. Sitten siellä voi olla ehkä tällaista yhteiskehittämistä, co-design,

hackathon-tyylistä toimintaa. Ja sitten ehkä myös tällaisia kansalaisraateja, deliberatiivista puntarointia, tällaisia demokratiainnovaatioita, joissa sitten todella kansalaiset pääsee ääneen. Ja ne on sitten tosiaan varmaan kaikkein tehokkaimpia siellä kehityksen alussa, kun määritellään niitä järjestelmän tavoitteita ja reiluutta, ja keskitytään sitten niihin paikallisiin yhteisöihin, joissa järjestelmä ehkä mahdollisesti eniten vaikuttaa. Mutta toki näen, että sitten tässä voisi olla paljon tilaa myös kansalaisjärjestöille ja muille, tässä järjestelmien auditoinnissa, ja niiden toiminnan seuraamisessa. Ja ehkä ylipäätään, tämä on kuitenkin laaja yhteiskunnallinen ongelma ja keskustelun paikka, eli ylipäätään tällainen keskustelu siitä, että missä mielessä me halutaan automaattisen päätöksenteon ja profiloinnin ja muun lisääntyvän yhteiskunnassa, ja milloin se on hyväksyttävää, niin on ehkä tällainen laajemman keskustelun paikka.

[tunnusmusiikkia]

Ville: Eduskunnan oikeusasiamies on huomionnut, että automatisoitu päätöksentekomenettely ei täytä perustuslain vaatimuksia, koska se on avoimen, täsmällisen ja hyvän hallinnon periaatteiden vastainen. Mutta sitten taas Kelan, Verohallinnon ja Maahanmuuttoviraston mukaan automaatio on välttämätöntä nykyaikana, koska sillä säästetään rekrytoinneissa. Ja tämä on varmasti just näin, tämän takiahan näitä isoja järjestelmiä hommataan. Mutta jaatko sä tämän saman huolen oikeusasiamiehen kanssa, ja osuuko hän naulankantaan tässä?

Atte: No kyllä mä jaan. Ja mielestäni se on ihan hyvä linjaus. Ja oikeastaan nythän on syksyllä tulossa tämä yleislainsäädäntö automaattisesta päätöksenteosta eduskuntaan, ja käsittääkseni siellä on juuri vähän rajattu näin, että tällaiset kehittyneet tekoälyjärjestelmät on sitten rajattu pois, tai että niitä ei sitten voitaisi automaattisen päätöksenteon kohdalla hyödyntää samanlailla, että siihen vaadittaisiin kuitenkin ihmisvalvontaa. Mutta sanottaisiinko, että toki sitten kaikki automaattiset päätökset ei välttämättä ole samanlaista, esimerkiksi Kelan sisällä. Että siellä on joitain päätöksiä, missä ne riskit on suuremmat. Että esimerkiksi joissain päätöksissä on henkilökohtaista harkintaa, ja sitten jotkut päätökset voi olla tällaisia, joissa on sama etuus, joka kaikille kuuluu, niin sitten ehkä esimerkiksi jälkimmäisessä ei välttämättä ole sinällään riskejä, jos se vaikka automatisoitaisiin se päätöksenteko, ainakaan samanlaisia.

Ville: Vaikka lapsilisiä tai joku tämmöinen.

Atte: Niin. Eli tässä voi olla relevantteja eroja siinä mielessä. Yleisestihän tehdään tällainen erottelu automaattisen päätöksenteon välille, jossa se tekoälyjärjestelmä toimii käytännössä sen päätöksenteon päätepisteenä. Ja sitten algoritmiavusteinen päätöksenteko, jossa se tekoälyjärjestelmä toimii tukena sille päätökselle. Eli siinä on ihminen vielä sitten mukana. Ja tosiaan se ehkä rajaveto näiden välille voi olla ajoittain myös vaikeaa, että missä mielessä se sitten vaikuttaa siihen lopulliseen päätökseen. Mutta ehkä käytännön tasolla, niin ehkä omasta näkökulmasta esimerkiksi joku esitetyt veroilmoitus nyt ei sinällään, en näe siinä mitään suuria riskejä sinällään. Mutta sitten kun päästään vielä tällaiseen, vielä suoremmin perusoikeuksiin liittyviin päätöksiin, niin sitten ollaan enemmän asian ytimessä.

Ville: Mutta palatakseni vielä tuohon oikeusasiamiehen kantaan, niin miten vakava kannanotto tämä sun mielestä on? Kun tässä puhutaan kuitenkin perustuslain vaatimuksista.

Atte: Niin, onhan se sinällään vakava. Ja nähdäkseni näiden samojen asioiden takia juuri tässä yleislainsäädännössä automaattisesta päätöksenteosta, niin ollaan linjattu aika yhteneväisesti, että siellä on juuri ne hyvän hallinnon periaatteet, joita pitää noudattaa. Ja sitten vaatii myös tällaista henkilökohtaista harkintaa ja virkahenkilöiden silmiä näihin asioihin.

Ville: No näihin tekoälyn käytön mahdollisuuksiin ja haittoihinkin on EU:ssa myös herätty. Ja siellä komissio julkaisi viime vuonna ehdotuksen tekoälyn harmonisoidun säätelyn asetuksesta. Miten tarpeellisenä sä pidät EU-tason säätelyä näissä tekoälyn riskiasioissa?

Atte: No kyllä mä näkisin sen kuitenkin erittäin tärkeänä. Että kyllä nämä tekoälyn ja digitalisaatioon liittyvät asiat kuitenkin läpileikkaa jäsenmaita, ja myös näiden globaalien arvoketjujen kautta. Ja oikeastaan mun on aika vaikea sillein miettiä jotakin muuta politiikka-alaa, jolla olisi pienemmät ne erot, että kyllähän ne tosiaan leikkaa kaikkia jäsenmaita. Ja sitten voidaan miettiä myös tätä, näitä digijättien valtaa verrattuna yksittäisiin jäsenmaitiin, että se on varmasti hyvä, että tässä on EU:lla merkittävä rooli. Ja kyllä näen tämän EU-tason säätelyn sinällään tervetulleena, mutta on siellä toki myös ongelmia. Että jos katsotaan vaikka tätä ehdotettua AI Actia, niin ehkä tämän yhdenvertaisuuden kulmasta siinä on toki osittain ehkä ongelmallista jossain mielessä se, että se on kuitenkin, se AI Act perustuu tällaiseen tuotelainsäädäntöön ja sisämarkkinoihin, että halutaan varmistaa kuluttajien luottamus näihin järjestelmiin. Mutta toisin kuin esimerkiksi GDPR, niin se ei kuitenkaan perustu ihmisoikeuksiin ja niiden turvaamiseen, mikä sitten on kuitenkin algoritmisen syrjinnän kannalta se ydin tavallaan.

Ville: Niin, kyllä.

Atte: Toki ne on siellä pyritty huomiomaan, mutta ei ehkä yhtä voimakkaasti kuin GDPR:n tapauksessa. Ja sitten toki myös tässä AI Actissa, ja muissakin EU-tason ulostuloissa, on ehkä vähän liikaa tällaista uskoa teknisiin menetelmiin. Eli esimerkiksi luotetaan siihen, että jos se data on laadukasta ja siihen käytetään jotain oikomismenetelmiä, niin sitten ne lopputuloksetkin tulevat olemaan vinoutumattomia, syrjimättömiä. Mutta ihan näin yksinkertaista se ei ole. Mutta tosiaan, niin kuin sanoit, niin se tavoite toki tässä on EU:lla, että se tekoälyn säätely tulisi kokonaan EU-tasolta, ja koskisi kaikkia maita yhtenäisesti. Ja sinällään näen, että se on ihan hyvä tavoite. Mutta toki yhdenvertaisuuden kannalta siellä voisi olla joitakin kohtia, mitkä on paremminkin huomioitu.

Ville: Entäs tässä Suomessa, millaisiin asioihin sun mielestä lainsäätäjien, päättäjien pitäisi täällä keskittyä näissä tekoälyasioissa?

Atte: No ehkä tällain ylätasolla, niin me kehoitetaan esimerkiksi, että ministeriöt voisi ottaa paremmin tämän tekoälyn syrjimättömyyden kehittämisen ja käyttöönoton, tukevien työkalujen, kuten vaikka arviointikehikkojen, viemisen käytäntöön omalla hallinnonalallaan, vaikka strategiatasolla. Ja sitten että integroisi tällaiset työkalut osaksi niitä hallinnon ohjausjärjestelmiä, kun näitä järjestelmiä suunnitellaan. Että esimerkiksi käytännön tasolla vaikka valtiovarainministeriö, niin heidän rooli tällaisessa avoimen julkisen hallinnon kehittäjänä voisi olla aika tärkeä tässä. Eli vaikkapa tällaisten virastojen tulohajauksen kautta voitaisiin vaikuttaa siihen, että tällaiset arviointikehikot ja muut otetaan käyttöön sitten. Ja tosiaan sitten tässä alemmalla tasolla, niiden vaikka virastojen kohdalla, tai vaikka kaupunkitasolla, niin sitten näkisin, että tällaisia työkaluja voitaisiin käyttää just näiden tekoälyjärjestelmien vaikka kilpailutuksessa, ja siinä, että varmistetaan, että siellä on oikeasti huomioitu nämä syrjimättömyys ja yhdenvertaisuuden edistäminen.

Ville: Kyllä. Että tulohajauksen kautta, se olisi teidän mielestä hyvä työkalu?

Atte: Se on yksi potentiaalinen työkalu ainakin, mikä tässä on noussut keskusteluissa esiin, mikä on nyt ajankohtainen.

Ville: Kyllä. Onko tästä siis jotain näyttöä, että tämä voisi toimia, tai onko tämä vain ikään kuin yksi idea?

Atte: No se on ehkä yksi idea, se on yksi tällainen vipuvarsi. Koska sitä on käsittäkseni, nyt on tulohajausneuvotteluja muun muassa meneillään ministeriön ja virastojen välillä.

Ville: Kyllä. No te kehitte osana tätä selvitystä tämmöisen arviointikehikon, jota voidaan käyttää tekoälyjärjestelmien mahdollisten syrjivien vaikutusten tunnistamiseen ja arvioimiseen. Onko tässä kyse siis jonkunlaisesta auditointityökalusta?

Atte: No näin voisi myös kuvailla. Joskin ehkä, me ollaan tämä arviointikehikko nähty enemmän sellaisena työkaluna, jota on tarkoitus käyttää ennakkoivasti näitä järjestelmiä hankittaessa, kehitettäessä ja käyttöä suunniteltaessa. Että ei ehkä ideaalisti sinällään jälkikäteen. Mutta sekin on toki mahdollista. Mutta ensisijaisesti kyse on tosiaan ehkä algoritmisten vaikutusten arvioinnin työkalusta. Ja se toimii ikään kuin ohjenuorana, jota sitten voidaan seurata läpi tekoälyjärjestelmien suunnittelu- ja kehitysvaiheiden, näitä yhdenvertaisuusvaikutuksia huomioitaessa. Ja tosiaan toimijoiden välinen yhteistyö on siinä myös keskiössä. Eli se tosiaan tukee tällaista yleistä riskienhallintaa tekoälyn kontekstissa, ja korostaa näitä vastuullisia käytäntöjä siinä kehityksessä. Ja myös organisaatiossa tällaista moninaisuuden lisäämistä ja näin edelleen.

[tunnusmusiikkia]

Ville: Seuraavaksi on hyvän kysymyksen aika. Tällä kertaa studion ulkopuolelta kysymyksen esittää europarlamentin jäsen Miapetra Kumpula-Natri, joka on toiminut tekoälyä käsittelevän erityisvaliokunnan varapuheenjohtajana.

Miapetra: Hei. Minun nimeni on Miapetra Kumpula-Natri, olen Euroopan parlamentin jäsen, joka saa myös uuteen tekoälylainsäädäntöön osallistua. Syksyllä saatelemme loppuun parlamentin kantaa, ja sitten vielä neuvottelut jäsenmaiden kanssa, niin Euroopassa olisi sitten tekoälyä, sen riskikäyttöä säätelevä lainsäädäntö. Kun tekoälyä otetaan käyttöön myös julkisiin järjestelmiin parantamaan ihmisten oikeusturvaa ja tehostamaan hallintoa, niin miten tutkijana näet, täytyykö siinä aina olla ennen päätöksen toimeenpanoa kuinka tarkkaan viranomaisen vastuu, tämmöinen common sense, että onko tämä päätös järkevä ollenkaan? Tietenkin, jos joka väliin laitetaan, niin se syö tehokkuutta, mutta onko se sitten oikeusturvan kannalta tarkempi, miten tutkijana näet tämän?

Atte: Joku toki voisi ehkä kyseenalaistaa toki tämän, ehkä tämän ihmisten oikeusturvan ja hallinnon tehostamisen yhdistämisen. Tai siis mun mielestä se on hyvä tavoite, mutta toki ainakin historiallisesti nämähän on usein ollut ehkä vähän ristiriidassa. Mutta toki se on tavoiteltavaa sinällään. Mutta yleisesti toki kyllä, että kyllä mä näen, että tällainen ihmisen rooli, tällainen human in the loop -idea, se voidaan toki toteuttaa monin tavoin riippuen siitä järjestelmän luonteesta ja perusoikeudellisista vaikutuksista. Että ehkä se ei välttämättä vaadi sitä, että se virkahenkilö toimii jokaisen päätöksen varmistamisessa ja leimaa ne tavallaan. Että ehkä se sitten liittyy just tällaiseen pääsyyn auditoimaan niitä ja tarkistamaan niitä päätöksiä, joitain tällaisia järjestelmiä, jotka mahdollisesti flagaa sieltä jotain kyseenalaisia asioita, eli erinäisiä indikaattoreita, jotka huomauttaa näistä ongelmista. Sekä sitten selkeitä oikeussuojamekanismeja näitä vääriä päätöksiä kohtaan, ja niihin puuttumiseksi myös kansalaisten suuntaan. Mutta ehkä tämä korostaa just tällaisten selkeiden standardien luomista myös auditoimiseen ja näiden järjestelmien seurantaan.

Ville: Eli olet sitä mieltä, että tästä pitäisi jotain päättää, tai joku malli pitäisi luoda?

Atte: No kyllä se olisi toivottavaa.

Ville: Toivotko siis, että ihminen pysyy mukana näissä päätöksissä?

Atte: No kyllä näin voisi sanoa. Ja se on jossain mielessä ehkä myös se lainsäädännön näyttämä suunta.

Ville: Hei kiitos haastattelusta Demos Helsingin Atte Ojanen.

Atte: Kiitoksia. Oli mukava päästä keskustelemaan.

[tunnusmusiikkia]

Ville: Tämä oli Hyvä kysymys -podcast, joka auttaa päättäjiä tekemään parempia ratkaisuja. Kiitos, että kuuntelit.

[tunnusmusiikkia]